



University of
Massachusetts
Amherst

The Effects of Multiple External Mandates on Curriculum, Pedagogy and Child Activity in the Preschool Classroom

Item Type	dissertation
Authors	Strub-Richards, Kathleen A
DOI	10.7275/2176936
Download date	2024-12-13 20:52:52
Link to Item	https://hdl.handle.net/20.500.14394/38855

THE IMPLICATIONS OF STANDARDIZED TESTING
AT HEAD START

June 2007

Kathleen Strub-Richards
Comprehensive Exam

Committee Members

Professor Grace Craig, Committee Chair

Professor Claire Hamilton

Professor Alfred Karlson

INTRODUCTION.....	3
<i>Child Development and Learning.....</i>	<i>5</i>
Young Children and Assessment.....	7
Formal Assessment	7
<i>Advantages of Standardized Assessments.....</i>	<i>8</i>
<i>Criticisms of Norm Referenced Standardized Tests.....</i>	<i>8</i>
<i>Deleterious Effects Associated with the Practice of Testing</i>	<i>9</i>
<i>Formal Assessment in Early Childhood Programs</i>	<i>13</i>
<i>Discussion of Head Start’s NRS</i>	<i>13</i>
<i>Arguments for NRS</i>	<i>17</i>
<i>Criticism of the NRS</i>	<i>22</i>
Informal Assessment.....	37
<i>Advantages in Using Informal Assessment.....</i>	<i>38</i>
<i>Disadvantages in Using Informal Assessment.....</i>	<i>38</i>
<i>Informal Assessment in Early Childhood Programs</i>	<i>38</i>
<i>EEC Approved Assessment Tools</i>	<i>41</i>
CONCLUSION	Error! Bookmark not defined.
References.....	49

INTRODUCTION

In 2003 high stake tests reached its pinnacle in the state of Massachusetts when the government declared that every public school student must pass the Massachusetts Comprehensive Assessment System (MCAS) in order to receive a high school diploma. The MCAS test, was not unexpected, it had been ten years in development, having originated during the 1993 Education Reform Act (McElhenny, 2003) yet state wide debates ensued, teachers, parents and students were in an uproar and protests of various sorts were launched throughout the state but the law remained, high school graduation requirements now included passing the MCAS (McElhenny, 2003; Baumann & Perlman, 2006).

In that same year another mandated high-stake test was implemented to a specific population within the entire nation. Though the public protest for this test was not as loud and powerful as that for the MCAS the repercussions could be just as alarming (Zigler & Styfco, 2004; Meisels, 2006). The recipients for this test were not high school teenagers who had been exposed to various forms of testing throughout their school years but instead our youngest school children, the preschoolers. While some high school students were grappling with the anxiety of having to pass the MCAS in order to graduate 450,000 four and five year old Head Start students were administered their first mandated standardized test that would continue biannually until they moved up to kindergarten. Contrary to the MCAS, this standardized test, entitled the National Reporting System (NRS) was not officially being used to evaluate and take action on student learning though it was subjecting preschoolers to formalized tests in order to evaluate teacher and program effectiveness. Amidst various forms of controversy Massachusetts public high school students still need to pass the MCAS in order to graduate and Head Start preschoolers nationwide are still being administered the NRS. My concentration will not be on the experienced high school students but instead I will focus on how and why standardized tests are being administered to novice preschoolers and on the potential impact of this practice on teachers and children.

Why are Head Start preschoolers being subjected to biannual standardized tests? It is directly related to accountability and this is clearly visible in the “The No Child Left Behind Act” of 2001. Accountability has been a major issue in the public school systems for decades but it wasn’t until the 1990s that the term accountability became nationally mandated for public education. The movement toward greater and greater accountability in publicly funded education has had many roots. Many states have had statewide achievement tests such as the New York State

Regents Exams, which were first authorized for New York high school students in 1876 (New York State Education Department, 1987).

The No Child Left Behind Act of 2001 (NCLB) is the current version of the Elementary and Secondary Education Act (ESEA) of 1965. Established by President Lyndon B. Johnson as part of his *War on Poverty* the ESEA provided funds to local education agencies (school districts) to meet “the special education needs of educationally deprived children” (Kafer, 2004, p.1). It also provided funds for educator's professional development, instructional materials, resources to support educational programs, and parental involvement promotion. The act was originally authorized through 1970, however the government has reauthorized the ESEA every five years since its enactment. As a result of the reauthorizations, the act has undergone numerous name changes and presidencies (NEA, 2002-2006).

In 1994 the Reauthorization of the Elementary and Secondary Education Act (ESEA) was amended to require states to create performance-based accountability systems for schools. The reauthorized ESEA required states to develop academic standards, assessments based on standards, and progress goals for schools and school districts. The merger between state and federal accountability policies was slated to occur by the year 2000 (Elmore, 2002). Yet by the proposed date the provisions had not been strongly enforced and fewer than half the states had met the requirements. While most states had some level of standards and testing, school sanctions were rare. By the end of the Clinton administration only seventeen states were in complete alliance even though all were receiving funds (Kafer, 2004).

When President George W. Bush came into office he sought to close the achievement gap by toughening the accountability provision. He proposed that the federal government provide states with additional funding and flexibility in return for implementing rigorous accountability. His proposal, the “No Child Left Behind Act (NCLB)” became a law in 2001. The demands of the NCLB included that states add content standards in history and science to their previously required reading and math standards; that annual assessments be administered to every child in grades 3-8; and adequate yearly progress be demonstrated for disadvantaged children, as well as the overall student population. Schools that made significant progress would receive additional funding while schools that failed to make adequate yearly progress would receive sanctions (The White House, 2002). This act, which affected all public school children, had its repercussions in the federally funded preschool program, Head Start.

The accountability measures administered to Head Start preschoolers through the National Reporting System is a prime example of how standardized measures for accountability were grossly misused. Yet this wasn't the first time that assessment measures of young children, for the sake of accountability, were mishandled nor will it be the last (Meisels, 2006; Wagner, 2003; Shepard, Kagan, and Wurtz, 1998). It is important that parents and schools know that their students are learning, but how can we measure what young children have learned without using an assessment tool that may be harmful to them?

Public schools need to be accountable to the community they serve. The public needs to know that teachers are teaching what they purport to teach and that all children are learning what the community values. It is essential that academic standards lay the groundwork for the curriculum to ascertain that all children are learning basic language and mathematics skills. It is necessary for all children to attain a certain level of knowledge before they leave school and enter the adult world. Yet the way that we measure how and what children are learning is developmentally different between teenagers and four and five-year-olds (Kostelnik, Soderman, and Whiren, 2004). What works for older children or adults will not work for younger children; they have developmental capabilities and needs that we, as adults, are obliged to recognize if we are to optimize their development (Shepard et al, 1998, p.4). Different tools of measurement must be applied for different age groups. Unfortunately young children are repeatedly the recipients of measures that should only be administered to older students (Helm, 1997; Katz, 2000; Gullo, 2005; Meisels, 2006; Shepard et al, 1998). How are young children developmentally different than older children? This is a very important issue to discuss before one can measure how and what knowledge and competencies young children have learned.

Child Development and Learning

In order to assess and evaluate young children it is imperative that one has an understanding of some of the principles of child development and learning. The National Association for the Education of Young Children's principles of child development and learning state that:

- Domains of children's development - social, emotional, cognitive, and physical - are closely related. Development influences and is influenced by development in other domains.
- Development occurs in a relatively orderly sequence, with later abilities, skills and knowledge building on those already acquired.

- Development proceeds at varying rates from child to child as well as unevenly within different areas of each child's functioning.
- Early experiences have both cumulative and delayed effects on individual children's development. Optimal periods exist for certain types of development and learning.
- Development proceeds in predictable directions toward greater complexity, organization, and internalization.
- Development and learning occur in and are influenced by multiple social and cultural contexts.
- Children are active learners, drawing on direct physical and social experience as well as culturally transmitted knowledge to construct their own understanding of the world around them.
- Development and learning result from interaction of biological maturation and the environment, which includes both the physical and social worlds that children live in
- Play is an important vehicle for children's social, emotional, and cognitive development, as well as a reflection of their development.
- Development advances when children have opportunities to practice newly acquired skills as well as when they experience a challenge just below the level of their current mastery.
- Children demonstrate different modes of knowing and learning and different ways of representing what they know.
- Children develop and learn best in the context of a community where they are safe and valued, their physical needs are met, and they feel psychologically secure (Bredekamp & Copple, 1997, p.10-15).

True learning occurs when children make a relatively permanent change in their thinking or behavior as a result of the interaction between maturation and experience (Kostelnik et al, 2004). Learning has been attained if a person is able to transfer what she/he has learned in school to everyday settings of home, community and workplace. Successful transfer is influenced by the degree of mastery of the original subject and by the degree to which it is learned with understanding rather than just through memorization. Motivation highly influences the desire and ability to learn. Children that are motivated to learn will spend large amounts of time mastering the subject (National Research Council, 2000).

YOUNG CHILDREN AND ASSESSMENT

It makes logical sense to evaluate children and see what they are learning but we need to ask ourselves if the forms of assessment that we are employing are appropriate for young children. There are various types of assessment tools that are being utilized in various preschool programs. Which of these tools are developmentally appropriate and which are not? I will begin with a brief discussion of the reasons for assessing young children and the types of assessment that are being used in early childhood classrooms.

Why do we assess young children? Young children are assessed in order to gain an understanding of a child's overall development; to plan instruction; to identify program improvement and staff development needs; to help teachers understand how well a child is progressing within the program; and to identify children who are at risk of academic failure or are potentially in need of special services (Gullo, 2005; Epstein, Shweinhart, DeBruin-Parecki & Robin, 2004; Appl, 2000). Two overall types of assessment are typically employed in early childhood education these are formal and informal assessment. I will briefly discuss each.

FORMAL ASSESSMENT

Formal assessment generally refers to standardized tests. Standardized tests were introduced in the early 1900's when lawmakers were hoping to measure the effectiveness of learning institutions. During this great influx of mass education the challenge of providing group instruction was often compared to mass production in the factories (National Research Council, 2000). In this analogy, school administrators wanted to operate their classrooms in the same efficient manner of a factory. Children were viewed as raw materials to be efficiently processed by technical workers (teachers) to reach the end product. This method attempted to sort the raw materials (children) so that they could be treated much like an assembly line. Teachers were regarded as workers whose job was to carry out orders from their superiors – the efficiency experts of schooling (administrators and researchers). The emulation of factory efficiency fostered the development of standardized tests for measurement of the “product,” of clerical work by teachers to keep records of costs and progress, and of the “management” of teaching by central district authorities that had little knowledge of educational practice or philosophy (National Research Council, 2000, p. 132)

Standardized tests are controlled tests, which require single-answer responses, usually in multiple-choice format, within specified time frames. These tests allow educators to compare an

individual child's performance on the test to the performance of other children, who have similar characteristics. Formal assessments include developmental screening tests, diagnostic tests, readiness tests and achievement tests (Gullo, 2005).

Advantages of Standardized Assessments

There are several advantages to using standardized tests. One advantage is that standardized tests have standard administration procedures. All children receive the same instructions, are given the same amount of time and specific objective scoring procedures are used. Another advantage is that the tests scores are numerical and are thus quantifiable. Quantifiable scores, usually in the form of raw scores can be statistically transformed into standard scores or norms. This facilitates comparing a child's performance to that of other children of the same age or that same child if the test is administered more than once over time. Another advantage is that most standardized tests are norm referenced. The test has determined, on average, what score children of a comparison group should achieve and how deviations (either up or down) can be interpreted. Finally, validity and reliability can be measured in standardized tests. If a child were given the test at a different time, or in a different situation, the results would be the same (Gullo, 2005).

Criticisms of Norm Referenced Standardized Tests

Standard test administration can be a benefit of standardized testing but it can also be a detriment. Children in the early childhood classroom represent vast differences in developmental levels, prior experience, and approaches to learning, motivation and individual needs. Due to the nature of norm-referenced assessments the instructions for test administration must be strictly followed. The same procedures are utilized for all children and there is little room for modification of those procedures to meet special needs in individual children (Gullo, 2005. p. 66).

The composition of the sample on which the norm-referenced assessment was standardized must be taken into consideration. Often the assessment is biased against children of different cultural or linguistic backgrounds or against children with developmental delays or special needs (Gullo, 2005. p. 66).

Many published norm-referenced assessments do not take into account contemporary approaches to curriculum and instruction in early childhood education. They are often based on skill development approaches and reflect a theoretical perspective that is more behaviorist than constructivist. They assess specific skills or knowledge learned rather than the process of learning.

This often leads to teachers teaching to the test and thus the norm-referenced assessment has the effect of narrowing the curriculum (Gullo, 2005, p. 67).

Deleterious Effects Associated with the Practice of Testing

There are several harmful effects associated with formalized testing. The first is that *teachers may intentionally or unintentionally teach to the test in order to raise their students' scores on the tests* (Gullo, 2005, p. 70). Learning information on the test may become the primary goal. Schools may look to the test to determine curricular goals and objectives thus relinquishing their rights of curriculum development to the test writers and publishers. Assessment and evaluation can actually determine the entire program. This type of teaching is called measurement driven instruction. Some schools use test results for teacher and program accountability. Some teachers will base their decisions on what and how to teach in order to increase the likeliness of a higher end of the school year test score instead of teaching according to the needs of the students (Gullo, 2005; Meisels, 2006; Ziegler & Styfco, 2004).

Another harmful effect is that *inappropriate use of assessment information can have a detrimental effect on children* (Gullo, 2005, p. 72). Testing can lead to inappropriate labeling. Children can develop a negative sense of self-worth or perception based on test scores results (Meisels, 2006). The younger the children are the greater the risk of assigning false labels to them. The longer the children live a label, be it true or false, the more difficult it may become to discard it (Katz, 1997). The test may be as restrictive as to which particular skill or developmental domain it is measuring (Gullo, 2005).

An additional harmful effect is the *practical problems of measurement* (Meisels, 2006, p. 8). Young children have restricted ability to understand assessment clues such as verbal instructions, auditory stimuli, situational cues or written instructions. Questions that require complex information-processing skills may cause the child to give the wrong answer. Young children may also have difficulty meeting the demand characteristics of the assessment situation, such as sitting still for the time required (Meisels, 2006). These tests, which require single-answer responses within specified time frames, put enormous pressure on young children. This pressure can inhibit thinking and decrease the accuracy of the assessment. Standardized tests are insufficiently sensitive to the ways young learners demonstrate competencies. On demand assessment can interrupt the learning process in active engaging classrooms (Helm, 2000). Young children are poor test takers, perhaps because they might be confused about answering questions

that they think the test takers already knows the answers to (Katz, 1997). Young children come to know things through doing as well as through listening; they often represent their knowledge better by showing than by talking or writing thus paper-and-pencil tests are not adequate. They do not have the experience to understand what the goals of formal testing are thus testing interactions may be very difficult or impossible to structure appropriately (Shepard et al, 1998, p.3).

An additional deleterious effect is the *opportunity to learn*. Tests given at the onset of instruction do not take into consideration the variability in what children have been taught prior to entering school (Meisels, 2006, p. 9). Testing may cause educators to view the nature of children's progress in fragmented compartmentalized ways rather than in a coherent cohesive fashion. Tests may measure a restrictive range of learning and development so the teachers understanding about the child is as narrow as the behavior assessed. Young children develop and learn so fast that tests given at one point in time may not give a complete picture of their learning. Young children's achievements at any point are the result of a complex mix of their ability to learn and past learning opportunities, it is thus a mistake to interpret measures of past learning as evidence of what could be learned (Shepard et al, 1998).

A final deleterious effect is *variability and predictability* (Meisels, 2006, p. 9). There is extensive variability and change throughout childhood. Before third grade differences in developmental timetables and other factors that contribute to performance are still too unstable, impressionable, and varied to achieve reliability (Katz, 1997).

La Paro and Pianta (2002) contend that how a child performs in preschool or kindergarten is not a valid indicator of how they will perform in first or second grade. La Paro and Pianta designed a meta-analysis to estimate cross-time relations between children's early school outcomes and measures of their skills and abilities in preschool or kindergarten. They provided a quantitative summary of findings from available studies between 1985 and 1998 that predicted early school outcomes from preschool or kindergarten measures in academic/cognitive and social/behavioral domains. Skills and abilities that were measured within the academic/cognitive domain were general knowledge, intellectual development, language development and skills, literacy, numeracy, and perceptual motor skills. Constructs such as attention and work habits were considered within the academic/cognitive domain since they reflect cognitive foundations of the academic learning that takes place in the classroom (La Paro & Pianta, 2000, p. 446). The social/behavioral domain included problem behavior, peer relations and social competence. To identify relevant articles the authors compiled a list of search terms using the time frame of

interest (preschool, kindergarten and early elementary grades) and characteristics of the domains and assessments of interest. Computerized literature searches of ERIC and PsycLit databases from 1985-1998 were conducted; as well as additional studies through the reference sections and the author's personal files. This search strategy produced 757 published reports for consideration. Selection criteria was chosen to ensure that the studies sample would reflect the larger literature on school readiness as well as provide data on the relations between readiness assessment status and early school academic and behavioral status. Studies included in the meta-analysis had to meet the following criteria: involve a longitudinal design; report at least one zero-order correlation between a predictive variable and a criterion variable measuring academic/cognitive or social/behavioral outcomes, or both; academic/cognitive predictor and outcomes measures had to address general knowledge, intellectual development, language development and skills, literacy, numeracy, and perceptual-motor skills; social/behavioral predictor and outcomes measures had to focus on problem behavior, peer relations and social competence; the initial assessment of academic/cognitive and/or social/behavioral status had to have occurred when the children were at least 36 months old; the follow up assessment had to have occurred no later than second grade or when the children were seven years of age; and the study had to be published in an English-language peer-reviewed journal. ERIC documents were excluded (p La Paro & Pianta, 2000, p. 448). A total of 414 studies were eliminated. Of the 343 remaining studies closer examination found that 273 articles did not meet the criteria and were thus excluded from further review. 70 published reports presenting information on stability of individual differences in academic/cognitive and social/behavioral domains were selected for inclusion in this meta-analysis. These reports provided data from 62 independent samples that were the focus of subsequent analysis.

A wide variety of measures were used as predictor variables and outcome measures representing the range of components included in various definitions of readiness. Standardized assessments instruments such as the Developmental Indicators for the Assessment of Learning, the Peabody Picture Vocabulary Test, the McCarthy Scales of Children's Development, and the Wechsler Preschool and Primary Scale of Intelligence were the most commonly used academic/cognitive predictor and outcomes measure. Social/behavioral measures included parent reports of behavior and teacher rankings and reports of classroom behavior.

La Paro and Pianta found that the estimates for the academic/cognitive domain indicate that, on average, 25% of the variance in early school academic/cognitive performance is predicted

from preschool or kindergarten academic/cognitive status (La Paro & Pianta, 2000p.474). This meta-analysis provided quantitative evidence that results from early assessments make at best only small to moderate contributions to the predictability to children's early school success.

In contrast social/behavioral assessments at preschool or kindergarten account for 10% or less of the variance in social /behavioral measures in kindergarten, first or second grade. The results of this analysis indicate that the prediction of social behavioral outcomes is relatively poor. The difficulty in prediction is probably due to the fact that social behavior is often quite different in dissimilar contexts and unstable overtime. Behavior in school is also affected by many factors including home experience, teaching styles and the child's temperament. The moderate to small effect size for academic/cognitive and social/behavioral outcomes indicates "instability or change may be the exception rather than the rule during this time period (La Paro & Pianta, 2000, p. 476). La Paro & Pianta concluded that tests used to make predictions are insufficiently stable to assigning stakes based on them. This study demonstrates that how a child performs in preschool or kindergarten is not a valid gauge of how they will perform in first or second grade (Meisels, 2006).

The 2002 joint position statement of the National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education, states that "Broad, significant content cannot be assessed with narrow instruments. Beyond the important requirements of technical adequacy (reliability and validity), assessments must also be developmentally valid, including observations by knowledgeable adults in real-life early childhood contexts, with multiple, varied opportunities for assessment over time. Of special importance when developing standards-related assessments are the needs of culturally diverse children and children with disabilities. In addition, the information yielded by these assessments must be useful to practitioners and families.

The National Education Goals Panel's *Principles and recommendations for early childhood assessments* states that before age 8, standardized achievement measures are not sufficiently accurate to be used for high-stake decisions about individual children and schools. The misuse of assessment and accountability systems has the potential to do significant educational and developmental harm to vulnerable young children. Children's failure to meet standards cannot be used to deny them services, to exclude them from beneficial learning opportunities, or to categorize them on the basis of a single test score. Therefore, high-stake assessments intended for accountability purposes should be delayed until the end of third grade (Shepard et al, 1998).

Many experts in the field of Early Childhood Education strongly oppose administering standardized tests to young children. Research has shown that standardized tests can be harmful to young children. The assessment is often biased against children from different cultures or learning abilities. Testing can lead to inappropriate labeling. If children do poorly on a test they can develop a negative sense of self-worth. Children cannot be measured based upon one moment in time. There are too many other environmental and social/emotional that must be taken into consideration. Preschoolers have restricted ability to understand assessment clues such as verbal instructions, auditory stimuli, situational cues or written instructions. Standardized tests assess specific skills or knowledge learned rather than the process of learning. Teachers may teach to the test instead of focusing on the needs of the children. Standardized tests are developmentally inappropriate to how young children learn. Why, then, when there is so much opposition to the formalized testing of young children, did Head Start choose to mandate this form of testing biannually to all its four and five year olds?

Formal Assessment in Early Childhood Programs

Various types of formal assessment are used in preschool programs. There are developmental screening tests which are used to identify children who may be in need of special services; diagnostic tests that are used to identify the existence of a disability or specific area of academic weakness in a child; readiness tests which assess the degree to which children are prepared for an academic or pre-academic program and achievement tests that measure the extent to which an individual has achieved certain information or attained skills that are identified within curricular objectives (Gullo, 2005, p.46). There are also formalized tests that are administered to young children to evaluate teacher and program effectiveness for the sake of accountability. The most controversial example of this is the standardized achievement test, entitled the National Reporting System, which is being administered biannually to every English and Spanish speaking Head Start preschooler. Why did Head Start agree to implement such a test? I will give a brief review of Head Start and then discuss the rationale behind the implementation of the NRS.

Discussion of Head Start's NRS

Project Head Start was designed as part of President Johnson's War on Poverty in the mid-1960s. Due to 1954 *Brown v. the School of Education* decision and the Civil Rights Act of 1964, President Johnson believed that it was the nation's duty to provide not just legal equality but also

equality of opportunity to all children regardless of race or socioeconomic status. Johnson's War on Poverty included many initiatives to bring blacks and other disadvantaged Americans on equal ground with the skills and abilities necessary to compete at the same level. The idea for Head Start came about from the observation that, on average, poor and minority children arrive at school already behind their peers in the intellectual skills and abilities required for academic achievement. These deficits lead to poor performance in school, which thus narrows the economic opportunities underprivileged children encounter when they become adults (Haskins, 2004; Vinovskies, 2005).

In 1964, the Federal Government asked a panel of child development experts, from John Hopkins, to draw up a program to help communities meet the needs of disadvantaged preschool children. The panel report became the blueprint for Project Head Start (PHS). PHS, launched as an eight-week summer program by the Office of Economic Opportunity in 1965, was designed to help break the cycle of poverty by providing preschool children of low-income families with a comprehensive program to meet their emotional, social, health, nutritional, and psychological needs. In 1969, Head Start was transferred from the Office of Economic Opportunity to the Office of Child Development in the U.S. Department of Health, Education and Welfare, and has now become a program within the Administration on Children, Youth and Families in the Department of Health and Human Services. The program is locally administered by community-based non-profit organizations and school systems (Head Start Bureau, 2002).

Head Start provides grants to local public and private agencies to provide comprehensive child development services to children and families. Intended primarily for preschoolers from low-income families, Head Start's mission was to promote school readiness to enable each child to develop to his or her fullest potential. Head Start children receive comprehensive health services, including immunizations, physical and dental exams and treatment, and nutritional services. Head Start engages parents in their children's learning and helps them make progress toward their educational and employment goals. By the year 2002, Head Start's budget had risen to \$6.5 billion. The program was serving 915,000 children, approximately 65 percent of all eligible 3- and 4-year-olds through a network of 1,545 local grantees (GSGS, 2002).

Wake of the National Reporting System

During one cycle of re-authorization the federal government exhibited its influence over Head Start more powerfully. In 1999 Congress mandated that Head Start programs implement standards of learning in early literacy, language, and numeracy skills in order to be reauthorized. These

standards of learning goals included developing phonemic, print, and numeracy awareness; understanding and using language to communicate for various purposes; understanding and using increasingly complex and varied vocabulary; (for non-English background children) to progress toward acquisition of the English language; recognize a word as a unit of print; identify at least 10 letters of the alphabet and associate sounds with written words (Schumacher, Greenberg & Mezey, 2003).

In 2001 President George Bush, in an attempt to close the achievement gap initiated the 2001 reauthorization of Title 1 entitled *No Child Left Behind (NCLB)*. This federal act required that all states, school districts, and schools be accountable for ensuring that all students meet high academic standards. It was imperative that each state develop a system of sanctions and rewards to hold districts and schools accountable for improving academic achievement; annual reading and math assessments were to be administered; and specific consequences would be bestowed upon schools that failed to make adequate yearly progress (The White House, 2002).

In 2002, in the wake of the NCLB, the Bush Administration alleged that the standards of learning specified in the Reauthorization of Head Start 1999 had not been fully implemented. In order to ensure Head Start programs were effectively preparing children to meet standards of learning an early childhood initiative, entitled “Good Start, Grow Smart (GSGS),” was implemented. In an attempt to strengthen Head Start the GSGS wanted to ensure that Head Start programs were evaluated on whether they effectively prepared children to meet standards of learning. All 3-5 year old children would be assessed two times a year to analyze the assessment data on the progress and accomplishments of all enrolled children. Head Start teachers would be trained to use the best methods of early reading and language skills instruction in order to better teach to these standards. *Early Learning Guidelines* (state guidelines on literacy, language, and pre-reading skills activities for children ages 3 to 5 that align with State K-12 standards) would be strongly encouraged (GSGS, 2002).

As part of the 2003 re-authorization an attempt to improve the school readiness of disadvantaged children was initiated. Known as the School Readiness Act of 2003, significant changes were once again administered to Head Start. The new provisions for Title 1 included an increased teacher formal education qualification (50% of center-based teachers would have to have at least a baccalaureate degree in early Childhood Education (ECE) ... by September 30, 2008. Within 3 years all new Head Start teachers would have to have at least an associates degree in ECE or be currently enrolled in a program to earn an associates degree in ECE within three

years of the date of hire). The bill would insert language emphasizing prereading, premathematics and language skills as a priority focus of Head Start programs as key educational standards by which performance of grantees would be measured, though the bill would delete social and emotional development as a factor in enhancing school readiness. The new provisions for Title II stated that in order to be eligible a state must have implemented standards for school readiness that include standards for prereading, premathematics and language skills development for PreK that were aligned with state K-12 academic content standards (Schumacher et al, 2003). On May 22, 2003 the Chairman of the Subcommittee on Education Reform, House Committee on Education and the Workforce, filed the first reauthorization of the process. On June 12, 2003 the Subcommittee passed a substitute version of that original legislation and made further amendments. On June 19, 2003 the committee passed the final bill (Schumacher et al, 2003, p.1). The House approved the bill to reauthorize the Head Start preschool program on September 22, 2005 (Davis, 2005).

Launching of the National Reporting System

In the fall of 2003 the largest administration of a single standardized test, the National Reporting System (NRS), was launched. At an estimated cost of \$25 million annually, 450,000 English or Spanish speaking Head Start 4-year-olds, from every state and nearly every local in the nation began to be administered the NRS twice yearly. The stated purpose of the test is three-fold:

1. To enable programs to engage in self-assessment and improvement
2. To target needed training and technical assistance efforts
3. To monitor program's performance in order to determine if public funding should be continued.

The test is a top-down policy initiative that high-level government officials directed the Department of Health and Human Services (HHS) the Administration for Children and Families (ACF) bureaucrats to put into place post haste. In less than 9 months it was developed and piloted on a small number of children and programs. The NRS is a standardized test that relies extensively on multiple-choice items (Meisels, 2006).

The test is composed of five individually administered subtests: two language screeners (to determine if the child is English or Spanish speaking), one receptive vocabulary knowledge subtest, letter-naming skills subtest, and the final subtest that focuses on early math skills. The first language screener, entitled, Simon Says, is based on the game of Simon Says. The examiner

says, “Simon says touch your ear. Simon says point to the door.” A child is supposed to do exactly what Simon says (NRS, 2005-2006, A1, A2). The second language screener, entitled Art Show, requires that the examiner to point to one of two pictures on the page and the child, “What is this?” or “What can you do with this?” The correct answer would be one of two or three correct answers. For example the teacher points to the picture of a pig and asks, “What is this?” The correct answer is “Pig, piggy, or hog.” In another example, the examiner points to a picture of a book and asks, “What can you do with this?” The correct answer would be “read, look at it, or tell stories” (NRS, 2005-2006, B3, B6).

There are two parts to the vocabulary subtest, the listening comprehension, with 24 items, and letter-naming subtest with 26 items. The listening comprehension section is adapted from the *Peabody Picture Vocabulary Test-III* (PPVT-III). Children are shown a page with four pictures: the examiner says a word and asks the child to point to the picture that “best shows what the word means.” For example the child is shown a page with four pictures, these being a stethoscope, a set of binoculars, a microscope, and a set of head phones: the examiner says to the child, “Point to the binoculars” (NRS, 2005-2006, C11). The letter naming section shows a box of 18 letters. There is lower case and upper case of representation of each letter. The examiner circles all of the letters and says. “Point to the letters that you know and tell me the name of each one). If necessary the examiner may say, “Point to the letter and tell me the name out loud” (NRS, 2005-2006, D1). The Early Math Skills subtest consists of 20 early math skills questions. For example there are four geometric shapes on the page, the examiner says to the child, “Look at these shapes. One of them is a square. Point to the square” (NRS, 2005-2006, E9). Another example of the Early Math Skills subtest is a picture of three books. The examiner circles the books and asks the child, “If you gave a child one of these books, how many books would you have left” (NRS, 2005-2006, E14)?

Arguments for NRS

As part of the Head Start Program Performance Measures Initiative, Head Start launched the Family and Child Experiences Survey (FACES), in fall 1997. With a nationally representative sample of 3,200 children and their families in 40 programs, FACES describes the characteristics, experiences, and outcomes for children and families served by Head Start, as well as the relationships among family and program characteristics and outcomes. In fall 2000, Head Start began data collection on a new national cohort of FACES, called FACES 2000. The FACES 2000

sample includes 2,800 children and their families in 43 different Head Start programs across the nation. FACES provides critical information for the Head Start program on important aspects of outcomes, quality, and practices beyond the aggregated, administrative data previously collected. Through the ongoing, longitudinal FACES study, Head Start can examine key facets of program quality and children's school readiness on successive, scientifically representative samples of children, families, teachers, classrooms, and programs. While these data are crucial for decision making at the national level, there are important limitations on the questions they can answer. They do not provide information on every child in each program, nor do they provide information on or comparisons to children recruited but not served by Head Start. These questions are being answered via the Head Start National Reporting System and the Head Start Impact Study, respectively (Zill, Resnick, Kim, O'Donnell, Sorongon, McKey, Pai-Samant, Clark, O'Brien, & D'Elio, 2003, p.1).

Each cohort of FACES consists of a nationally stratified sample of Head Start programs, centers, classrooms, children, and parents. FACES 2000 features four phases of data collection and follows 3- and 4-year-old children from program entry, through one or two years of program experience, with follow-up in the spring of Kindergarten. FACES 2000 employs four main components: the direct child assessment, parent interview, teacher and staff interviews, and classroom observations (Zill et al, 2003). FACES revealed that most children entered Head Start at a great disadvantage, with early literacy and math skills substantially below national averages. Children who entered the program with lower levels of knowledge and skill showed larger gains during the program year, yet still lagged considerably behind national averages. Spanish-speaking children in Head Start showed significant gains in English vocabulary skills without declines in their Spanish vocabulary skills. Head Start graduates showed further progress toward national averages during kindergarten yet Head Start graduates remained behind their more advantaged peers in early achievement. The size of gains that children made while in Head Start were predictive of their achievement levels and behavioral adjustment and performance in early elementary school. Children showed growth in social skills and reduction in hyperactive behavior during the Head Start year. Children with high levels of shy, aggressive, or hyperactive behavior showed significant reductions in these problem behaviors in Head Start (Zill et al, 2003, p.iii).

The great majority of Head Start programs use a curriculum, as mandated by the Head Start Program Performance Standards. A wide variety of curricula are used, with a majority of programs selecting an integrated curriculum such as Creative Curriculum or High/Scope (Zill et al,

2003). Head Start quality has been observed to be consistently good over time, using a variety of indicators including child-adult ratio, teacher-child interactions, and classroom activities and materials. FACES shows that Head Start has a better, more limited range of quality than that seen in child care centers and preschools in several other national studies. More teachers in 2000, studied Early Childhood Education or Child Development for their highest degree, compared with those in 1997-1998. Classrooms with higher levels of quality have teachers with higher levels of education, experience, and positive attitudes and knowledge about early childhood education practice. Teacher education and attitudes are linked to classroom quality (Zill et al, 2003, p. iv).

Higher teacher salaries and use of an integrated curriculum are linked to greater gains in several cognitive and social-emotional areas, including letter identification, oral communication of basic social information, and cooperative classroom behavior, these children also showed greater improvement in hyperactive problem behavior during the Head Start year. Teachers' educational credentials are linked to greater gains in early writing skills. Children taught by Head Start teachers with Bachelors' Degrees or Associates' Degrees showed gains toward national averages in an assessment of early writing skills, whereas children taught by teachers with lesser credentials merely held their own against national norms. Children whose parents report reading to them every day show larger gains in vocabulary knowledge and letter recognition skills than children whose parents report reading once or twice or less frequently per week (Zill et al, 2003, p. v).

Children's scores on FACES assessments at the end of Head Start, as well as the gains they make during the Head Start year, strongly predict their performance at the end of Kindergarten. As an indicator of pre-literacy skills, the cognitive measures show strong associations with reading ability at the end of the kindergarten year. As an indicator of school adjustment and social competence, the behavior ratings demonstrate ability to predict kindergarten behaviors that promote learning and those that impede learning (Zill et al, 2003, p.vi).

As stated within the FACES data, this Family and Child Experiences Survey does not provide information on every child in every program. This is one reason why the NRS is assessing every child. But the purpose of the NRS is to evaluate teacher and program effectiveness. This is contradictory to the FACES data that states that Head Start quality has been consistently good and that some Head Start teachers are even better than other preschool teachers. FACES also reveals that most classrooms are using the standards. So why is the NRS evaluating teacher and program effectiveness when we already know that that Head Start teachers and programs are effective?

Some National Reporting Systems supporters believe that accountability in Head Start needs standardized tests. Ron Haskins (2004) co-director of the Center on Children and Families and Senior Consultant at the Annie E. Casey Foundation, states that data on school readiness for children entering Head Start in 1997 and 2001 showed that children start the program with scores far below average. Haskins notes that Head Start produces an initial boost in children's test scores but these effects fade within a year or two after children enter school. These striking differences continue in school age test scores, high school graduation rates, college attendance, and earnings in the work force. These facts convinced the Bush administration that Head Start needed to be retooled to focus on getting children ready to learn. The plan began with the *Good Start, Grow Smart* initiative, whose plan was to retrain Head Start teachers and bring accountability to the program. Bush was making school readiness, an easily measurable outcome, and the single most important goal of Head Start. The shift in emphasis from comprehensive services to intellectual development has provoked strong opposition from Head Start stakeholders. Haskins says that Bush administration should be commended for taking on a politically difficult issue and sticking to its agenda, despite Head Start's overwhelming popularity. "For too long, Head Start has been merrily rolling along, enjoying ever more increases in funding, without demonstrating its value." Stakeholders take comfort in by spreading false claims of success or in excusing its ineffectiveness by lowering expectations by saying that, "no one should expect a year of preschool to overcome the difficulties of being raised in a troubled family or neighborhood (Haskins, 2004, p.33).

The administration of the NRS poses ethical problems for many teachers who are asked to administer the test. Marla Susman Israel, Ed. D., a member of the advisory group updating NAEYC's *Code of Conduct*, discusses the ethical dilemma of whether or not early childhood educators should administer the NRS test. Numerous educators recounting stories of young children reduced to tears when administered the mandatory NRS test have contacted members of the advisory group responsible for updating NAEYC's Code of Ethical Conduct and Statement of Commitment. In some instances early childhood professionals have refused to administer the test to the upset children. Directors dealing with those refusals have weighed the ethical options of the situation and have responded by (a) removing the upset child from the testing situation; (b) assigning administration of the test to another staff member; (c) writing letters and articles advocating a better assessment system; and/or (d) firing the offending staff member for insubordination. Concerned educators have asked the advisory group this question: Given the

NAEYC's Code of Ethical Conduct and our concerns regarding the NRS child assessment, how should we respond and how have others responded? (Israel, 2004, p.27).”

Israel analyses the values concerning this moral dilemma in order to help guide teachers. To provide objectivity Israel suggests applying the three ethical perspectives that Feeney and Freeman (1999) describe in *Ethics and the Early Childhood Educator: Using the NAEYC Code* to help clarify values. These values are Utilitarianism (providing the greatest good for the greatest number of people); Kantianism (establishing the rightness of the action by determining if it could be universally applied), and; the ethics of care (being respectful of people and relationships (p.30).

Using the framework of utilitarianism, the ethical dilemma under discussion is stated as thus: *Between administering the NRS child assessment and not administering it in order to cause possible harm to children, which alternative provides the greatest good for the greatest number of people?* (Strike, Haller, & Soltis, 1998 as quoted in Israel, 2004, p. 30) Within the mandates of the NRS it is stipulated that opportunities for training and knowledge be provided. Early childhood administrators can insure that a child development expert, a family specialist or the director herself can administer the NRS assessment to the preschoolers. Thus a familiar person can administer the assessment. The administrator can have a fearful child observe a less fearful child as he is assessed so as to relieve anxiety. In this way early childhood professionals can provide a comfortable environment for the majority of children while complying with the federal mandate and ensuring that adequate funding is provided to the greatest number of families and children needing services. Results from the NRS may also be used to provide direction for future professional and curriculum development.

Using the teachings of Immanuel Kant the ethical dilemma can be stated this way: *If early childhood professionals ensure to the best of their ability that children are not harmed, is their work best served by administering the Head Start NRS assessment?* (Beck 1997; Rachels 2003 as quoted in Israel, 2004, p. 30) The policy of mandated assessment in Head Start has also put the national spotlight on the need of quality early education, qualified early childhood educators and the question of what constitutes school readiness. Highlighting quality and strengthening pedagogy has strengthened the field of early childhood. It may benefit the Head Start children who are currently being assessed by the NRS. If research were to establish that the Head Start NRS is not an acceptable form of assessment that same research could provide direction for determining what is appropriate assessment of young children's learning and development so that early childhood teachers could help young children prepare for success in

school. Some educators believe the NRS has done a service for both present and future Head Start participants by increasing the early childhood field's attention to and discussion about what it means to be accountable and to do no harm to children.

Using the final perspective that Feeney and Freeman (1999) suggest, the ethics of care, the ethical dilemma is now stated: *In deciding whether or not to administer the NRS assessment as mandated, and knowing that the test may have serious problems and possible cause harm to children, which alternative is more respectful of people and relationships?* (Kidder 1995; Rachels 2003 as quoted in Israel, 2004, p. 31) Israel states that is more respectful to follow the mandate, administering the test, while continuing to question its validity and reliability. Early childhood educators are responsible for explaining to children, families, colleagues and communities why the mandate must be followed and how through continued research and advocacy policies can be changed to achieve even better results for young children and families. Conversations surrounding the Head Start NRS child assessment can provide real opportunities for families to be involved in policy decisions and advocacy.

Israel concludes that “even though NRS does not use multiple forms of assessment, which is considered to be best practice in understanding children's learning, it may be an opportunity for meaningful discourse, professional development, family education, and advocacy for quality early intervention in programs like Head Start (Israel, 2004, p.32).”

Edward Zigler, PhD and Sally Styfco, 2004, from Yale University's Center in Child Development and Social Policy are confident that if Head Start officials rely on the expertise that already exists in this area and build on the considerable work that has been done already to develop scientifically based protocols, the NRS can become a credible, useful means of helping Head Start providers deliver better services that enable Head Start children to be better prepared for school (Zigler & Styfco, 2004, p.859).

Criticism of the NRS

Several early childhood educators and assessment experts are hoping to stop Head Start's National Reporting System (NRS) or persuade the Bush Administration to make significant changes in its design before it begins its second round of testing 4 and 5-year-old Head Start students. Linda Epinosa, associate professor at the University Missouri-Columbia and member of the technical work group states that numerous assessment experts say standard testing for accountability purposes should not begin before the 3rd grade. Critics of the test say that assessments in preschool

can be unreliable because young children's development can fluctuate from day to day, and that results from the NRS should not be used to make decisions about whether programs are effective (Jacobson, 2004).

In order to ascertain whether or not an assessment tool is developmentally appropriate we need to review two factors in particular; the assessment itself and the needs of the participants. When reviewing the assessment we have to make sure that the test is reliable and valid. Has it been tested on a sample population? We also need to consider the participants; in the case of the NRS this is the preschooler. In order to be developmentally appropriate we must consider the whole child. It is imperative that we consider all developmental domains of the child, cognitive as well as, physical, social and emotional. We need to view the child in terms of their home environment, such as their family and their community. It is impossible to assess a young child's abilities without taking these various factors into consideration.

I will begin with a discussion of NAEYC's concerns about the implementation of the NRS. I will then discuss various psychometric problems with the NRS. I will present several research studies that center on Head Start children. These reviews either focus on the appropriateness of the test or on the specific aspects of young children that need to be considered when evaluating a young child. I will review a study by Hawken (2005) that examined 500 Head Start teachers' views on emerging literacy instruction and how this is reflected in their classrooms. I will review a study by McWayne, Fantuzzo & McDermott (2004) that examines classroom competence and approaches to learning study as an indicator of early academic success. I will discuss the home environment and factors such as relations between family involvement and behavioral and learning competencies for children (Fantuzzo, McWayne, Perry & Childs, 2004), and the relationship between family conflict, community violence and preschoolers' socioemotional functioning (Farver, Xu, Eppe, Fernandez & Schwartz, 2005). I will review a study on preschool quality and the longer-term effects of Head Start (Currie & Thomas, 2000).

NAEYC's Concerns about the NRS

Adele Robinson (2003) Senior Director, Public Policy & Communications at National Association for Education of Young Children discusses NAEYC's multiple concerns about the process, content, and timing for Head Start's new assessment, the National Reporting System (NRS). NAEYC is concerned because a dramatic change in Head Start's approach to assessment was instituted without public discussion or sufficient involvement of experts and stakeholders in early

childhood assessment and education. The assessment was implemented on a time frame that disregarded critical comments from the field, and at best, failed to allow time to determine both the concurrent and predictive validity of all parts of the planned assessment, and to identify the intended and unintended consequences of its use.

Another concern is that the assessment, which was developed for one purpose, is being used for a different purpose. The measures used for the purpose of helping to identify children's learning and development and to help teachers improve instruction and services will also be used for the purpose of program accountability.

NAEYC is worried that the NRS is narrowing the areas of child assessment and that this will carry consequences for programs. The NRS states that the assessment will be limited to three areas: (1) language (2) literacy and (3) numeracy. As noted in the assessment, this is a significant shift from the current requirement to assess children in all 8 domains. By using a 20-minute, limited battery of assessment tasks to determine technical assistance, improve program services, and monitor program effectiveness, the new assessment plan runs the risk of causing programs to dilute other important areas of instruction and services. Research is clear that children's academic development is not separate or apart from other areas of development. Further, research also substantiates that health, nutrition, and other services, including genuine family involvement and support, are particularly important in helping children from very-low income families reach standards for development and learning. By eliminating these areas from the NRS assessment, the proposal disregards evidence of what is needed for high-quality early childhood programs and may lead, in its narrowness, to inadequate attention to children's overall development and readiness for school. Other areas critical to determining program effectiveness would not be captured in this assessment, such as: children's self-regulation and socioemotional competence, creative arts and science, motivation and approaches to learning, problem-solving and other aspects of cognitive development.

The NRS has the potential problem of programs "teaching to the test" and diluting other critical areas of children's development and learning as well as family supports, in an effort to "do well" on the new assessments. The NAEYC suggests that the trainers should guard against assuming a "one size fits all" approach as if every teacher or potential assessor has the same experience and knowledge. The training also must use a variety of approaches to match the learning styles of a varied audience.

The NAEYC states that the NRS gives insufficient attention to the needs of children with

disabilities and children whose native language is not English. The document states that the assessment will be used in English and Spanish. Yet many children in Head Start speak other non-English languages. Will children who speak a language other than English or Spanish be excluded from this field test? The NRS appears to have no discussion of how this assessment and the field test will be used with children with disabilities. At least ten percent of children enrolled in Head Start have disabilities. Will these children, or children with certain kinds of disabilities, be excluded from the field test? To what degree will assessors be trained in how to conduct and interpret the assessment with children with cognitive, physical, or emotional disabilities or children with developmental delays?

The NAEYC has additional concerns about the appropriateness of the subset of items selected for the shortened version of the *Peabody Picture Vocabulary* (PPVT), especially in cultural contexts. Items like “porcupine,” “juggling,” and “camper” suggest a middle-class, European culture. The assessments’ focus on “Progress in Speaking English” is also disturbing given that an important goal with 4-year-olds may be to develop their foundations in the home language, so as to provide a good conceptual basis for second language acquisition.

The NRS is assessing every child, rather than sampling, as a method of conducting a national reporting system. Such sampling is routinely used by NAEP and others and is widely viewed as having methodological advantages. The document appears to assume that only center-based programs are included in the sample, not home-based program option or combination model agencies, nor family childcare homes. If the NRS will be employed by all Head Start grantees in the future, then appropriateness of the assessment methods is an issue for field-testing in home-based environments, as well.

The document states that analysis of the data will take into account certain demographic information. For a fully informed analysis, there should also be collection of data on the educational level of the Head Start parents, the child’s previous enrollment in Head Start, center-based childcare, family childcare, relative care or other setting,

Based on these numerous concerns NAEYC recommends that it would be in the best interest of Head Start children and programs to more fully examine, through a public process, the challenges of this national reporting system and this assessment before engaging in a pilot or full implementation. At a minimum, the instruments being used by the National Reporting System should be provided to the public and reviewed thoroughly. Time is also needed to verify that all instruments, including those adapted from measures used with older children, are valid and useful

with low-income Head Start children of diverse cultures, languages, and disabilities. Second, assessment of individual children in the limited areas of literacy and mathematics should not be used as the primary determinants of Head Start program effectiveness. Instead, the assessment results should be used in conjunction with other evidence to provide additional resources and technical assistance to programs to improve teaching and services. Third, there should be a determination of the full cost of implementing the National Reporting System. The document states that programs will not have direct costs except their “time to participate in the field test.” This is a direct cost to participants in that it diverts time away from other activities. Further, the document fails to account for how this system, including professional development and technology needs, will be paid for under the President’s budget request, which provides only a cost of living increase annually for Head Start. Further, the President’s Head Start proposal makes no mention of the cost to states under his schema for the redesign of the Head Start program (Robinson, 2003, p.7).

Psychometric problems

Samuel Meisels, Ed.D, president of The Erikson Institute lists several psychometric problems with the NRS. These include:

1. *External validity of the subtests*: There is no evidence to demonstrate that these subtests measure what they purport to measure (Meisels, 2004).
2. *Construct underrepresentation*: Both the math and literacy items fail to capture important aspects of the construct that the test is intended to measure. In the letter naming subtest the child does not receive credit for merely producing the sound that a letter represents if he does not name the letter. Creating the sound a letter makes is an indicator of phonemic awareness, the cornerstone of oral language development. In the math subtest there are no items that assess comprehension of pattern, number constancy, matching, classification or estimation and spatial reasoning or recognition of more or less all of which are central to the development of a young child’s mathematical thinking (Meisels & Atkins-Burnett, 2004).
3. *Construct-irrelevant variance*: The test scores are very likely influenced by factors that are irrelevant to the constructs the test is intended to measure, such as choice of vocabulary, selection of illustrations, language burden of math items, or how the items appear on the page. For the receptive vocabulary knowledge subtest the preschoolers are asked to identify appropriate items such as body parts, animals and actions and some less appropriate, class-biased items. For

example, the child is asked to choose the picture that looks like a vase yet every item that is shown could be used as a vase. In another example the children are asked to choose the facial expression for horrified. All four facial expressions are Caucasian ignoring the fact that facial expressions differ in different cultural and ethnic groups. The expression for horrified is often confused with anger or rage and it is not recognized by most children until later in life (Meisels & Atkins-Burnett, 2004).

The letter-naming task is totally removed from children's natural use and experience with letters. Except on a school bulletin board children rarely see an alphabet with both upper and lower case letters displayed in pairs. No appreciation of context is recognized on this subtest. The child receives credit only for naming the letter not for providing the sound the letter represents (Meisels & Atkins-Burnett, 2004). The letter name task is misconceived and reflects a lack of understanding about what rapid letter naming teaches us about young children's skills in early literacy (Meisels, 2006).

The math items assume that Head Start 4-year-olds can attribute causality, do subtraction, use standard metric units, and understand the subjunctive case (Meisels, 2006). The math skills subtest focuses more on language competence than quantitative skills. Only 25% of this subtest involves naming a number or pointing to a specific shape. The items are poorly designed, for example, a child is asked to count grapes some which are in groups the others are individual. Another example is that the test asks which crayon from an array of four crayons is longer than the brush, but the brush that is shown is shorter than all the crayons. It requires the child to know that the word brush refers to the handle and brush as a single item. These two examples can both be very confusing to a young child. Another item on the math subtest shows several coins and asks, "Which coin is smaller in size than the penny?" This question is difficult because it requires that the child make a distinction between the physical size of a coin and its monetary value, be familiar with five different coins (one of which is a relatively rare 50 cent piece), and know the meaning of "which." The test contains several if-then statements and comparative terms. An example of an if-then construction is "If you give a friend one of these books, how many would you have left?" This question is particularly language driven and is developmentally difficult for a child to understand before elementary school age (Meisels & Atkins-Burnett, 2004).

4. *Lack of sampling strategies*: There is no justification for testing every child in every program (except for non-English and non-Spanish speaking children) when a sampling strategy could save time and money or prevent teaching to the test (Meisels, 2004, p. 1401). When the NRS was

announced many in the field suggested that only a small sample of Head Start children be sampled. Yet the Department of Health and Human Services (HHS) administrators wanted to test the population, not the sample. Without testing every child in every program the HHS felt that it would be impossible to know whether the children were learning or if public funds were being used wisely (Meisels, 2006). Only a small sample of Head Start children should be tested as is done in other national assessments such as the National Assessment for Educational Progress. Instead every child who is the right age and speaks either English or Spanish is tested (Meisels, 2004).

Meisels (2004) discusses two other issues that depict more serious problems with the NRS. The first is that the model of pedagogy that is implicit in this test is a model of passive reception. Rather than recognizing learning as active and teaching as a joint process between teacher and child the child is treated like an empty vessel that is being filled with knowledge and skills that are needed for competence. Some teachers will alter their teaching to conform to the pedagogical model implicit in this test. When a teacher knows that that the results of a test will be used to make decisions that may effect the programs continuation (as is such with the NRS) they are tempted to begin teaching to the test.

The second issue includes the overall rational for the NRS. Policy makers in Washington have long recognized that poor children, particularly those in Head Start do not start school with skills equivalent to more affluent children. Yet just because children are poor does not mean that they are all the same. Moreover development is not linear. In the first 5 to 8 years of life change is more the rule than the exception. This is one of the reasons that so little of the variance in outcomes at first or second grade is accounted for by preschool tests. The variance is ~25% for cognition and only 10% for socioemotional predictions (LaParo & Pianta, 2000). To believe that a test of this kind can tell us enough to improve programs and enhance children's learning is to assume a homogeneity in children that is entirely unjustified (Meisels, 2004, p. 1401).

The vocabulary subtest for the National Reporting System is adapted from the *Peabody Picture Vocabulary Test-III* (PPVT-III), a test of listening receptive language skills such as listening comprehension. There is has much controversy about the developmental appropriateness and cultural bias portrayed in this test. I will first discuss The NAEYC's concerns about the PPVT-III then I will review a study about concerns about African American children who have been administered this test.

The Peabody Picture Vocabulary Test

Adele Robinson, from the National Association for Education of Young Children, is concerned about the cultural appropriateness of the subset of items selected for the shortened version of the *Peabody Picture Vocabulary* (PPVT). Items like “porcupine,” “juggling,” and “camper” suggest a middle-class, European culture. The likelihood that the nation’s poorest urban and rural children would have familiarity with porcupines, juggling and many of the other items is slim; failure to recognize these words may say less about preschoolers’ language ability than about opportunities to learn this culturally- and economically-laden information (Robinson, 2003).

Champion, Hyter, McCabe, & Bland-Stewart (2003) assessed the receptive vocabulary skills of typically developing African American preschoolers using the Peabody Picture Vocabulary Test-III (PPVT-III). The participants were 49 impoverished (annual incomes less than 9,000) Head Start children (24 girls, 25 boys) between 3.2 and 5.9 years old. The participants were enrolled in 3 different Head Start extended-day programs in the Tampa Bay area. All of the children were typically developing 3-, 4-, 5-year olds. The children were tested in the summer of 2001. An African American female speech-language pathologist using the published guidelines as the instrument administered form A of the PPVT-III. Two African American female speech-language pathologists supervised her. Each participant’s protocol was scored according to published scoring guidelines. Raw scores were converted to percentiles and standard scores. Interrater agreement was established by having a certified speech-language pathologist rescore a randomly selected subset of raw-scores, standard scores and percentile ranks of 20% of the participants. Scoring agreement was 100% between raters. The 49 children in the sample reported a mean of 86.4 (SD of 10.96) on the PPVT-III, which was significantly different than the normative sample for that test (mean =100, SD = 15), despite the fact that that the test’s normative sample included minority children. There was also no significant correlation between the age of the preschooler and the PPVT-III standard, which is what one would expect, since the standard scores are age-normed. These children had disproportionately lower scores (range 57-71) with three children scoring below 70 (2 standard deviations below the mean) while 11 scored 78 or below (1.5 standard deviations below the mean). Twenty students (41%) scored more than one standard deviation below the mean. 80% of the sample clustered within one standard deviation of its own mean, which indicates that the sample itself was not normally distributed. An item analysis revealed that most children systematically missed few items. Instead performance seemed

reflective of socioeconomic and/or patterns of vocabulary usage. Children's culture plays a role in their vocabulary. African American adults were asked about the 75 words missed by the African American children, the adults had strong alternative responses for 11 of the words. Which implies that the children might also have strong alternative meanings for certain vocabulary words presented in the PPVT-III.

The results of this test indicate that the PPVT-III may not be a culturally appropriate measure of African American preschoolers language acquisition. Since these children have unreasonably lower scores than the average preschooler one would need to consider the reliability of assessing African American children using this assessment tool especially since the parents of these children also have different meanings for certain words that are used in the test. If children are assuming different meanings to words at home than at school their definitions for those words would be that of their home culture not the school culture. Since the correct answers on the PPVT-III are indicative of the school culture some of the answers that the children supply for the language subtest of the NRS would be incorrect.

Language/Literacy

Hawken, Johnston & McDonnell (2005) discuss Head Start's emphasis on the importance of children acquiring emerging literacy skills in preschool. The reauthorization of the Head Start Act (1998) stressed the need for highly qualified teachers who could implement scientifically based emergent literacy skill instruction. Head Start developed the *Head Start Child Outcomes Framework* to provide guidelines for teachers to use in assessing critical skills for school readiness, with a focus on emergent literacy skills. The emphasis on emerging literacy was further signified by the formation of Head Start's National Reporting System (NRS) in 2003. This system requires Head Start centers to assess all children on key skills outlined in the *Outcomes Framework* and to report the results to the national foundation. Although several of the areas outlined are considered optional for the NRS, three of the four legislatively mandated domains of assessment focus on emerging literacy skills (p.232).

Upon entering kindergarten Head Start children are significantly below national averages on most measures of academic readiness, including emerging literacy skills. Studies have shown that teacher views/beliefs shape classroom practice related to implementing literacy instruction. Hawken et al (2005) conducted a study to examine Head Start teachers' views on emerging literacy instruction and the practices and strategies currently being implemented in Head Start

classrooms to promote emerging literacy. Emerging literacy skills as defined in the *Outcomes Framework* were used to structure the investigation. The *Head Start Child Outcomes Framework* divides emerging literacy into the following skills: book knowledge/appreciation, print awareness, phonological awareness, alphabet knowledge and early writing (Head Start, 2000). 500 surveys were mailed to a stratified, random sample of Head Start preschool teachers. The 10-page survey was based on a review of the emerging literacy research, policy, and practice literature and included skills outlined in the *Head Start Child Outcomes Framework*. The survey consisted of 24 items, many with subparts, and used a variety of question formats. The items were divided into four major sections: (1) demographic information, (2) needs related to emergent literacy, (3) frequency and types of evidence-based strategies used to promote emergent literacy, and (4) perceptions regarding literacy and language instruction (p. 234). An initial mailing and two follow up mailings yielded 273 completed surveys. The emergent literacy strategies presented in the survey were arranged according to the literacy domain elements from the *Head Start Child Outcomes Framework*, these being: book knowledge/appreciation, print awareness, phonological awareness, alphabet knowledge and early writing.

The survey results summarized the responses to these five domains. The most frequently used strategy Head Start teachers reported to improve book knowledge was to have the children practice holding books and turning pages correctly (89%) on a daily basis. Having children predict stories on a daily basis was another frequently implemented strategy (67%). Providing opportunities for children to retell stories and taking dictation from children was used at least once or twice a week by more than 75% of the teachers surveyed. Rereading stories to children was implemented one or two times per month (59.7%). The most frequently used strategy by the teachers that occurred on a daily basis to improve print awareness and concepts were encouraging children to use literacy related props during dramatic play (83.2 %). Teaching children that reading occurs from left to right, pointing to print while reading aloud, using a written schedule, and displaying children's writing around the room was implemented daily in more than 70% of the classrooms.

The most frequently used strategy to promote phonological awareness was having children practice identifying initial sounds in words (e.g. "f" in fish) daily or one or two times a week (80.3%). More than 70% of the teachers reported using nursery rhymes and practicing rhythm games either daily or one or two times per week.

The most frequently used strategy to improve alphabet knowledge was encouraging play with alphabet puzzles and magnetic letters on a daily basis (81.3%). Strategies such as reading alphabet books, introducing new letters, playing games to teach letter-word recognition, using flannel boards with letters/words, and practicing sounds during read aloud time were used either daily or at least once or twice a week by more than 70% of the teachers surveyed. The most frequently used strategy to improve early writing skills was presenting children with opportunities to use a variety of writing tools and have children practice writing their names. More than 97% of the teachers reported providing opportunities to use a variety of writing tools at least once or twice a week and 76.2 % of the teachers had children practice writing their names on a daily basis, whereas 20% implemented this strategy at least once or twice a week (Hawken et al, 2005).

This survey, based on the *Head Start Child Outcomes Framework*, demonstrates that these Head Start teachers are employing the emergency literacy strategies that are suggested in Framework. Yet the language based subtests in the NRS, the “Receptive Vocabulary Knowledge” and “Letter Naming Skills” subtest do not measure much of what these teachers are teaching. When a child sits and answers a multiple-choice question like those in the NRS we can’t ascertain if she/he knows how to hold a book. The preschoolers do not have the opportunity to predict the ending of a story, nor retell a story through multiple-choice questions. Yet teachers know that the children understand and are thus developing emerging literacy skills when they predict an ending or retell a story.

More than 80% of the surveyed teachers promote phonological awareness at least twice a week by having the children make the initial sounds of words. Yet in the NRS the children receive no credit if they say the sound of the letter instead of the name of the letter. A multiple-choice question removes a letter from its context. If a child looked at the alphabet in chronological order or read the letters in a known word they would be able to express the name of the letter in relationship to its symbol. An isolated letter may be harder to recognize, as is such with the NRS.

97% of the teachers supply the children with various writing tools to encourage them to write yet when preschoolers are administered the NRS they sit and point to images instead of holding a writing implement and demonstrating how they write. If the preschoolers were given the opportunity to write and then repeat the letters out loud it would express whether or not they know specific letters. The NRS does not allow them to do this.

How can we ascertain if preschoolers are beginning to develop the skills for reading and writing if they are not supplied with the tools in which to express themselves? The evaluator

could read the child a story and ask him or her to retell the story or give the child a pencil and ask the child to write his/her name. This would tell us a great deal more about a child's emerging literacy skills than pointing to a letter or picture in the NRS assessment booklet.

The Head Start teachers in his study were employing different strategies in which to incorporate Head Start's *Child Outcomes Framework*. Though the study does not tell us how well the preschoolers acquired language and literacy skills using these strategies it does tell us that the recommended strategies are being employed. The teachers are engaging the children in language and literacy development through meaningful interactions with books and spoken language. Through everyday experiences the teachers are able to listen to the children and ascertain whether or not the children are conceptualized their new language and literacy skills.

Classroom competence and approaches to learning

McWayne, Fantuzzo & McDermott (2004) combined both developmental and ecological consideration as they examined the unique contribution of multiple preschool classroom competencies to an indicator of early academic success. Dimensional analyses were used to analyze classroom competency for low-income preschoolers and to determine the contribution of these classroom competencies to children's early school success while controlling for preschool classroom quality and neighborhood characteristics. Typological analyses were then used to determine the nature of distinctive classroom competency profiles among the children. 195 Head Start children participated in the study. Children ranged in age from 55 months to 69 months (mean = 62.52 months) and all were expected to enter kindergarten the following fall. 53% of the children were boys, 81% were African-American, 12% Caucasian, 7% Latino or Asian. All of the children were English speaking. Participants were recruited from a representative set of 32 classrooms in 17 centers across a large, urban school district Head Start program. Three girls and three boys from each classroom who were expected to proceed to kindergarten the next year were selected.

Participant classrooms were selected randomly from the six clusters that compose the school district. The school districts Head Start program serves 4,628 children in 238 classrooms across 78 centers. Demographic data indicate that the program serves a predominantly Africa American population (76%), with annual income below \$15,000 for 83% of the families. The measures employed were the Child Observation Record (COR), Penn Interactive Peer Play (PIPPS), Preschool Learning Behaviors Scale (PLBS), Early Screening Inventory Revised-

Kindergarten Version (ESI-K), Classroom Quality, and Neighborhood Context. Individually administered direct assessments were conducted with participant children using the ESI-K at the end of their Head Start experience (during May and June). Data were linked from program-wide teacher performance assessments. Data from COR are routinely collected for Head Start evaluation purposes during the months of fall and spring. The PIPPS-TA and PLBS measures were distributed to teachers and teacher assistants in late spring, concurrent with the COR observation period and the ESI-K administration.

In order to take into account ecological influences on Head Start children's development two additional forms of data were obtained: classroom quality information and neighborhood characteristics. Two different empirical approaches identified the salience of classroom competency constructs for this sample – dimensional (variable centered) and typological (person-centered). Three distinct dimensions emerged: General classroom competence (including emergent cognitive literacy and numeracy, motor and social skills), specific approaches to learning (task orientation behaviors related to persistence, attention, motivation, and responses to instructional situations) and to learning and interpersonal classroom behavior problems (reflecting children's problematic interactions with peers during play and difficulty with interpersonal aspects of the classroom learning process). These empirically derived dimensions were analyzed using hierarchical set wise regression to determine their unique relations to children's early academic success. General classroom competence and specific approaches to learning were found to be uniquely associated with academic success (McWayne et al, 2004, p. 633).

General classroom competence and specific approaches to learning included cognitive, social engagement, movement and coordination, competence motivation, attention/persistence, attitude toward learning and play interaction. While these factors are uniquely associated with academic success the NRS only analyzes the cognitive factor. According to this study we need to analyze social interactions and physical characteristics. We need to recognize how a child feels about him/herself and how attentive or persistent that child is. We need to understand the child's attitude toward learning. Knowledge in and of itself is not an appropriate indicator of future academic success. Knowledge needs to be assessed in context of everyday situations.

Home Environment

Fantuzzo, McWayne, Perry & Childs (2004) conducted research on the relations between multiple dimensions of family involvement and their relations to behavioral and learning competencies for

urban, low-income children. Parental report of family involvement was gathered in the fall using a multidimensional assessment. Relations between family involvement dimensions and end of the year outcomes of approaches to learning, conduct problems, and receptive vocabulary were investigated. Participants in the study were 144 preschoolers enrolled in central city Head Start programs in large urban cities in the Northeast. Children ranged in age from 46 to 68 months and 46% of the children were male. 96% of the participant families were African American. Parent participants ranged in age from 18 years to 74 years (mean = 33, SD =11.5). 73% were mothers, 8% were fathers, 19% were other relatives or foster parents. Reports show an average of three children per household. The sample consisted of only English speaking caregivers. The demographic composition of the participating Head Start centers matches national proportions for urban Head Start programs with 90% of the families below \$12,000 and most families (64%) below \$9,000.

Family members involvement in children's education was measured using the Family Involvement Questionnaire (ask family member the nature and extent to which they are involved with their children). Approaches to learning were assessed using the Preschool Learning Behaviors Scale (a multi-dimensional teacher's report of children's behaviors). Classroom problem behaviors were evaluated using the Conner's Teacher Rating Scale-28 (utilizes three subscales [Conduct Problems subscale, Hyperactivity subscale, Inattention-Passivity subscale] to categorize patterns of troublesome child behavior during classroom activities). Receptive vocabulary skills were assessed using the Peabody Picture Vocabulary Test – III. Results indicated that for this study parent involvement dimensions were differentially associated with student learning and classroom behavioral adjustment outcomes. Home-based involvement activities (reading at home, providing a place for educational activities, and asking a child about school) evidenced the strongest relationships to later preschool classroom competencies. These activities were related to children's approaches to learning, especially motivation and attention/persistence, and were found to relate positively to receptive vocabulary. Higher levels of home-based involvement were associated with significantly lower levels of classroom behavior problems. These study indicated the importance of including family involvement in the preschool classroom to elevate the child's academic achievement and lower potential and behavioral problems (Fantuzzo et al, 2004).

This study indicates that parent involvement could lead to preschool classroom success yet parental relationships are not considered in the NRS. If parent involvement can be a contributor to

preschool success then the NRS should consider the role of the family within its assessment tool instead of focusing only on academics.

Farver, Xu, Eppe, Fernandez & Schwartz (2005) examine the relation among family conflict, community violence and preschoolers' socioemotional functioning and explored how children's social cognition and mother's physiological functioning may mediate the outcomes associated with this exposure. Mothers of 431 Head Start preschoolers were completed questionnaires about their family demography, exposure to community violence, family conflict and children's distress symptoms. Children were administered a social cognition assessment and teachers rated their behavior. Results show that mother's reports of children's co-witnessing of community violence were positively associated with police department crime rates, children's distress symptoms, and teacher's ratings of aggression. A path analysis revealed that children's social awareness and mother's depressive symptoms partially mediated the effects of community violence and family conflict on outcomes for children (p. 160).

Preschoolers are administered the NRS at specific times twice a year. Environmental factors can play a deciding role on how a child may feel on the day of the test. If a child has witnessed some type of abuse or the child's caregiver has been the recipient of abuse the child may have difficulty identifying the correct response for a multiple-choice question. A distressed child will not be as alert and receptive as a child who is not distressed. A child's home and community environment must be considered when assessing a child. The NRS does not take the child home environment into consideration.

Preschool quality

Currie & Thomas (2000) reviewed school quality and the longer-term effects of Head Start. They found that the effects from test scores for Head Start black children fade out more quickly than the effects from test scores for Head Start white children. Currie & Thomas state that Head Start black children go on to attend schools of worse quality than other black children. There is not a similar pattern among white children. In relation to school quality gaps in test scores are very similar between white children and black children. They surmise that effects of Head Start may fade out more rapidly among black students at least partially because black Head Start children are more likely to subsequently attend inferior schools.

According to this study the effects of Head Start fade out more quickly for Head Start black children than Head Start white children because Head Start black children go on to inferior

schools. Perhaps a study should be conducted on the effects of Head Start for both white and black children who attend the same elementary school. Only then we could see if the quality of the school affects the knowledge that the child has gained through Head Start.

The NRS is not officially administering the assessment to evaluate preschoolers' learning. It is being utilized to evaluate teacher and program effectiveness. Yet it is the preschoolers' scores that are being used to evaluate both teachers and programs effectiveness. If the test is administered for this reason it should be an accurate measure of preschooler's learning. Yet these studies have just illustrated that there is much more to consider when testing a child than just the test score itself. We need to consider all aspects of children's development, the child's home environment, and quality of the preschool before we can measure their knowledge.

Why does Head Start allow its preschoolers to be evaluated using such a limited summative assessment tool? There is a wide variety of alternative forms of formative, authentic assessment available that are currently being administered in various early childhood programs, including Head Start. These alternative forms of assessment are meaningful and take place within context of the classroom. In the following section I will discuss developmentally appropriate formative assessment alternatives that could be used instead of the NRS.

INFORMAL ASSESSMENT

Informal assessments, which are sometimes referred to as authentic, alternative or performance based assessment engage or evaluate children on ongoing tasks that are individually meaningful, take place in real life contexts and are grounded in naturally occurring instructional activities (Epstein et al, 2004). Informal assessment can effortlessly be incorporated into classroom routines and learning activities. These unobtrusive techniques can be used at anytime without interfering with instructional time. Methods for informal assessment can be divided into two main types: unstructured (such as student work samples, journals, anecdotal records, homework) and structured (e.g., checklists, observations, and rating scales). Examples of unstructured informal assessment include samples of student work, journals, anecdotal records, and homework. Examples of structured informal assessment include checklists, direct observations, and rating scales (Navarete, Wilde, Nelson, Martínez & Hargett, 1990; Gullo, 2005)

Advantages in Using Informal Assessment

There are several advantages of using informal assessments. One advantage is that the assessment is derived directly from the curriculum and the teacher's instructional objectives or from commercial published criteria. The teachers are able to choose and assess processes, skills and knowledge. Another advantage is that the assessment maintains the integrity of a constructivist approach to teaching and learning. It assesses the process of what children learn and how they use the knowledge and skills that they have acquired within the context of activities embedded in the curriculum. The third advantage is that if designed and used properly informal assessments can be correlated with diagnostic needs. Finally, informal assessments reflect a flexible approach to assessing children. The assessment can be used to determine mastery as well as the level that the child is performing at on his way to mastery. Thus informal assessment can yield information that the teacher can use to design and implement curriculum activities that lead to individual mastery (Gullo, 2005. p. 84).

Disadvantages in Using Informal Assessment

There are also disadvantages of using informal assessments. One disadvantage is the improper development and implementation of the assessment measure. There is a lack of reliability and validity with informal assessment. Since it is used teachers within school districts interrelated reliability can be established. A number of teachers can use the same instrument to determine whether or not they get the same results. Another disadvantage is the misuse of the information that is gathered with the assessment, if the assessment is not measuring what it purports to measure then it is being misused. Finally teachers are not adequately prepared to develop informal assessments or use the information effectively for curriculum development (Gullo, 2005).

Informal Assessment in Early Childhood Programs

Informal ongoing assessment supports children's learning and leads to meaningful curriculum. It is the practice of gathering information in the context of everyday class activities and using this to obtain a descriptive picture of children's abilities and progress. Observing and documenting what children do and say, collecting samples of children's work over time, talking with the children and exchanging information with parents are examples of ongoing assessment.

To be successful with ongoing, systematic assessment teachers need to facilitate comprehensive curriculum, that which is grounded in research and child development theory.

Teachers must be knowledgeable about their curriculum and know how to build a comfortable learning environment. After purposefully observing children and gathering rich data teachers can set goals and objectives for the children addressing all areas of development (cognitive, language, social/emotional and physical). These goals should reflect content standards for different disciplines, including literacy, math, social studies, and address the outcome requirements and state expectations. After determining a child's progress in relation to an objective teachers can make decisions about next steps (Dodge, Heroman, Charles & Maiorca, 2004).

Informal Assessment Techniques

There are many different types of informal assessment that are utilized in early childhood programs. Some of the most commonly used approaches are checklists, rating scales, samples of children's work, direct observation, anecdotal records, running records, time and event sampling, the project approach, and portfolios (Gullo, 2005). I will briefly discuss a few of these.

Checklists and Rating Scales

A checklist is a list of chronological skills or behaviors arranged into categories and used to determine whether the child exhibits the skills or behaviors listed. Teachers can quickly observe the children in groups and check to see the behavior or skill that the child is demonstrating at the moment (Ratcliff, 2001/2002).

Rating scales are used to describe the degree to which behaviors or traits are alleged to be present in the individual. Rating scales are often used to measure traits not easily expressed using other assessment procedures. For example, on report cards children's conduct, motivation and effort are often described using a type of rating scale (Gullo, 2005).

Direct Observation

Direct observation occurs in the context of the daily classroom routine while the children are involved in curricular activities. The teacher actively participates as she closely her students' interactions with classmates, classroom materials, and the environment. There are different types of direct observations these include anecdotal records, time and event sampling (Gullo, 2005). Anecdotal Records are written descriptions that provide a brief objective account of an event or incident. Only the facts are recorded, what happened and when and where it occurred (Ratcliff, 2001/2002). Running Records are oral-reading tests that are used to assess a child's reading skills. A child reads an age appropriate story while the teacher reads a photocopy of that same story. The

teacher listens to evaluate the quality of the child's reading, noting mistakes, number of words read, self-corrections, words omitted, words added and words reversed. Notes are made on comprehension, fluency, and expressiveness in reading and the nature of the child's mistakes (Kostelnik et al, 2004, p. 194). Time sampling occurs when there is an interest in determining the frequency of a behavior. Children are observed for a predetermined period during which the specified behavior is recorded every time it occurs. Event sampling occurs when there is an interest in determining the frequency of an event. Children are observed for a predetermined period during which the specified event is recorded every time it occurs (Gullo, 2005).

The Project Approach

The project approach is in-depth investigation of a topic worth learning more about. The investigation is usually undertaken by a small group of children within a class, the whole class, or by an individual child. The key feature of a project is that it is a research effort deliberately focused on finding answers to questions about a topic posed either by the children or the teacher. The goal of a project is to learn more about the topic rather than to seek right answers to questions posed by the teacher (Katz, 1994; Helm, 2003).

Projects can be described as having three distinct phases: Phase I: Planning and Getting Started, in which a common ground among the participants is established by pooling the information, ideas and experiences the children already have about the topic. A topic, either initiated by the teacher or emerging from the child's interest, is selected, and questions are formulated to answer during the investigation. Phase II: Projects in Progress, in which new information is introduced. This is done by visiting field sites, talking to visitors and other experts, by examining artifacts, and conducting experiments. Phase III: Reflections and Conclusions, in which the project is brought to completion through either group or individual work and learning is summarized. During this phase a culminating event or activity is both planned and completed followed by a review of the project and assessment of achievement of goals (Katz & Chard, 1997; Helm & Katz, 2001).

Portfolio Assessment

Portfolio assessment is a collection of information about an individual child that is drawn from a variety of sources. A portfolio contains an accumulation of materials, compiled by the teacher and the students, representing the child's efforts, achievement and progress. Portfolios provide a

comprehensive view of individual children over time within the context of the school program. Portfolios help teachers integrate assessment and instruction and they help children learn to evaluate their own work (Appl, 2000). Portfolio assessment is a purposeful, multifaceted process of collecting documentation of children's growth, progress and effort over time. There are several criteria assessment portfolios must meet. The portfolio must be clearly linked with instructional objectives; must be an ongoing assessment system that allows teachers to observe the continuous, dynamic movement of children's growth; avoid becoming a teacher-manufactured document; be performance based; emphasize purposeful learning; be ongoing in all cultural contexts of school, home and community; and celebrate, support, and encourage a child's development and learning (Hanson & Gilkerson, 1999, p. 81).

EEC Approved Assessment Tools

The Commonwealth of Massachusetts' Department of Early Education and Care (Department of Early Education and Care, 2007) has approved four assessment tools that are appropriate for evaluating young children. I have selected these assessment tools in particular because these are the assessment tools that a program must be utilizing in order to be eligible for a Universal Prekindergarten (UPK) Pilot Assessment Planning Grant. I believe that if the EEC has listed these tools as criteria for the UPK grant they must be highly reflective of the EEC preferred assessment tools for gauging young children's understanding. These EEC approved assessment tools are Work Sampling, High Scope Child Observation Record, Creative Curriculum Developmental Continuum and Ages and Stages.

Work Sampling System

The Work Sampling Observational Assessment System is a validated, research-based observational assessment designed to enhance instruction and improve learning. It is a low stakes, nonstigmatizing system, which relies on extensive sampling of children's academic, personal, and social progress over the school year. It provides a rich source of information about students' strengths and weaknesses. Through checklists, samples of children's work and other structures teachers can systematically assess student's progress in seven curricular areas (Meisels, 1996/1997; EEC, 2007).

The Work Sampling System (WSS) consists of three complementary components: developmental guidelines and checklists, portfolios, and summary reports. The developmental

guidelines and checklists cover seven curriculum areas, each of which are broken down into specific areas within that domain. For preschoolers the domains are distributed into the following categories. Personal and social development includes self-concept, self-control, approaches to learning, interactions with others and social problem solving. Language and literacy includes listening, speaking, reading and writing. Mathematical thinking includes mathematical processes, number and operations, patterns, relationships and functions, geometry and spatial relations, and measurement. Scientific thinking focuses on inquiry. Social studies includes people, past and present, human interdependence, citizenship and government and people and where they live. The arts include expression and representation and understanding and appreciation. Physical development and health includes gross motor development, fine motor development, and personal health and safety. These guidelines and checklists give teachers a set of observational criteria that are based on state and national standards. Skills, behaviors, and academic expectations are presented in the form of one-sentence performance indicators. Each of the developmental guidelines explains the meaning and the significance of the performance indicator by providing a rationale, reasonable expectations, and classroom examples for children of that age and grade. (Meisels, Jablon, Marsden, Dictelmiller and Dorfman, 2005; Meisels, 2006).

The WSS advocates a structured approach to portfolio collection that relies on the collection of two types of work: core items (representations of a particular area of learning within a domain that are selected three times a year); and individualized items (unique examples of a child's work that capture the child's passions and achievements and reflect integrated learning across domains). Collecting portfolio items on multiple occasions allows the Portfolio to become a tool for documenting, analyzing, and summarizing the child's growth and development through the entire school year (Meisels, 1995; Meisels et al., 2005; Meisels, 2006).

The final component of the WSS is the summary report, completed three times a year for each child. Teachers evaluate children's performance and progress by using information from the developmental checklists and student work collected in the portfolios. Any deviation from growth is explained by the teacher in the general comments section. The summary report provides information about the current level of the child's performance and about the child's growth over time (Meisels et al., 2005).

High Scope Child Observation Record

The High/Scope Child Observation Record (COR) is an observational assessment tool that charts children's development and progress over time. Teachers collect information about each child throughout the school year and in many different learning areas (High Scope, 2005). COR is widely used in Head Start programs. Head Start teachers who complete it several times a year can assess how well their program contributes to children's development (Schweinhart & Storer, 2001). The COR rates a child's behaviors and activities in six developmental curriculum areas. The COR assessment booklet is divided into six domains which in turn are broken down into specific areas within that domain. The six curriculum areas and their subsequent items are: Initiative, which includes, expressing choices, solving problems, engaging in complex play and cooperating in program routines. Social relations, which consists of, relating to adults, relating to other children, making friends with other children, engaging in social problem solving and understanding and expressing feelings. Creative representation includes making and building, drawing and painting and pretending. Music and Movement includes exhibiting body coordination, exhibiting manual coordination, imitating movement to a steady beat and following music and movement directions. Language and literacy consists of understanding speech, speaking, showing interest in reading activities, demonstrating knowledge about books, beginning reading and beginning writing. Logic and mathematics includes sorting, using the words *not*, *some* and *all*, arranging materials in graduated order, using comparison words, comparing numbers of objects, counting objects, describing spatial relations and describing sequence and time (High/Scope Educational Research Foundation, 1992).

Five statements describing the child's level of behavior, numbered 1 to 5, are listed under each item. Based upon observations of the child, the observer chooses the statement under each item that best represents the highest level of behavior characteristics of the child. A COR Ratings booklet can be used in conjunction with the COR. The Rating Booklet provides general descriptions of each behavior and offers examples of the typical responses for each of the items within the six curriculum areas. In most cases the behaviors are ones that the child initiates unless they specifically relate to the child's ability to follow directions or otherwise respond to adults. These observations are recorded three times during the year. On the last page of the COR is a ratings and summary sheet. Each of the developmental areas and their subcategories are listed with three time slots for each item. The staff member tallies the scores, which were based on the numbered 1 through 5 level of behavior. Completing a summary sheet for each child is helpful for

processing data for groups of children or classrooms (High/Scope Educational Research Foundation, 1992).

Information for the COR ratings comes from the daily, ongoing, anecdotal notes kept by the program staff. High Scope offers anecdotal note cards that are divided into the six developmental areas. High/Scope also offers High/Scope Key Experience Note Forms. On these forms the COR category of logic and mathematics is broken into six subcategories (classification, seriation, number, space and time) that correspond to High/Scope key experience categories. Other useful sources of information are samples of children's work that may be kept in an individual portfolio (High/Scope Educational Research Foundation, 1992).

High /Scope offers a Parent Report Form that is designed to aid in the preparation of a periodic report for parents based on COR. The report is divided into the six developmental categories. Space is provided for program staff to insert anecdotal notes on behaviors that are characteristic of the child's experiences or abilities in each of the domains (High/Scope Educational Research Foundation, 1992).

Creative Curriculum Developmental Continuum

The Creative Curriculum Developmental Continuum is an assessment instrument that uses multiple methods of data collection to help guide teachers in observing what preschool children can do and how they do it over the course of the year. The teacher observes the child's learning in relation to the goals set by the Creative Curriculum framework. This recorded information is then used to rate children's development on various indicators (EEC, 2007).

The teacher uses the Creative Curriculum Developmental Continuum to integrate the curriculum and assessment. The teacher begins by devises a systematic way to observe, document and organize her notes on each child using the Creative Curriculum Developmental Continuum as a guide. She also creates a portfolio for each child. The continuum, which lists 50 goals and objectives for the early childhood curriculum, provides the teacher with direction for planning the curriculum and provides a framework for what each child knows and how each child is developing. Since early childhood focuses on the whole child all areas of development are covered in the continuum. The development domains are then divided into specific characteristics, which are again divided into items within these subcategories. The developmental domains and their subcategories are social/emotional development, which includes sense of self, responsibility for self and others, and prosocial behavior; physical development, both gross motor and fine motor;

cognitive development, which includes learning and problem solving, logical thinking and representation and symbolic thinking, and; language development, which includes listening and speaking and reading and writing (Dodge, Colker, & Heroman, 2001).

Once a system has been established the teacher begins collecting facts about the children by observing and documenting the children's learning. She composes short informal anecdotal notes, makes formal observations, documents these observations, and collects samples of children's work in the individual portfolios (Dodge et al, 2001).

After the data has been collected and organized the teacher evaluates the individual child's progress by reviewing the observation notes and portfolio items. She then uses the Developmental Continuum to identify what step each child has reached for each of the 50 objectives. She uses the Class Summary Worksheet to keep track of children's progress at three checkpoints during the year. The Class Summary Worksheet lists the developmental domains and specific items within those domains. The teacher keeps track of the individual child by indicating where they are on the rating scale: Forerunner, Level I, Level II, or Level III. This same information is entered on the Individual Child Profile at each checkpoint (Dodge et al, 2001).

The teacher next plans how best to support each child's development and learning. She summarizes each child's progress on Child Progress and Planning Report. She meets with the families to share the information and plan next steps with them. She implements this plan as she uses what she has learned to help make decisions to extend learning. She continues to observe the child's progress. The teacher also plans for the group. Using the Class Summary Worksheet, a summary of developmental progress for each child within each domain, she reflects on the progress of the group. She decides what objectives to target for the whole group and for selected children. She plans strategies to support children's learning. She then implements her plan as she continues to observe the children's progress (Dodge et al, 2001).

Finally the teacher reports on the children's progress. She generates reports as needed. This can be done using software that analyzes and reports on individual and group progress. Through these reports she is able to identify aspects of the program that need strengthening and can thus develop a program improvement plan (Dodge et al, 2001).

Ages and Stages (ASQ)

The Ages and Stages Questionnaire (ASQ) is a parent completed child-monitoring system designed for specific age groups. This assessment tool screens infants and young children to assess

whether or not they have developmental delays and thus need early intervention. Parents or caregivers complete the questionnaires at designated intervals, assessing children in their natural environments so as to ensure valid results (EEC, 2007).

The ASQ is composed of a 30-item questionnaire, divided into five developmental areas: communication, gross and fine motor, problem solving and personal-social. There are six specific age relevant questions about a child's activity in each domain. These questions assess typical skills of a child at that particular point in time. An example of a communication query for a 48 month old is: Does your child use endings of words, such as "s," "ed," and "ing"? For example, does your child say things like, "I see two cats," "I am playing," or "I kicked the ball"? An example of a gross motor query for a 54 month old is: Does your child hop up and down on either his right foot or left foot at least one time without losing his balance or falling?"

The parent answers each question by choosing one of three choices: yes, sometimes, or not yet. The scores are then tallied, 10 points for yes, 5 points for sometimes and 0 points for not yet. The accumulated score is then compared with a chart that depicts typical development or ambiguous development. If the child's score falls into the typical developing area the ASQ suggests that the child is doing well at this point in time. If the child's score falls into the ambiguous developing area the ASQ suggests that the parent talk with a professional since the child may need further evaluation.

Each questionnaire also asks 8 overall questions that address specific biological development of the child. These same 8 questions, which are repeated for each age group, ask about the child's hearing, speech, vision, gross motor ability, and medical history. These questions require a yes or no answer (Bricker et al, 1999).

The developmental domains measured by COR, WSS, ASQ and Creative Curriculum assessment tools are social/emotional, physical, cognitive and language development. COR also includes approaches to learning such as initiative and creativity while WSS includes math and scientific thinking. The teacher in COR and WSS interprets the results, using them to guide activities and instruction, and provide information to parents. The information learned in Creative Curriculum is integrated into daily decisions regarding curriculum and individualization of instruction. Each ASQ serves as a summary of the child's performance on the questionnaire. The teachers can use this information to incorporate modifications for children with different learning abilities. These EEC approved tools are used to enhance learning in all developmental domains.

CONCLUSION

There are several concerns about the use of standardized tests for preschoolers that I focused on in this paper. I discussed how and why standardized tests are being administered to preschoolers. The potential impact of this practice on teachers and children and how to measure what young children have learned without using an assessment tool that may be harmful to them.

Standardized tests are being administered to preschoolers to identify children who may be in need of special services or to identify the existence of a disability, to assess the degree to which children are prepared for an academic or pre-academic program, and to measure the extent to which an individual has achieved certain information or attained skills that are identified within curricular objectives (Gullo, 2005). There are also formalized tests that are administered to young children to evaluate teacher and program effectiveness for accountability, in particular, the NRS.

Standardized tests employ a standard procedure for procuring results. In the case of the NRS preschoolers are asked a specific question and then are asked to answer the question by either pointing to the correct answer (from four images) or, in the case of the math skills subtest, saying the correct answer. The correct answer must be one of two or three specified answers.

The potential harm that standardized tests, such as the NRS, have on children is that the results can be used to deny services to children, to make the children repeat a grade, and they can make children feel inadequate and set them up for failure (especially since many standardized tests are culturally and developmentally biased).

“Children aren’t learning to think, and they are learning that a lot of school is useless. I worry that they are being turned off to learning. It’s almost like we want to create an underclass. If we want them to drop out of high school fail them when they are four.” Carol Seefeldt, an early childhood education professor at Hofstra University, author of *Guidelines for Preschool Learning and Teaching* (Vail, 2003, p.17).

The potential misuse of standardized tests, such as the NRS, by teachers is that the teachers may begin teaching to the test instead of teaching to the individual needs of the children. Teachers may narrow their curriculum and focus only on the topics that are covered on the test. Formalized tests can cause teachers to become less creative if they are expected to teach certain topics that are not necessarily based on the interests of the children.

How can we measure what young children have learned without using an assessment tool that may be harmful to them? It is imperative that we consider informal assessment techniques for measuring the achievement of young children. A wide range of developmentally appropriate ongoing assessments are available and are currently being used in various preschools. Many Head Start programs employ Creative Curriculum and High/Scope. These could be used for accountability purposes. We as educators need to be deliberately aware of the types of assessment tools that we administer to our children. We need to consciously object if an unsuitable assessment tool is being used in another early childhood program. Only then can we make sure that our nation's children are being educated and tested in a way that will enhance their educational experience.

On May 3, 2007 the House passed the latest five-year re-authorization of Head Start, and eliminated the requirement for using the NRS for testing 4 and 5 year olds. The Senate is expected to pass it in the next few weeks (PRS, 2007; Sunday Republican, 2007). Public and professional protest have been heard and responded to. We, as conscientious educators, can make a difference. We need to maintain this vigilance and do all we can to prevent harmful testing of our nation's children.

REFERENCES

- Appl, D. (2000) Clarifying the preschool assessment process: Traditional practices and alternative approaches. *Early Childhood Education Journal*, 27 (4), 219-225.
- Baumann, M. & Perlman, H. (2006) Settlement allows MCAS to stand as a graduation requirement. The Office of the Massachusetts Attorney General, May 26, 2006. <http://www.ago.state.ma.us/sp.cfm?pageid=986&id=1679>
- Bredenkamp, S. & Copple C., Eds. (1997) *Developmentally appropriate practice in early childhood programs*. Washington D.C.: The National Association for the Education of Young Children.
- Bricker, D. & Squire, J. (1999) *Ages & Stages Questionnaires: A parent completed, child-monitoring system, 2nd Edition*. Paul H. Brookes Publishing Co.
- The Center for Academic Excellence, Office of Instructional Services. *The Work Sampling System*. Retrieved February 26, 2007 from <http://www.thecenter.spps.org/wss.html>
- Champion, T., Hyter, Y., McCabe, A. & Bland-Stewart, L. (2003) A matter of vocabulary: Performances of low-income African American Head Start children on the Peabody Picture Vocabulary Test—III. *Communication Disorders Quarterly*, 24 (3), 121-127.
- Currie, J. & Thomas, D. (2000) School quality and the longer-term effects of Head Start. *Journal of Human Resources*, 35 (4), 755-774.
- Davis, M. (2005) House OKs Head Start reauthorization. *Education Week*, 25 (5), 5.
- Dodge, D., Colker, L. & Heroman, K. (2001) *A Teacher's guide to using the Creative Curriculum Developmental Continuum Assessment System*. Washington, D.C. : Teaching Strategies.
- Dodge, D., Colker, L. & Heroman, K. (2002) *The Creative Curriculum for preschool*. Washington, D.C. : Teaching Strategies.
- Dodge, D., Heroman, C., Charles, J. & Maiorca, J. (2004) Beyond outcomes: How ongoing assessment supports children's learning and leads to meaningful curriculum. *Young Children*, 59 (1), 20-29.
- Elmore, R. (2002) Unwarranted intrusion. *Education Next*, 1, 30-35. Retrieved May 9, 2007 from <http://www.hoover.org/publications/ednext/3367491.html>
- Epstein, A., Shweinhart, L., DeBruin-Parecki, A. & Robin, K. (2004) Preschool assessment: A guide to developing a balanced approach. *Preschool Policy Matters*, 7, 1-11.

- Fantuzzo, J., McWayne, C., Perry, M. & Childs, S. (2004) Multiple dimensions of family involvement and their relations to behavioral and learning competencies for urban, low-income children. *School Psychology Review*, 33 (4), 467-480.
- Farver, J. Xu, Y., Eppe, S., Fernandez, A. & Schwartz, D. (2005) Community violence, family conflict, and preschoolers' socioemotional functioning. *Developmental Psychology*, 41 (1), 160-170.
- Goffin, S.C. (2004) Moving into the next phase: Developing new accreditation assessment tools and addressing questions. *Young Children*, 59 (3), 70–73.
- Gormley, W. Jr. (2005) The universal pre-K bandwagon. *Phi Delta Kappan*, 87 (3), 246-249.
- Gullo, D. (2005) *Assessment and evaluation in early childhood education*. New York: Teachers College Press.
- Hanson, M.F. & Gilkerson, D. (1999) Portfolio assessment: More than ABCs and 123s. *Early Childhood Education Journal*, 27(2), 81-86.
- Haskins, R. (2004) Competing visions. *Education Next* 4, (1) 26-33.
- Hawken, L., Johnston, S. & McDonnell, A. (2005) Emergent literacy views and practices: Results from a national survey of Head Start preschool teachers. *Topics in Early Childhood Special Education*, 25 (4), 232-242.
- Head Start. (2000) *Head Start child outcomes framework*. Retrieved May 9, 2006, from http://www.headstartinfo.org/pdf/im00_18a.pdf
- Head Start Reauthorization: A Side-by-Side Comparison of H.R. 1429, S. 556, and Current Law April 18, 2007. Washington D.C.: *Congressional Research Reports for the People*. Retrieved May 9, 2007 from <http://opencrs.cdt.org/document/RL33968>
- Helm, J. H. (2003). Beyond the basics: Using the project approach in standards-based classrooms. *Dimensions of Early Childhood*, 31 (3), 6-12.
- Helm, J. H. & Gronlund, G. (2000) Linking standards and engaged learning in the early years. *Early Childhood Research & Practice*, 2 (1) (ERIC Document Reproduction Service ED439855)
- Helm, J. & Katz L. (2001). *Young investigators: The project approach in the early years*. New York: Teachers College Press.
- High/Scope Educational Research Foundation (1992) *High/Scope Child Observation Record (COR) for ages 2 ½-6*. Ypsilanti, Michigan.

- High/Scope Educational Research Foundation (2005). *Educational Programs: Early Childhood. High Scope Preschool Curriculum*. Retrieved February 26, 2007 from <http://www.highscope.org/EducationalPrograms/EarlyChildhood/homepage.htm>
- Israel, M. (2004) Ethical dilemmas for early childhood educators: The ethics of being accountable. *Young Children*, 59 (6), 24-32.
- Jacobson, L. (2004) Debate continues over Head Start assessment. *Education Week*, 23 (26), 10.
- Kafer, K. (2004) No child left behind: Where do we go from here? *Backgrounder*, 1775, 1-8. The Heritage Foundation, July 6, 2004. Retrieved May 9, 2007 from <http://www.heritage.org/Research/Education/bg1775.cfm>
- Katz, L. (1997) A developmental approach to the assessment of young children. (ERIC Document Reproduction Service ED407172)
- Katz, L. & Chard, S. (1989) *Engaging children's minds: The project approach*. Norwood, NJ: Ablex.
- Kostelnik, M., Soderman, A. and Whiren, A. (2004) *Developmentally appropriate curriculum: best practices in early childhood education*. Prentice Hall: Upper Saddle River, N.J.
- Kreisman, M. (2003) Evaluating academic outcomes of Head Start: An application of general growth mixture modeling. *Early Childhood Research Quarterly*, 18 (2), 238-254.
- La Paro K. & Pianta, R. (2000) Predicting children's competence in the early school years. A meta-analytic review. *Review of Educational Research*, 70 (4), 443-484.
- La Paro, K., Pianta, R. C. & Stuhlman, M. (2004) The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, 104 (5), 409-426.
- Magnuson, K., Meyers, M., Ruhm, C. & Waldfogel, J. (2004) Inequality in preschool education and school readiness. *American Educational Research Journal*, 41, (1), 115-157.
- Mantzicopoulos, P. & Knutson, D. (2000) Head Start children: School mobility and achievement in the early grades. *Journal of Educational Research*, 93 (5), 305-311.
- Mashburn, A. & Henry G. (2004) Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice*, 23 (4), 16-30.
- Massachusetts Department of Early Education and Care (2007). *Universal Pre-Kindergarten (UPK) Program Pilot Assessment Planning Grant: Appendix C Description Of EEC Approved Assessment Instruments*. Retrieved March 1, 2007 from

- http://www.eec.state.ma.us/docs/AppendixC_DescriptionOfEECAApprovedAssessmentInstruments.doc
- Massachusetts Department of Early Education and Care (2007). *Universal Pre-Kindergarten (UPK) Program Pilot Assessment Planning Grant*. Retrieved February 20, 2007 from <http://www.eec.state.ma.us/docs/UPKPilotAssessmentPlanningRFR.doc>
- Massachusetts Department of Education (2006) *Settlement Allows MCAS To Stand As Graduation Requirement*. Retrieved March 22, 2007 from <http://www.doe.mass.edu/news/news.asp?id=2907>
- Maxwell, K. & Clifford, R (2004) School readiness assessment. *Young Children*, 59 (1), 42-46.
- McElhenny, J. (2003) Judge rejects plea to shelve MCAS graduation requirement. *Boston Globe*, May 16, 2003. Retrieved May 11, 2006 from <http://www.boston.com/news/globe/>
- McWayne, C., Fantuzzo, J. & McDermott, P. (2004) Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, 40 (4), 633-645.
- Meisels, S. (2006) Accountability in early childhood: no easy answers. *Herr Research Center/Erickson Institute Applied Research in Child Development*, 6,1-28.
- Meisels, S. (2004) Should we test 4-year-olds? *Pediatrics*, 113, 1401-1402.
Retrieved from www.pediatrics.org by on April 9, 2007.
- Meisels, S. (1996/1997) Using work samples in authentic assessments. *Educational Leadership*, 54 (4), 60-65.
- Meisels, S. and Atkins-Burnett S. (2004) The National Reporting System: A critique. *Young Children*, 59 (1), 64-66.
- Meisels, S. (1995) Performance Assessment in Early Childhood Education: The Work Sampling System. Urbana IL: ERIC Clearinghouse on Elementary and Early Childhood Education. Retrieved March 22, 2007 from <http://www.ericdigests.org/1996-1/early.htm>
- Meisels, S., Jablon, J., Marsden, D., Dictelmiller, M. and Dorfman, A. (2005) *The Work Sampling System: Preschool through Grade 6. Program Overview*. Lebanon, IN: Pearson Learning Group.
- National Association for the Education of Young Children (1988) NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children*, 4 (3), 42-47.

National Association for the Education of Young Children (2005) *Code of Ethical Conduct and Statement of Commitment*. Retrieved May 12, 2006 from

<http://www.naeyc.org/about/positions/PSETH05.asp>

National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) Joint Position Statement (2002) *Early Learning Standards Creating the Conditions for Success*. Retrieved April 18, 2007 from

http://www.naeyc.org/...positions/pdf/position_statement.pdf

National Education Association (2002-2006) No Child Left Behind/ESEA: It's Time for a Change! Retrieved May 30, 2007 from <http://www.nea.org/lac/esea/index.html>

National Research Council Committee on Learning Research and Educational Practice, National Research Council: Bransford, J., Brown A. and Cocking R. editors (2000) *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, D.C.: National Academy Press.

Head Start National Reporting System Direct Child Assessment (2005-2006) *Baseline Assessment*. U.S. Department of Health & Human Services. Administration for Children and Families Administration for Children, Youth and Families Head Start Bureau.

Navarete, C., Wilde, J., Nelson, C., Martínez, R. and Hargett, G. (1990) Informal assessment in educational evaluation: Implications for Bilingual Education Programs. *National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs* Retrieved May 12, 2007 from <http://www.ncela.gwu.edu/pubs/pigs/pig3.htm>

Neisworth, J, & Bagnato S. (2004) The MisMeasure of young children: The authentic assessment alternative. *Infants and Young Children*, 17 (3), 198-212.

The New York State Education Department (1987) History of Regents Examinations: 1865 to 1987. *Bureau of Elementary and Secondary Testing Programs*, Albany, New York 12234 Retrieved May 31, 2007 from <http://www.emsc.nysed.gov/osa/hsinfogen/hsinfogenarch/rehistory.htm>

Paulson, L., Kelly, K., Jepson, S., van den Pol, R., Ashmore, R., Farrier, M. & Guilfoyle, S. (2004) The effects of an early reading curriculum on language and literacy development of Head Start children. *Journal of Research in Childhood Education*, 18 (3), 169-178.

Pretti-Frontczak, K. ,Kowalski, K.& Brown, R.D. (2002) Preschool teachers' use of assessments and curricula: A statewide examination. *Exceptional Children*, 69, (1), 109–124.

- PR Newswire (2007) House Head Start Reauthorization Vote Lauded by National Head Start Association. New York: *PR Newswire*. Retrieved May 9, 2007 from <http://www.prnewswire.com/cgi-bin/stories.pl?ACCT=109&STORY=/www/story/05-03-2007/0004580586&EDATE>
- Raab, 2005-2007, Teaching Research Institute, Western Oregon University Retrieved March 14, 2007 from <http://www.tr.wou.edu/train/cdc.htm>
- Ratcliff, N. (2001/2002) Using authentic assessment to document the emerging literacy skills of young children. *Childhood Education*, 78 (2), 66-69.
- Raver, C. & Zigler, E. (2004) Another Step Back? Assessing readiness in Head Start. *Young Children*, 59, (1), 58-63.
- Raver, C. & Zigler, E. (1991) Three steps forward, two steps back: Head Start and the measurement of social competence. *Young Children*, 46, (4), 3-8.
- Robinson, N., Lanzi, R., Weinberg, R., Ramey, S. & Ramey, C. (2002) Family factors associated with high academic competence in former Head Start children at third grade. *Gifted Child Quarterly*, 46 (4), 278-90.
- Robinson, A. (2003) *Letter to Head Start Clearance Officer (Administration for Children and Families Office of Information Services) on NAEYC's multiple concerns about the process, content, and timing for the new assessment of four-year old children in Head Start*. Retrieved March 13, 2007 from http://www.naeyc.org/policy/federal/pdf/comments_final.pdf
- Samuels, C. (2006) President seeking to hold line on Head Start, Child-Care Grants *Education Week*, 25 (23), 24-25.
- Schweinhart, L. and Storer, E. (2001) Using the High/Scope Preschool Child Observation Record (COR). *Head Start Bulletin*, 70. Retrieved April 23, 2007 from http://www.headstartinfo.org/publications/hsbulletin70/hsb70_06.htm
- Schumacher, R., Greenberg, M. & Mezey, J. (2003) *Head Start reauthorization: A preliminary analysis of HR 2210 the "School Readiness Act of 2003."* Washington, DC: Center for Law and Social Policy.
- Shepard, L., Kagan, S. & Wurtz, E. Eds. (1998) *Principles and Recommendations for Early Childhood Assessments*. Washington D.C.: National Education Goals Panel. Retrieved May 12, 2006 from <http://govinfo.library.unt.edu/negp/reports/prinrec.pdf>

- Spark, S. (2006) Early education programs get little support in budget. *Education Daily*, 39 (25), 6.
- Sunday Republican (May 6, 2007) *Head Start Renewal*. Springfield, MA.
- Vail, K. (2003) Ready to learn. *American School Board Journal*, 190 (1), 14-17.
- Vinovskies, M. (2005) *The birth of Head Start: Preschool education policies in the Kennedy and Johnson administrations*. Chicago: University of Chicago Press.
- Wagner, S. (2003) Assessment in the early childhood classroom: Asking the right questions, acting on the answers. *Herr Research Center/Erickson Institute Applied Research in Child Development*, 4, 1-8.
- The White House (2002) *Good Start, Grow Smart: The Bush Administration's Early Childhood Initiative*. Retrieved May 12, 2005 from <http://www.whitehouse.gov/infocus/earlychildhood/earlychildhood.html>
- The White House (2002) *No Child Left Behind*. Retrieved May 12, 2005 from <http://www.whitehouse.gov/news/reports/no-child-left-behind.html>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2006) Head Start Bureau. *Head Start History*. Retrieved May 9, 2006, from <http://www.acf.hhs.gov/programs/hsb/about/history.htm>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2006) Head Start Information and Publication Center. *National Reporting System Information and Resources*. Retrieved May 9, 2006, from http://www.headstartinfo.org/nrs_i&r.htm
- Wenning, R., Herdman, P., Smith, N., McMahon, N. & Washington, K. (2000) *No Child Left Behind: Testing, reporting, and accountability*. ERIC Clearinghouse on Urban Education, Institute for Urban and Minority Education Retrieved March 3, 2007 from <http://www.ericdigests.org/2004-2/behind.html>
- Zigler, E. and Styfco, S. (2004) Head Start's National Reporting System: A work in progress. *Pediatrics*, Vol. 114 No. 3 September 2004, pp. 858-859. Retrieved March 3, 2007 from <http://pediatrics.aappublications.org/cgi/content/full/114/3/858>
- Zill, N., Resnick, G., Kim, K., O'Donnell, K. and Sorongon, A. [Westat]; McKey, R.H., Pai-Samant, S., and Clark, C. [Xtria]; O'Brien, R. and D'Elio, M. [The CDM Group] (2003) *Head Start Faces 2000: A whole-child perspective on program performance*. The Administration for Child and Families, U.S. Department of Health and Human Services.