

May 2005

HEAD START

Further Development Could Allow Results of New Test to Be Used for Decision Making



GAO
Accountability · Integrity · Reliability

Highlights

Highlights of [GAO-05-343](#), a report to congressional requesters

Why GAO Did This Study

In September 2003, the Head Start Bureau, in the Department of Health and Human Services (HHS) Administration for Children and Families (ACF), implemented the National Reporting System (NRS), the first nationwide skills test of over 400,000 4- and 5-year-old children. The NRS is intended to provide information on how well Head Start grantees are helping children progress.

Given the importance of the NRS, this report examines: what information the NRS is designed to provide; how the Head Start Bureau has responded to concerns raised by grantees and experts during the first year of implementation; and whether the NRS provides the Head Start Bureau with quality information.

What GAO Recommends

GAO recommends the HHS Assistant Secretary for ACF, in collaboration with the Head Start Bureau, determine how NRS data will be used for accountability and targeting technical assistance; monitor the effects of the NRS on local Head Start practices; use first year NRS results to conduct further study of the reliability and validity of the NRS; compile a detailed, well-organized document on the technical quality of the NRS; improve management of its data on NRS participation; and study the costs and benefits of sampling in administering the NRS. ACF generally agreed with our recommendations.

www.gao.gov/cgi-bin/getrpt?GAO-05-343.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Marnie S. Shaull at (202) 512-7215 or shaullm@gao.gov.

HEAD START

Further Development Could Allow Results of New Test To Be Used for Decision Making

What GAO Found

The Head Start Bureau developed the NRS to gauge the extent to which Head Start grantees help children progress in specific skill areas, including understanding spoken English, recognizing letters, vocabulary, and early math. Due to time constraints and technical matters, the Head Start Bureau adapted portions of other assessments for use in the NRS.

Head Start Bureau officials have responded to some concerns raised during the first year of NRS implementation, but other issues remain. For example, the Head Start Bureau has modified training materials and is exploring the feasibility of sampling. However, it is not monitoring whether grantees are inappropriately changing instruction to emphasize areas covered in the NRS.

Head Start Bureau officials have said NRS results will eventually be used for program improvement, targeting training and technical assistance, and program accountability; however, the Head Start Bureau has not stated how NRS results will be used to realize these purposes. Currently, results from the first year of the NRS are of limited value for accountability purposes because the Head Start Bureau has not shown that the NRS meets professional standards for such uses, namely that (1) the NRS provides reliable information on children's progress during the Head Start program year, especially for Spanish-speaking children, and (2) its results are valid measures of the learning that takes place. The NRS also may not provide sufficient information to target technical assistance to the Head Start centers and classrooms that need it most.

An Assessor and Head Start Student Demonstrate the NRS Assessment.



Source: GAO.

Contents

Letter		1
	Results in Brief	3
	Background	5
	NRS Assesses Selected Skills Using Adaptations of Other Assessments	11
	The Head Start Bureau Has Been Responsive to Some Implementation Issues Raised during First Year of NRS, but Others Remain	17
	The Head Start Bureau Has Not Specified How NRS Results Will Be Used and Important Analyses Remain to Be Done	21
	Conclusions	28
	Recommendations for Executive Action	29
	Agency Comments and Our Evaluation	30
Appendix I	Objectives, Scope and Methodology	32
Appendix II	Survey Instrument	40
Appendix III	Comments from the Department of Health and Human Services	45
Appendix IV	GAO Contacts and Staff Acknowledgments	53
Tables		
	Table 1: Examples of Information Included in Computer-Based Reporting System (CBRS)	10
	Table 2: Description of NRS Components and Their Modifications	13
	Table 3: Sample Disposition	34
Figures		
	Figure 1: Head Start Grantees, Delegate Agencies, and Centers	6
	Figure 2: Timeline of Events Leading to Implementation of NRS	9
	Figure 3: Example of NRS Letter Naming Instructions and Task	14

Figure 4: Example of NRS Early Math Skills Instructions and Task	15
Figure 5: Example of Type of Vocabulary Instructions and Task Used in the NRS	16

Abbreviations

ACF	Administration for Children and Families
CBRS	Computer-Based Reporting System
ECLS-K	Early Childhood Longitudinal Study of a Kindergarten cohort
HHS	U.S. Department of Health and Human Services
HSB	Head Start Bureau
NAEYC	National Association for the Education of Young Children
NAS	National Academy of Sciences
NHSA	National Head Start Association
NRS	National Reporting System
OLDS	Oral Language Development Scale
PPVT	Peabody Picture Vocabulary Test
Pre-LAS 2000	Pre-Language Assessment Scale 2000
QRC	Head Start Quality Research Centers
TWG	Technical Work Group

This is a work of the U.S. government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



United States Government Accountability Office
Washington, DC 20548

May 17, 2005

The Honorable Edward M. Kennedy
Ranking Minority Member
Committee on Health, Education, Labor and Pensions
United States Senate

The Honorable Christopher J. Dodd
Ranking Minority Member
Subcommittee on Education and Early Childhood Development
Committee on Health, Education, Labor and Pensions
United States Senate

In fall 2003, the federal Head Start program initiated a nationwide skills test of over 400,000 4- and 5-year-old children. This test, called the Head Start National Reporting System (NRS), is intended to meet a long-standing need for systematic information on how well specific Head Start grantees are helping children learn. Head Start is designed to promote school readiness and healthy development among poor preschool children and provides services to nearly 1 million children, generally between the ages of 3 and 5, through nearly 1700 grantees. These grantees or their delegates provide services at about 19,000 Head Start centers nationally, with each grantee having from 1 to over 100 centers. For nearly a decade the Head Start Bureau (HSB) and the U.S. Department of Health and Human Services (HHS) have been engaged in promoting accountability and moving toward a results-oriented evaluation of Head Start. The NRS builds on this work. The NRS was developed in response to President Bush's April 2002 announcement of the "Good Start, Grow Smart" early childhood initiative that directed HHS to develop a national accountability system to ensure that every Head Start grantee will assess the progress made by children in early literacy, language, and numeracy skills.

Head Start teachers, or others trained as NRS assessors, administer the NRS to children individually in the fall and spring of the Head Start year. The NRS begins with a game of "Simon Says," lasts about 15 minutes, and includes four sub-tests designed to screen for understanding of spoken English and to assess skills in recognizing letters, vocabulary, and early math. During the test, an assessor sits across from a child at a table and asks scripted questions of the child, and the child responds by verbally identifying or pointing to pictures, numbers, or letters that are contained in a 3-ring binder. The assessor marks the child's responses on a

computer-readable scoring sheet. While all of the children are given at least the portion of the English-language assessment that screens for understanding of spoken English, children whose primary language is Spanish are also assessed using a Spanish version of the NRS. Children who speak both English and Spanish are given both versions of the NRS and scores from both tests are reported separately.

Although other evaluations of children's skills and Head Start performance exist, the NRS differs from them in its scale, type, and purpose. The NRS is a standardized test intended for all prekindergarten Head Start children. It represents the first time that HSB will use children's performance on a standardized test to measure how well specific Head Start grantees are helping children progress. Many in the Head Start community and beyond agree that it is a laudable goal to look at Head Start at the national and grantee levels to determine whether Head Start achieves its stated objectives. However, there have been significant concerns about whether the NRS, as currently composed, is the right way to accomplish this goal.

Given the importance HSB places on measuring Head Start performance and the concerns about the NRS, we examined (1) what information the NRS is designed to provide, (2) how HSB has responded to implementation issues raised by the Head Start grantees and experts during the first year of NRS implementation, and what issues remain to be addressed, and (3) whether the NRS provides HSB with the quality of information it needs to meet its purposes.

To answer these questions, we collected and analyzed information from multiple sources. To determine what information the NRS is designed to provide, we interviewed representatives from HSB, its contractors, and early childhood professional organizations and we reviewed documents chronicling the steps HSB took in developing the NRS. To examine how HSB responded to implementation issues raised by Head Start grantees and experts during the first year of NRS implementation and what issues remain to be addressed, we interviewed representatives from HSB and randomly sampled Head Start grantees and delegates from the population of all Head Start grantees and delegates during the 2003-2004 school year. We received responses from 80 percent of the grantees and delegates we surveyed. We also visited 12 Head Start grantees in 5 states (Colorado, Maryland, Massachusetts, Rhode Island, and Virginia), to interview staff who conducted the assessments and to observe them administering the NRS to children. The states and grantees chosen for site visits were judgmentally selected to include a range of enrollment sizes, types of program, rural and urban locations, and linguistic populations. Finally, to

examine whether the NRS provides HSB with the quality of information it needs to meet its goals, we reviewed the professionally accepted standards for test development, interviewed all of the members of the Technical Work Group—a team of experts convened to assist HSB and its contractors in the design and implementation of the NRS—and consulted with individuals recommended by the National Academy of Sciences as experts in the areas of test design and the educational testing of Spanish-speaking and bilingual children. These independent experts reviewed documents provided by HSB and its contractors pertaining to the adequacy and appropriateness of the assessment. See appendix I for additional information on our scope and methodology. We conducted our work between May 2004 and February 2005 in accordance with generally accepted government auditing standards.

Results in Brief

HSB developed the NRS to gauge the extent to which Head Start grantees help children progress in specific academic skill areas. The NRS includes materials adapted from other tests and is designed to provide information on selected academic skills of children in Head Start. Specifically, the NRS probes children's understanding of spoken English and skills in vocabulary, letter recognition, and simple math through the use of pictures, letters, and numbers. For example, children are asked to count marbles pictured on a page and identify the height of a teddy bear pictured beside a simple ruler. Children's skills in the selected areas are assessed to determine how well participating children, as a group, are learning and to identify grantees where children are not making the expected progress.

In response to concerns raised during the first year of NRS implementation, HSB has made changes to how the NRS is implemented and is considering other changes, although other concerns have not yet been addressed. In response to assessors' feedback that the initial training instructed assessors to follow the assessment script too rigidly, HSB modified some of its training materials to better prepare assessors for the situations they encountered when implementing the test. In addition, in response to suggestions by Technical Work Group members, HSB changed the order in which the Spanish and English assessments are administered. HSB is also considering substantive changes like requiring only a sample of children to take the NRS and adding a social-emotional development component to the NRS. According to our survey, over 60 percent of grantees found it at least moderately challenging to find time to assess all children, and sampling may help to minimize this burden. Adding a measure of social-emotional development would help to address concerns about the narrow range of skills that the NRS tests. While these changes

demonstrate HSB's responsiveness to some concerns raised, the Bureau has yet to address other potential implementation problems, such as whether all 4- and 5-year-olds eligible to participate in the NRS are assessed and whether assessors have narrowed the curriculum they teach in response to the NRS.

Analysis of the NRS is currently incomplete to support its use for the purposes of accountability and targeting training and technical assistance. First, HSB has not articulated a strategy for how it will use information from the NRS to meet its purposes. For example, it has not articulated what level of progress is expected, how it will use NRS scores to target training and technical assistance, or how it will hold grantees accountable for achieving results. Such decisions are important first steps in any test development process. Further, results from the first year of the NRS currently cannot be used to hold grantees accountable or to target training and technical assistance because HSB analyses have not yet shown that the NRS provides the scope and quality of assessment information needed for these purposes. The usefulness of educational tests is dependent on their consistency of measurement (their reliability), along with whether they measure what they are designed to measure (their validity). HSB has asserted that the NRS meets these criteria because it borrows certain material from existing tests that have met them, but the agency has not shown the NRS itself to be valid and reliable over time. Test developers generally use a pilot test to establish reliability and validity, but due to time constraints, HSB did not conduct a full pilot test. In addition, language experts advising HSB have raised serious concerns about whether the Spanish version of the NRS adequately measures the skills of Spanish-speaking children and whether results from the English and Spanish versions are comparable. Responding in part to these concerns, HSB has not yet used first year results of the NRS for accountability decisions and has stated that future accountability decisions will not be based solely on NRS results, but will reflect other grantee information as well. The NRS also may not provide sufficient information to target training and technical assistance to the centers and classrooms that need it most. NRS results are aggregated across the many classrooms and centers that a grantee may operate and results are reported only at the grantee and delegate levels, because results are more reliable at these levels than at lower levels. However, a grantee's average score could mask variability among the multiple classrooms or centers and limit information on where technical assistance would be most effectively targeted. Furthermore, NRS results alone do not indicate why results may be high or low, or what type of training or technical assistance would be appropriate.

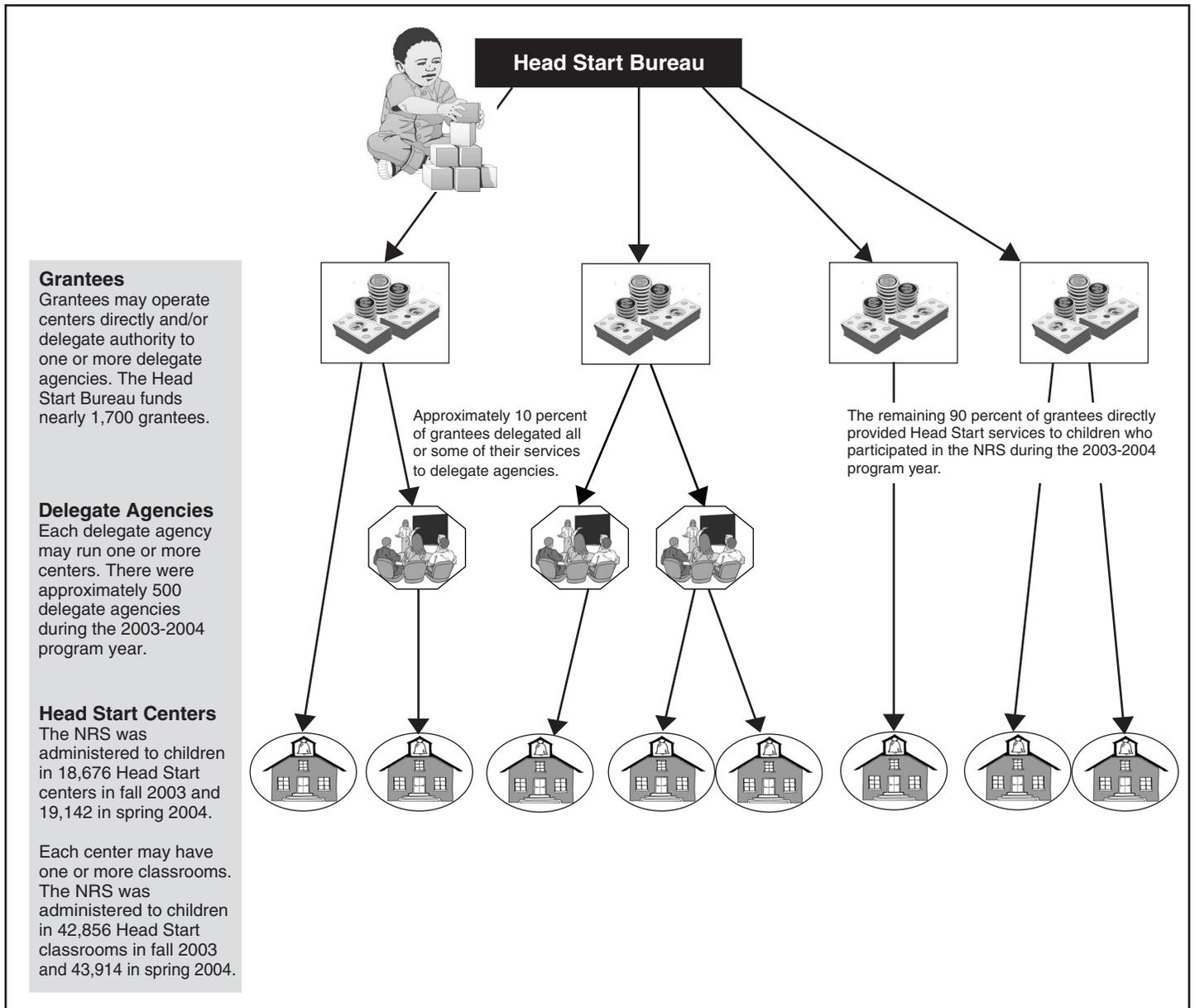
To help ensure that the NRS successfully and efficiently achieves its purposes, we are recommending that the HHS Assistant Secretary for the Administration for Children and Families (ACF) take several actions, including articulating plans for use of the NRS results, providing additional technical information on the test results, and conducting additional study of unintended effects and alternative ways for improving the test. ACF generally agreed with GAO's recommendations and described some of the actions it has already begun. In addition, ACF submitted detailed comments on certain aspects of the draft report, including comments concerning the level of evidence for the validity of the NRS.

Background

Established in 1965, Head Start is a federally funded early childhood development program that served over 900,000 children at a cost of \$6.8 billion in 2004. Head Start offers low-income children a broad range of services, including educational, medical, dental, mental health, nutritional, and social services.¹ Children enrolled in Head Start are generally between the ages of 3 and 5 and come from varying ethnic and racial backgrounds. Head Start is administered by HSB within ACF. HSB awards Head Start grants directly to local grantees. Grantees may develop or adopt their own curricula and practices within federal guidelines. Grantees may contract with other organizations—called delegate agencies—to run all or part of their local Head Start programs. Each grantee or delegate agency may have one or more centers, each containing one or more classrooms. In this report, the term “grantee” is used to refer to both grantees and delegate agencies. Figure 1 provides information on the numbers of Head Start grantees, delegate agencies, centers and classrooms.

¹Head Start regulations require that at least 90 percent of the children enrolled in Head Start come from families with incomes at or below the federal poverty guidelines, receiving public assistance, or caring for a foster child. In 2004, the federal poverty guideline for a family of four in the 48 contiguous states and the District of Columbia was \$18,850.

Figure 1: Head Start Grantees, Delegate Agencies, and Centers



Source: GAO analysis of HSB documents.

Since the inception of Head Start, questions have been raised about the effectiveness of the program. In 1998, we reported that Head Start lacked objective information on performance of individual grantees and Congress enacted legislation requiring HSB to establish specific educational standards applicable to all Head Start programs and allowed development of local assessments to measure whether the standards are met.² HSB implemented this legislation by developing the Child Outcomes Framework to guide Head Start grantees in their ongoing assessment of the progress of children. The Framework covers a broad range of child skill and development areas and incorporates each of the legislatively mandated goals, such as that children “use and understand an increasingly complex and varied vocabulary” and “identify at least 10 letters of the alphabet.”

Since 2000, HSB has required every Head Start grantee to include each of the areas in the Framework in the child assessments that each grantee adopts and implements. The eight broad areas included in the Framework are language development, literacy, mathematics, science, creative arts, social and emotional development, approaches to learning, and physical health and development. Grantees are permitted to determine how to assess children’s progress in these areas. These assessments are to align with the grantee’s curriculum; as a result the specific assessments vary across the grantees. The assessments occur 3 times each year and generally involve observing the children during normal classroom activities.³ The results of the assessments are used for the purposes of individual program improvement and instructional support and are not aggregated across grantees or systematically shared with federal officials. The NRS, prompted by the April 2002 announcement of President Bush’s Good Start, Grow Smart initiative, builds on the 1998 legislation by requiring all Head Start programs to implement the same assessment, twice a year, to all 4- and 5-year-old Head Start participants who will attend kindergarten the following year.

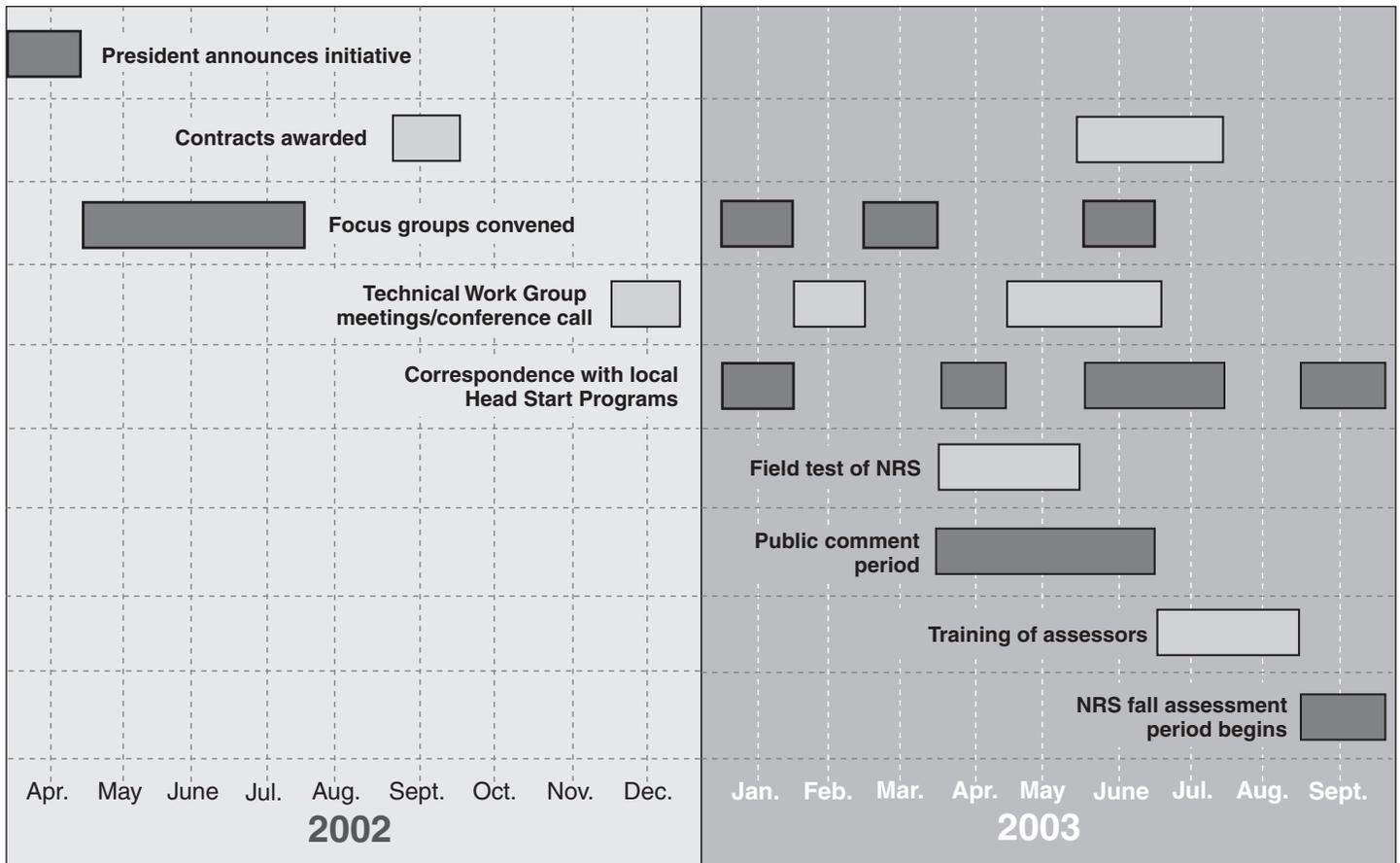
²See GAO, *Head Start: Challenges in Monitoring Program Quality and Demonstrating Results*, GAO/HEHS-98-186 (Washington, D.C.: June 1998), and *Head Start: Curriculum Use and Individual Child Assessment in Cognitive and Language Development*, GAO-03-1049 (Washington, D.C.: September 2003).

³ According to ACF officials, in addition to the assessments conducted as part of the Head Start Child Outcomes Framework, Head Start teachers must observe and record examples of children’s development and learning on an ongoing basis throughout the year.

When President Bush announced this initiative in April 2002, it called for full implementation in fall 2003; as a result the NRS was developed and preparations for implementation occurred within an 18-month period. See figure 2. Shortly after the President announced this initiative, HSB hired a contractor to assist it in developing and implementing the NRS. The contractor, working closely with HSB, was responsible for the design and field testing of the NRS, including developing training materials to support national implementation of the reporting system by grantees.⁴ HSB also worked with the Technical Work Group and others throughout implementation of the NRS. The Technical Work Group includes 16 experts in such areas as child development, educational testing, and bilingual education. They advised HSB on the selection of assessments, the appropriateness of the assessments in addressing the mandated indicators, the technical merit of the assessments, and the overall design of the NRS. While the Technical Work Group members offered advice, the group members were not always in agreement with each other and HSB was not obligated to act on any of the advice it received. A list of the Technical Work Group members and their professional affiliations is included in appendix I.

⁴ Analyses and actions taken by the Head Start Bureau's contractors are attributed to the Head Start Bureau itself.

Figure 2: Timeline of Events Leading to Implementation of NRS



Source: HSB documents and interviews with HSB officials.

Through focus groups, teleconferences, and various correspondences, HSB officials communicated to Head Start grantees the purpose of the NRS and their plans for administering the assessment. Focus groups and discussions were held with various interested parties, including Head Start managers and directors and experts from universities and the public sector, on issues ranging from strengths and limitations of various assessment tools to strategies for assessing non-English speaking children. HSB also received input through a 60-day public comment period, from mid-April to June 2003.

Another contractor developed a Computer-Based Reporting System (CBRS) for the NRS. Local Head Start staff use the CBRS to enter descriptive information about their grantees, centers, classrooms,

teachers, and children, as shown in table 1, as well as to keep track of which children have been assessed. HSB analyzes the descriptive information from the CBRS in conjunction with the child assessment data to develop reports on the progress of specific subgroups of children. For example, HSB can report separately on the average scores of children enrolled in part-day programs and those enrolled in full-day programs.

Table 1: Examples of Information Included in Computer-Based Reporting System (CBRS)

Program information	Center information	Classroom level information	Assessor information	Teacher information	Child information
<ul style="list-style-type: none"> • Program name • Director name • Number of delegates • Number of centers • Number of family day care centers • NRS lead for program 	<ul style="list-style-type: none"> • Center name • Center type • Enrollment year start date • Enrollment year end date • NRS center lead name 	<ul style="list-style-type: none"> • Teacher name • Classroom type • Day option • Total enrollment • Number of additional teaching staff • Teacher entry date to classroom 	<ul style="list-style-type: none"> • Name • Highest grade or year of school completed • Highest degree held in Early Childhood Education or related field 	<ul style="list-style-type: none"> • Teacher name • In what languages is teacher fluent? • Total years teaching • How many years teaching Head Start? • Highest grade or year of school completed • Child Development Associate credential 	<ul style="list-style-type: none"> • Child name • DOB • Date of entry into classroom • Child unique ID from center • Years in preschool Head Start • Does child have a disability? • Does child speaks a language other than English at home? • If yes, how well does child speak English? • If yes, what is primary language? • Ethnicity/race

Source: Head Start National Reporting System, Computer-Based Reporting System Train-the-Trainer Manual, Prepared by Xtria, LLC, February 2004.

HSB, with assistance from the contractors, worked to ensure local staff received adequate training on administering the assessment and using the CBRS, and provided guidance on how to obtain consent from parents. Training and certification of all assessors was required so that all assessors would administer the NRS in the same way. Two-and-a-half day training sessions were held at eight sites throughout the U.S. and Puerto Rico during July and August 2003. Roughly 2,800 individuals completed the training, of which 484 were certified in both English and Spanish. In turn, these certified trainers held training sessions locally to train and certify additional staff who would be able to administer assessments.

The development of educational tests is a science in itself, to which university departments, professional organizations, and private companies are devoted. Among the most important concepts in test development are validity and reliability. Validity refers to whether the test results mean what they are expected to mean and whether evidence supports the intended interpretations of test scores for a particular purpose. Reliability refers to whether or not a test yields consistent results. Validity and reliability are not properties of tests; rather, they are characteristics of the results obtained using the tests. For example, even if a test designed for 4th graders were shown to produce meaningful measures of their understanding of geometry, this wouldn't necessarily mean that it would do so when administered to 2nd or 6th graders or with a change in directions allowing use of a compass and ruler. Test developers typically implement "pilot" tests that represent the actual testing population and conditions and they use data from the pilot to evaluate the reliability and validity of a test. This process generally takes more than 1 year, especially if the test is designed to measure changes in performance.

In the remainder of the report, we will discuss how the focus of the NRS was determined and the assessment was developed, HSB's response to problems in initial implementation as well as some implementation issues that remain unaddressed, and the extent to which the assessment meets the professional and technical standards to support specific purposes identified by HSB.

NRS Assesses Selected Skills Using Adaptations of Other Assessments

The NRS assesses vocabulary, letter recognition, simple math skills, and screens for understanding of spoken English. As initially conceived by HSB, the NRS was to gauge the progress of Head Start children in 13 congressionally mandated indicators of learning. However, time constraints and technical matters precluded HSB from assessing children on all of the indicators and led HSB to consider, and eventually adopt, portions of other assessments for use in the NRS.

The 18 months from announcing the Good Start, Grow Smart initiative, of which the NRS is a part, to implementing the assessment was not enough time for HSB to develop a completely new assessment. Therefore, HSB, with the advice of its contractor and the Technical Work Group, chose to borrow material from existing assessments. Concerns raised by Technical Work Group members and the contractor about the length and complexity of the assessment and the technical adequacy of individual components eventually led to limiting the areas assessed in the NRS, from 13 skills to 6.

The six legislatively mandated skills that HSB targeted included whether children in Head Start:

- use increasingly complex and varied spoken vocabulary;
- understand increasingly complex and varied vocabulary;
- identify at least 10 letters of the alphabet;
- know numbers and simple math operations, such as addition and subtraction;
- for non-English speaking children, demonstrate progress in listening to and understanding English; and
- for non-English speaking children, show progress in speaking English.

In April and May of 2003 an assessment that included 5 components covering the 6 skills was field tested with 36 Head Start programs to examine the basic adequacy of the NRS, as well as the method for training assessors, and the use of the CBRS. The field test also included a Spanish version of the NRS. Based on the field test, one component—phonological awareness, or one’s ability to hear, identify, and manipulate sounds—was eliminated. While this component examined an area that experts have linked to prevention of reading difficulties, the test used to assess it was problematic. HSB moved forward with the other components of the NRS. The four components of the NRS each measure one or more of the six legislatively-mandated indicators.

The four components that comprise the NRS are from the following tests:

- Oral Language Development Scale (OLDS) of the Pre-Language Assessment Scale 2000 (Pre-LAS 2000),
- Third Edition of the Peabody Picture Vocabulary Test (PPVT-III),
- Head Start Quality Research Centers (QRC) letter-naming exercise, and
- Early Childhood Longitudinal Study of a kindergarten cohort (ECLS-K) math assessment.

Some or all of each test was previously used for other studies, and the PPVT and letter naming were previously used in studies of Head Start

children.⁵ Three of the four tests were modified from their original version, as shown in table 2. Figures 3 and 4 are examples from the letter naming and early math skills components of the NRS. Figure 5 is an example of the type of item used in the vocabulary (PPVT) component of the NRS.

Table 2: Description of NRS Components and Their Modifications

NRS components	Modifications to components	Description of components	Legislatively-mandated skill measured by component
Oral Language Development Scale (OLDS) of the PreLAS 2000 (comprehension of spoken English)	NRS includes two subtests from the original assessment	Simon Says—The child is asked to follow the instructions that “Simon says,” such as “Simon says, “Touch your toes.” Art Show—The child is presented with a series of 10 pictures and asked to name or explain what is in each picture.	Use increasingly complex and varied spoken vocabulary. For non-English speaking children, demonstrate progress in listening to and understanding English. For non-English speaking children, show progress in speaking English.
Third Edition of the Peabody Picture Vocabulary Test (PPVT-III)	NRS includes 24 items from what was originally a 144-item test	The child is asked to point to pictures to demonstrate understanding of words representing parts of the human body or their functions, activities of daily living, emotions and feelings, work/career-related activities, and plants, animals, and their habitats.	Understand increasingly complex and varied vocabulary.
Head Start Quality Research Centers (QRC) letter-naming exercise	None	The child is shown all 26 letters of the alphabet, divided into three groups of 8, 9, and 9 letters, and arranged in approximate order of item difficulty, and is asked to identify the letters they know by name	Identify at least 10 letters of the alphabet.
Early Childhood Longitudinal Study of a kindergarten cohort (ECLS-K) math assessment	NRS includes items in the easier range of the original assessment	Using pictures, the child is asked about a range of math skills: number recognition of 1-digit numerals, basic geometric shapes, matching number names with objects, counting, simple addition and subtraction, and interpreting simple measurements and graphic representations.	Know numbers and operations.

Source: GAO analysis of HHS documentation.

⁵Both the OLDS and the math assessment were used in the ECLS-K, and the PPVT-III was used with two cohorts of the Head Start Family and Child Experiences Survey (FACES). The Head Start Quality Research Centers letter-naming exercise was developed for use in Head Start curriculum studies. The ECLS-K is an ongoing study that focuses on children’s early school experiences beginning with kindergarten and following children through fifth grade. FACES is a national longitudinal study of the development of Head Start children, their families, and Head Start programs and staff in a small sample of programs.

Figure 3: Example of NRS Letter Naming Instructions and Task

Here are some letters of the alphabet.

GESTURE WITH A CIRCULAR MOTION AT LETTERS AND SAY:

Point to all the letters that you know and tell me the name of each one. Go slowly and show me which letter you're naming.

INDICATE ONLY CORRECTLY NAMED LETTERS ON ANSWER SHEET.

WHEN CHILD STOPS NAMING LETTERS, SAY:

Look carefully at all of them. Do you know any more?

KEEP ASKING UNTIL CHILD DOESN'T KNOW ANY MORE.

A a	O o	S s
B b	E e	C c
D d	X x	

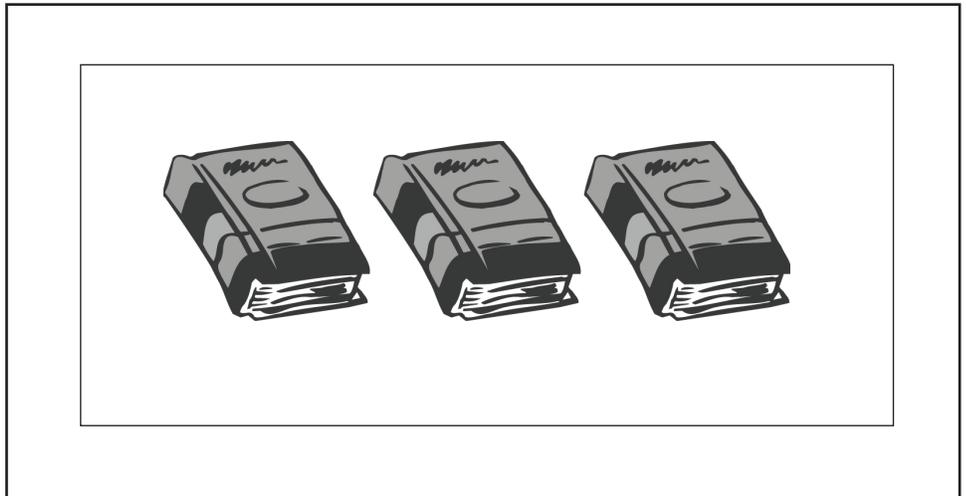
Source: U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, "Full National Implementation of the Head Start National Reporting System on Child Outcomes, Office of Management and Budget Clearance Package Supporting Statement and Data Collection Instruments," June 23, 2003.

Figure 4: Example of NRS Early Math Skills Instructions and Task

RUN YOUR FINGER ACROSS THE ITEM AND SAY:

If you gave a friend one of these books, how many books would you have left?

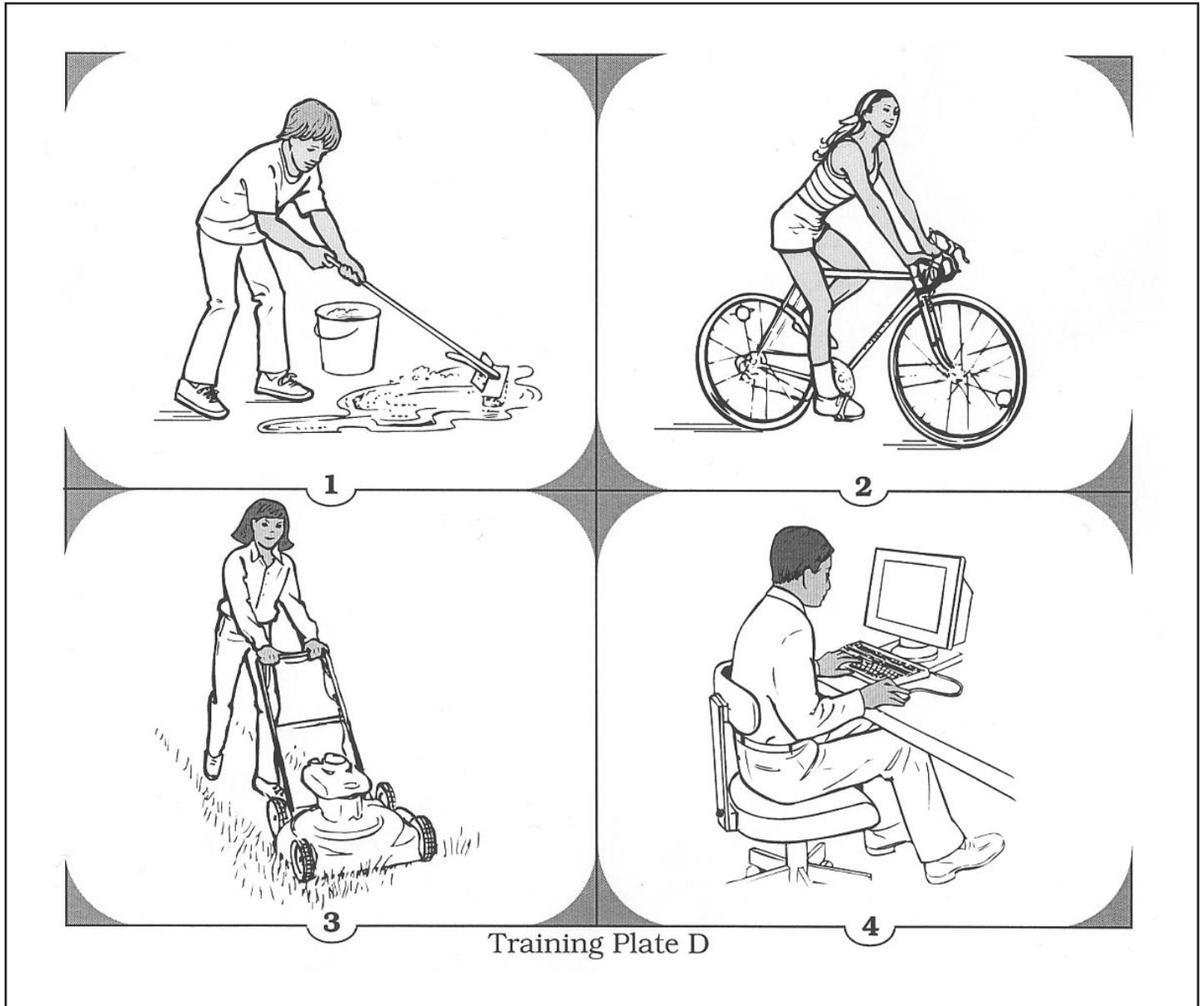
CORRECT: TWO (BOOKS)



Source: U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, "Full National Implementation of the Head Start National Reporting System on Child Outcomes, Office of Management and Budget Clearance Package Supporting Statement and Data Collection Instruments," June 23, 2003.

Figure 5: Example of Type of Vocabulary Instructions and Task Used in the NRS

Say: point to mowing.



Source: PPVT-III. ©1997 Lloyd M. Dunn, Leota M. Dunn and Doug M. Dunn.

The Head Start Bureau Has Been Responsive to Some Implementation Issues Raised during First Year of NRS, but Others Remain

HSB has been responsive to some specific implementation concerns about the NRS, but other issues remain that might pose problems in the future. HSB already has made modifications to NRS training materials, the CBRS, and how the Spanish NRS is administered. In addition, HSB is working with the Technical Work Group to explore the feasibility of adopting a sampling strategy and including a measure of social-emotional development in the NRS. HSB has told grantees not to make changes to their programs based on the first year of the NRS, but our survey found that some grantees have changed instruction to emphasize areas covered in the test.⁶ While some such change may be appropriate, HSB currently is not monitoring whether grantees are changing the content of instruction to de-emphasize areas not tested or adopting inappropriate styles of teaching.

HSB Has Responded to Some Implementation Issues That Arose during the First Year of NRS

Based on grantee feedback about their experiences during the first year of NRS implementation, HSB has already responded to some concerns by providing additional guidance on handling children's behavior, making it easier for Head Start staff to use the CBRS, and changing the order in which the Spanish and English versions of the NRS are administered to Spanish speaking children. These changes are, in part, a response to feedback from local assessors and concerns raised by Technical Work Group members. During our site visits, some assessors described the 2003 NRS training as rigid, with a lot of emphasis placed on following the script. HSB addressed these concerns in the 2004 spring refresher training video. Assessors agreed that this video better reflected the situations they encountered when assessing young children, such as a child who fidgets, has to go to the bathroom or wants a drink of water during an assessment.

In addition to changing training material, HSB added several new features to the CBRS in response to information contractors gleaned while fielding assessors' phone calls for technical assistance. For example, the CBRS initially required local Head Start staff to type in all necessary information about their students, but the fall 2004 version of the CBRS allowed local

⁶We use the terms "the test" and "the assessment" to make shortened reference to the NRS test battery. The NRS also incorporates a support infrastructure for the test battery, including a system for training staff to conduct the assessments and a computer-based reporting system. While the NRS may eventually be expanded to incorporate additional components,[0] we examined it as implemented through spring 2004. [0]

staff to update information about their children using information from the previous year or by transferring information from other computer systems.

Another change to the NRS is the order in which the Spanish and English assessments are administered to Spanish speaking children. Some TWG members suggested that by administering the NRS first in English and secondly in Spanish to Spanish-speaking children with limited English proficiency, the children will have experienced difficulty and frustration during the English test. These feelings of frustration or failure could affect a child's disposition—and a child's responses—when later taking the Spanish version. Thus, the validity of the Spanish assessment might be compromised. During summer 2004, Migrant and Seasonal Head Start Programs administered the assessment in Spanish first. Based on the positive response they received from local assessors, HSB instructed all programs to follow this format in fall of 2004.

HSB Is Considering Sampling Strategies and Broadening NRS to Include a Measure of Social-Emotional Development

HSB is considering ways to deal with two issues raised during the first year of implementation: the burden on grantees in dedicating staff for the assessments and the limited range of skills that were assessed in the NRS. In particular, HSB is considering the feasibility of sampling to minimize the burden that grantees experienced in assessing all 4- and 5-year-old Head Start participants who will attend kindergarten the following year. According to our survey, finding time to conduct assessments presented at least a moderate challenge to an estimated 63 percent of grantees and allocating staff to administer the NRS presented at least a moderate challenge for an estimated 42 percent of grantees during the first year of the NRS. According to most of the assessors we spoke to (8 of 12) during our site visits, local staff neglected other tasks, juggled tasks, or took work home because they were occupied with administering the NRS. Assessors also mentioned having to reschedule training and reallocate staff because of the NRS.

Several Technical Work Groups members and grantees have suggested sampling as a way for the NRS to provide better information while reducing the burden on grantees. Sampling would allow staff to spend more time in the classroom and would cost less. Responding to these suggestions, HSB is working with some members of the Technical Work Group to identify various sampling strategies and their practical implications. These sampling strategies include matrix sampling, which involves taking a subset of items from the larger assessment and randomly assigning them to test takers, thereby avoiding the need to administer all items to all test takers. Matrix sampling would allow for more items to be

included and, therefore, more in-depth assessment of the subjects covered by the test. Drawing an appropriate sample is complicated, however, and it might be difficult to learn how subgroups are doing, by region or subpopulation, using sampling or matrix sampling.

In addition to studying the feasibility of sampling, HSB is actively exploring ways to incorporate a measure of social-emotional development into the NRS. Technical Work Group members have argued that social-emotional development is critical to kindergarten success and adding a measure of social-emotional development would begin to address criticisms that the scope of the NRS currently is too narrow. A Technical Work Group subcommittee has identified eight measures of social-emotional development for possible field-testing. In addition, HSB has directed its contractor to conduct a small pilot to assess the feasibility of these measures and to conduct focus groups to obtain teacher feedback on the measures. Following the pilot test and focus groups, the contractor will conduct a field test with 30 Head Start programs to determine the appropriateness and technical adequacy of the measures.

HSB Has Not Yet Addressed Some Concerns

While HSB is addressing some issues associated with the NRS, additional implementation concerns have yet to be addressed. HSB currently lacks independent information to verify that grantees are assessing all of the children eligible to participate in the NRS. Thus, the potential exists for undetected errors or exclusion of children HSB intends to be assessed. HSB attempts to ensure it has accurate information in several ways. For example, HSB compares the number of 4- and 5-year-olds reported in the current year with information from the previous year and it analyzes the data for inconsistencies and discrepancies.⁷ However, beyond these checks, HSB does not have an independent way to confirm the number of children eligible to participate in the NRS.

There is also a concern that local Head Start programs will alter their teaching practices and curricula based on their participation in the NRS. These alterations, whether intended or unintended, might have positive and negative consequences. Local assessors are generally Head Start staff and it is expected that they want their children to perform well on the NRS and that they will teach their children the specific skills measured in the NRS. An increased focus on teaching these skills could be positive to the

⁷The current year's data are not available until December.

extent they have been neglected. However, this focus would be detrimental if it resulted in narrowing the curriculum to exclude skills that are not measured on the NRS but that experts believe are equally important for children's development. HSB specifically told grantees not to make changes to their programs based on their initial NRS results and has provided guidance on appropriate instruction. Nonetheless, according to our survey of assessors, at least an estimated 18 percent of grantees changed instruction during the first year of NRS implementation to emphasize areas covered in the NRS. One assessor we interviewed explained that despite being told during NRS training that programs should not adjust their curricula, it is human nature to try to correct areas in need of improvement. Without additional information, it is not possible to determine whether changes in instruction are positive or negative.

Despite HSB's assurances that it intends to use the NRS results only in the context of other information on performance, experts state that grantees' perception of the NRS as a "high stakes" test could compromise the test within a few years. Assessors are very involved in the scoring of the NRS, yet the NRS is evaluating the grantees that employ them; thus, they are not independent. Assessors' input and interpretations could make the grantee appear to accomplish its goals, whether it actually does or not. For example, one assessor commented that participating in the NRS had planted a seed that perhaps she should teach her children particular words that appear in the NRS, such as the word "altogether," which appears in the instructions. It is also worth noting that the words used to screen for understanding of English were exactly the same in fall 2003 and spring 2004, so that learning particular words would make a large difference. An independent expert argued that there needs to be continuous monitoring and retraining of NRS assessors, as there was during the first year of NRS implementation, to maintain quality control over the testing process. For the second year of the NRS, HSB has extended its effort to review the quality of assessment administration, but these efforts do not include monitoring of changes in classroom practices.

Additionally, in the absence of clear direction from HSB, local Head Start staff might misinterpret the results and use them inappropriately. The Technical Work Group has been clear that NRS scores for classrooms and individual children are not reliable and should not be used at the classroom level or for individual child evaluation or instruction. Yet, two of the Head Start grantees we visited stated that they photocopied each child's responses before returning the completed scoring sheets and one stated that the grantee intended to use the individual test results to evaluate its own performance at the classroom level. Technical Work

Group members have argued that local Head Start programs should be given clear information on how to interpret the NRS results and how to improve their programs if they are unhappy with their NRS scores; however, the Technical Work Group members themselves have expressed confusion about how to interpret NRS scores, given the technical issues that are discussed in detail in the next section.

The Head Start Bureau Has Not Specified How NRS Results Will Be Used and Important Analyses Remain to Be Done

HSB has not said specifically how it will use the NRS results and HSB currently lacks analyses showing that the NRS provides the scope and quality of information needed to hold Head Start grantees accountable or target training and technical assistance. To support these purposes, the NRS must produce valid and reliable results on children's performance that would allow for clear conclusions about Head Start grantees' effectiveness in improving the academic performance of children. Due to time constraints, HSB did not conduct a pilot test that could have provided information to establish the reliability and validity of changes in the NRS results over time. Experts have also questioned the technical merit of the Spanish-language NRS. Apart from these concerns, the NRS results alone do not provide enough contextual information to support accountability decisions. Acknowledging some of these issues, HSB has stated that accountability decisions will not be based solely on NRS results, and it will consider other grantee information, though it has not explicitly described how NRS results will be interpreted. Finally, because multiple classrooms are averaged to produce grantee results and this average may mask variability among different classrooms, NRS results are of limited use to target training and technical assistance to the classrooms where assistance is needed most.

Head Start Bureau Has Not Stated How It Will Use NRS Results to Achieve Its Purposes

Head Start Bureau officials have stated in general terms that they will use NRS results to improve program performance, target training and technical assistance and hold Head Start grantees accountable; however, it remains unclear whether the NRS' purposes will be realized because HSB has not explained how assessment results will be used. For example, as of February 2005, HSB had not specified what grantee scoring level constitutes adequate performance. In addition, it had not indicated whether HSB would adjust scores to account for age or other differences among the children grantees serve, how it would account for students with disabilities, or whether adequate performance would be measured in absolute terms (e.g., the average score or the percentage of children that score above a certain level) or by growth in performance (performance change from fall to spring assessment).

Professional standards for educational testing require that test developers specify how results will be used prior to developing a test so that judgments can be made about the appropriateness of the test. The specific uses of the NRS dictate the specific technical criteria it should meet. For example, if HSB intends to hold grantees accountable for increasing their assessment scores by a particular percentage, the NRS would need to be sensitive enough to reliably measure increases of that size. Several Technical Work Group members have emphasized the point that HSB should have determined exactly how it intended to use the NRS as a first step in the development of the NRS. As of February 2005, HSB officials had not indicated when they would make decisions about the specific uses of the NRS data or when they would provide this information to grantees.

This ambiguity has left some grantees wondering what the consequences could be of their assessment results. Assessors from 6 of the 12 Head Start grantees we visited said they were concerned about how HSB would use the NRS. Assessors from two grantees expressed apprehension that the results would be misinterpreted as evidence regarding the effectiveness of the program. One assessor suggested that HSB should share with local Head Start staff how it plans to use the data because it would generate greater support for the NRS among staff. These findings are consistent with recommendations from a quality assurance study, commissioned by HSB, that recommended HSB provide more information on how it will use the results of the NRS assessments, especially with respect to implications for training and technical assistance, program improvement, and funding, to alleviate the concerns of grantees.⁸ HSB has stated that it is focusing on how to work with grantees on understanding NRS results and how to use the information to make improvements through training and technical assistance.

⁸The Head Start Bureau awarded a contract to Mathematica Policy Research, Inc., to conduct an implementation study of the NRS in a randomly-selected set of 35 Head Start programs. The research team observed a total of 119 local assessors, interviewed Head Start directors, NRS trainers, and data managers, and held focus groups with staff conducting the assessments to learn about their experiences. Mathematica also planned to visit four Migrant and Seasonal Head Start programs during spring 2004 and fall 2005.

**Results from First Year
Cannot Be Used to Hold
Grantees Accountable
Because Important
Analyses Have yet to Be
Completed or Documented**

In order to use the NRS for the purpose of holding grantees accountable for children's progress, HSB needs to demonstrate that the NRS will provide reliable and valid information. As of February 2005, HSB had not, however, conducted certain analyses on NRS results to establish the validity and some aspects of the reliability of the assessment. A test is considered valid when it measures what it is supposed to measure and evidence supports the intended interpretations of test scores for a particular purpose. Reliability refers to whether or not a test yields consistent results, meaning that if a child in Head Start took the NRS on, say, a different day, that his or her score would be similar.

HSB tested the reliability of particular NRS items through a short field test, but given the time constraints on the development of the NRS, HSB did not run a more extensive "pilot" test prior to full implementation. The field test results provided some information on the reliability of the NRS components for one point in time, which generally was strong at the grantee level. However, HSB lacked information on the range of growth that children might experience over the course of a year and—consequently—did not have the data to show that the test produces valid and reliable results on change from fall to spring. Some assessors also have expressed doubt about whether the NRS accurately measures change over time. According to our survey of NRS assessors, about a quarter of assessors agree that the NRS accurately measures the progress of their Head Start children from fall to spring. Further, without additional data from a pilot test, HSB could not fully validate the NRS and ensure that its use for the intended purposes was appropriate.

Despite not conducting a pilot test, HSB stated that the NRS was technically sound in large part because it borrowed sections from tests that produced valid and reliable results in previous studies. Relying on this past work instead of conducting a new pilot test allowed HSB to develop the NRS within a very short time frame, but there are problems with this approach. The sample of children in these past studies is not always the same as the Head Start children with regard to age, home language, culture, or range of socio-economic status. Moreover, some of the tests used in the past were modified for use in the NRS by either limiting the questions asked or modifying the instructions. Without further analyses of the actual NRS implementation data, it is impossible to determine whether interpretations of the NRS results for the purpose of accountability are valid. Data from the first year of implementation could now be used to conduct some of these analyses and make determinations. For this reason, some Technical Work Group members have suggested that the first year of NRS implementation should have been considered a pilot test. HSB

officials stated recently that they would be working with the Technical Work Group and a new advisory committee to continue to review the quality, reliability, and validity of the NRS assessment.

Technical Work Group members have noted specific concerns with the approach and format of the NRS that may be threats to its validity. For example, Technical Work Group members have criticized the math section for asking children to refer to items pictured on a page rather than providing physical items (e.g., blocks) to handle and have argued that the instructions are complicated for 4- and 5-year-old children. They argue children might fail items due not to lack of math skills, but because they do not understand the instructions or they lack the ability to perform the math operations without items that can be manipulated. Technical Work Group members also questioned whether the letter-naming task is a valid measure of how many letters the children know. Given the layout of the letters on the page, a child can miss letters even if he or she actually knows the names of the letters, or may tire of naming them and seek to see what is on the next page. Several of the assessors we interviewed echoed these concerns and also raised concerns about the quality of the pictures and choice of vocabulary used in the PPVT component of the NRS. Due in part to these concerns, only about half of lead assessors believe that the NRS accurately portrays the majority of their children's abilities.

Currently, HSB cannot use the results from the Spanish version of the NRS for accountability purposes because it has not been demonstrated that this version produces reliable and valid results or that its results are comparable to those from children tested in English. While it is important that a Spanish version was developed due to the fact that 20 percent of Head Start children speak Spanish, experts have questioned the reliability of the Spanish NRS results and criticized other aspects of this version. First, the Spanish version of the NRS was not standardized for the Spanish-speaking Head Start population. Because the country of origin and class of a child's family affect the Spanish dialect he or she speaks, there are important language differences among subpopulations, making such standardization important. For example, the Spanish spoken in Puerto Rico differs from that in Mexico and children from these countries are likely to recognize and use different words in test questions and answers. A number of NRS assessors commented to us that the Spanish terms used in the NRS were unfamiliar to their children and, in some cases, unfamiliar to the staff as well. A second problem with the Spanish NRS is that the English and Spanish versions are scored differently in that English answers are acceptable on the Spanish version, but not vice versa. This presents a problem because bilingual children may know some things

in English and other things in Spanish. For example, a child might know the Spanish words for household items and the English words for numbers and math concepts. As an indication of this, one-third of Spanish-language NRS assessors found that on the Spanish version of the NRS many of their children responded correctly in English, but not in Spanish.

Members of the Technical Work Group and experts in bilingual testing have also questioned whether the Simon Says and Art Show components of the NRS can be used appropriately to track children's progress in English, as HSB intends. They express concerns that these components, designed simply as a screener to identify children who might have difficulty understanding English, do not provide useful information on the extent of English understood.

In addition to addressing concerns about the reliability and validity of the NRS directly, it is important that HSB's analyses and results are easy for other knowledgeable people to understand and use. Professional standards call for a technical manual addressing issues such as reliability and validity, as well as clearly specifying the intended uses and interpretations of the tests and cautioning against unintended misuses. According to all three of the independent experts who reviewed the technical aspects of the NRS at our request, the documentation of the reliability and validity of the NRS is not as well organized as would be desirable.⁹ They stated that given the importance of the validity of the NRS, a technical manual that brings all the evidence together in one place would be valuable. The expert reviewers reported that, in some cases, relevant material for evaluating the procedures and evidence to support the reliability and validity was provided, but was not organized in one place. For other areas, especially concerning the empirical work related to the Spanish version, documentation was not provided. For example, the information on the Spanish version of the test was limited to descriptions of procedures and summaries (e.g., "reliabilities were in the moderate to high range") and did not include documentation that would have made it possible for the reviewers to confirm the findings.

⁹See appendix I for a list of the expert reviewers and their affiliations.

HSB Acknowledges that NRS Alone Does Not Provide Range of Information and Context Needed for Making Accountability Decisions

The NRS by itself does not provide sufficient information to draw conclusions about the effects of Head Start grantees on children’s outcomes—information that would support use of the NRS for Head Start grantee accountability. The NRS does not measure all aspects of Head Start, but only a limited range of the areas on which Head Start focuses and which contribute to children’s school readiness. For example, the NRS does not include measures related to science, creative arts, approaches to learning, physical health and development, or social and emotional development, areas on which all Head Start programs are required to focus. Further, the cognitive areas included in the NRS are measured using a very narrow source of data that is not sufficient to evaluate the effects of Head Start grantees on the full range of child outcomes. For the area of literacy, the test measures how well children can identify letters, but not whether they can recognize rhymes or understand that letters make sounds—both aspects of “phonemic awareness,” which is believed to be an area critical for preventing reading difficulties. For the area of language development, the test measures how well children can identify pictures by name, but not grammar, usage, or expressive speech.

The Head Start Bureau has acknowledged the limited scope of the NRS and has expressly urged Head Start grantees to continue implementing their local assessments of the broader range of Head Start activities. The Associate Commissioner for the Head Start Bureau has stated that the Bureau does not intend to make decisions about grantees based solely on NRS data. Rather, the NRS information will be combined with comprehensive program level data collected on program designs and staff patterns; funded and actual enrollment; health, education, disability, and family services delivered; and demographic, social, and other trends.¹⁰ Many Technical Work Group Members have stated that this type of contextual information is necessary for the NRS to be a useful part of an overall program evaluation design.

In addition to measuring a limited range of the areas on which Head Start focuses, the NRS does not include all of the 4-year-old children who participate in Head Start. Most notably, children who speak neither English nor Spanish, about 4 percent of Head Start children otherwise eligible to participate in the NRS, are excluded from the NRS. Some

¹⁰See GAO, *Head Start: Comprehensive Approach to Identifying and Addressing Risks Could Help Prevent Grantee Financial Management Weaknesses*, [GAO-05-176](#) (Washington, D.C.: Feb. 28, 2005).

grantees do not have such children in their classrooms while others may include many such children. In addition, a number of children are excluded from the NRS due to prolonged absence and the scores of some children who do participate in the NRS are later excluded due to administrative reporting errors.

Application of NRS in Targeting Training and Technical Assistance Requires Further Development

NRS results are most reliable at the grantee level, but results at the grantee level are not the most useful for identifying where training and technical assistance should be targeted because some grantees include a large number of locations and classrooms. Using average scores at the grantee level to target training and technical assistance can mask the variability that underlies them. An average score gain for a grantee may be accounted for by high gains only of children in particular classrooms, while the scores of children in other classrooms did not change or actually lost points. The NRS data would allow for more effective targeting of training and technical assistance if the data could be used at the center and classroom levels, but currently the NRS cannot be used in this way. Given this limitation, HSB has stated that it might use NRS results to target training to a particular region of the country or to support a national training initiative in a particular skill area rather than to target specific grantees.

The NRS, by itself, cannot identify which particular aspects of the Head Start program, if any, contributed to a grantee's particular NRS results and this imposes some limitations on its utility for targeting training and technical assistance. The NRS does not directly assess the performance of Head Start grantees, such as by assessing the quality of the classroom environment or teacher-child interactions. Rather, the NRS assesses children's performance as an indirect measure of grantee performance. To ensure that the NRS can be used as a valid indicator of grantee performance (vs. variations in student age or other characteristics), experts believe it would be important to link NRS data to other observations known to distinguish more and less successful programs. In its quality assurance study of the NRS, HSB found that local Head Start staff were not sure how to use the fall 2003 results that were reported at the grantee level. Likewise, in our survey of NRS assessors we found that almost one-third of assessors believed the NRS did not provide useful information for their programs.

Some members of the Technical Work Group have suggested that HSB further investigate the assumption that targeting training and technical assistance at the grantee or broader level can affect the progress made by

children on certain academic skills. They argue that, if it is found that the classroom level matters, then the focus of analysis and reporting should be redirected and efforts could be made to increase the reliability of the scores at the classroom level.

Conclusions

The NRS is an important step toward meeting a long-standing need for systematic data on children's progress in Head Start and grantees' performance. Developing such a system is a challenging endeavor and considerable care and resources have gone into the project so far. At the same time, the technical standards applicable to HSB's planned uses for the assessment results need to be met. In addition, the system should be implemented with the greatest efficiency and caution against unintended negative consequences. The current NRS has strengths as well as areas in need of refinement, further investigation, and development.

While the NRS provides some information on child outcomes among Head Start grantees, HSB has not yet articulated how it intends to interpret and use this information for the purposes of informing decisions about Head Start accountability and targeting training and technical assistance. Without further guidance, there is confusion among Head Start grantees about what level of performance is expected of them and how NRS results from their programs might be used to hold them accountable. Out of anxiety about potential uses of the test, grantees may be inappropriately narrowing the educational activities provided through Head Start to match those included in the NRS, even though instructed not to do so. Thus far, HSB has not established an ongoing mechanism for monitoring the extent to which the NRS has such effects on instruction.

Other key steps that HSB has not taken include validating component tests and determining the reliability and validity of the NRS results across time. In addition, it has not compiled complete, well-organized documentation on the analyses conducted during test development and implementation, making it difficult for independent experts to evaluate the full technical merits of the English and Spanish versions of the NRS. Further, HSB lacks a mechanism for ensuring that all English and Spanish-speaking Head Start children who are eligible to participate in the NRS are assessed. Without such a mechanism and additional analyses, and the assurances they provide, the potential exists that the NRS will produce results that are not useful for program evaluation. Moreover, without further work on test validation, HSB cannot use the NRS for making decisions about grantees.

Finally, HSB's decision to assess all children with the full NRS assessment, rather than assessing a sample of children with a sample of items, has created a logistical challenge for many local Head Start grantees who must conduct the assessments, and limited the depth of information the NRS can provide about the learning of Head Start children in particular skill areas. At the same time, developing a sampling or matrix sampling strategy is complicated, especially for gathering information on the performance of subgroups of grantees, such as by region.

Recommendations for Executive Action

To help ensure that the NRS successfully and efficiently achieves its purposes, we are recommending that the HHS Assistant Secretary for ACF take steps to better monitor some aspects of NRS implementation and examine means of improving its efficiency, including steps to:

- monitor the effects of the NRS on local Head Start instructional practices;
- improve the management and accuracy of its data on the number of children eligible for and participating in the NRS; and
- work with the Technical Work Group to determine the feasibility of sampling options for administering the NRS, including documentation of their costs and benefits.

In addition, we are recommending that the Assistant Secretary for ACF reduce uncertainty about the appropriate uses of the NRS by taking additional steps to:

- determine how the NRS data will be used for the purposes of accountability and targeting training and technical assistance, and clearly communicate this information to grantees;
- use the first year of NRS results to conduct further study to ensure that the results are reliable and valid for both the English and Spanish versions and that the results are appropriate for the intended purposes; and
- compile detailed technical information on the NRS, including appropriate uses, in a single, well-organized document and make this information publicly available.

Agency Comments and Our Evaluation

ACF provided written comments on a draft of this report, which are reprinted in appendix III. ACF generally agreed with GAO's recommendations and stated that it had taken the following actions:

- ACF's contractors are conducting additional analyses of the first year NRS results to ensure that future results are reliable and valid.
- ACF's contractors are preparing a detailed technical report.
- ACF has engaged its contractors and TWG in the preparation of an options paper with recommendations for sampling.
- ACF is examining changes that occur in local curriculum implementation and teaching practices.

Further, ACF indicated that it will examine ways to improve the management and accuracy of its data on the number of children eligible for and participating in the NRS.

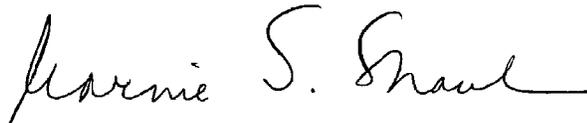
ACF's positions regarding the NRS evolved over the course of our review, as evidenced by ACF's decision not to include the 2003-2004 NRS results in the 2004-2005 program monitoring process, its modification of training materials, and changes ACF made to the CBRIS. ACF expressed in its comments a continued willingness to receive recommendations and advice.

While generally agreeing with our recommendations, ACF also submitted detailed comments on certain aspects of the draft report. Several of these comments concerned the level of evidence for the validity of the NRS. For example, ACF cited ongoing analyses of validity and noted that most of the tests in the NRS have been used in other studies. However, while further evidence of validity may be forthcoming, the data available at the time of our review did not fully document that the tests provide for valid inferences about program performance or children's progress from fall to spring. If the test is to be used as a measure of program performance or to assess changes in child outcomes, it is important to ensure that it is sensitive to the range of development typically demonstrated in Head Start. Based on our analysis and that of the TWG and independent experts, we continue to believe that further study is necessary to ensure that the NRS results are reliable and valid and that the results are appropriate for the intended purposes.

ACF also commented at length on our finding that, according to our survey of assessors, at least an estimated 18 percent of grantees "changed

instruction during the first year of NRS implementation to emphasize areas covered in the NRS.” ACF does not dispute that such changes were made, but suggests they may be appropriate, which we had noted in the draft report. In addition, ACF made a number of technical comments that we have incorporated as appropriate.

We are sending copies of this report to the Assistant Secretary for ACF, appropriate congressional committees, and other interested parties. We will also make copies available to others upon request. In addition, the report will be available at no charge on GAO’s Web site at <http://www.gao.gov>. Please contact me at (202) 512-7215 if you or your staff have any questions about this report. Other major contributors to this report are listed in appendix IV.



Marnie S. Shaul
Director, Education, Workforce
and Income Security Issues

Appendix I: Objectives, Scope and Methodology

We designed our study to examine (1) what information the National Reporting System (NRS) is designed to provide, (2) how the Head Start Bureau (HSB) has responded to implementation issues raised by the Head Start grantees and experts during the first year of NRS implementation, and what issues remain to be addressed, and (3) whether the NRS provides HSB with the quality of information it needs to meet its goals. We obtained information about these objectives through the following methods:

- Conducted in-person interviews with representatives from HSB, its contractors, and early childhood professional organizations.
- Reviewed documents chronicling the steps HSB took in developing and implementing the NRS and delineating the professionally accepted standards for test development.
- Conducted a mail survey of a nationally representative sample of Head Start grantees and delegates.
- Conducted in-person interviews with staff at 12 Head Start programs in 5 states.
- Conducted interviews with all of the members of the Technical Work Group.
- Contracted with individuals recommended by the National Academy of Sciences as experts in the areas of psychometrics and the educational testing of Spanish-speaking and bilingual children.

We conducted our work between May 2004 and February 2005 in accordance with generally accepted government auditing standards.

Interviews with Head Start Bureau and Relevant Parties

To obtain information on the steps HSB took in developing and implementing the NRS, we conducted in-person and/or telephone interviews with HSB and its contractors or subcontractors (Westat, Mathematica, and Xtria), using semi-structured interview protocols. A representative of HSB was present at each of the interviews with its contractors. We asked HSB officials' questions about the purpose of the NRS, reporting NRS results, revisions and updates to the NRS, reactions to NRS critics, and other related matters. We asked Westat staff questions regarding: (1) the validity, reliability, and other analyses of NRS results; (2) test development and revision; (3) test administration, scoring, and

reporting; (4) testing individuals of diverse linguistic backgrounds; and (5) testing individuals with disabilities. We asked Xtria staff about focus groups they conducted, Computer-Based Reporting System (CBRS) training, and the CBRS itself. We asked Mathematica staff about their Quality Assurance Study methodology and findings.

We interviewed representatives of the National Head Start Association (NHSA) to obtain information on what NHSA staff and their members learned from the first year of NRS implementation and to obtain their opinion on the extent to which the NRS comports with professional standards. We interviewed representatives of the National Association for the Education of Young Children (NAEYC) to learn how the NRS comports with their recommendations for assessing young children.

Review of Documents

To obtain information chronicling the steps HSB took in developing and implementing the NRS and information about the quality of the NRS results, we reviewed documents provided by HSB and its contractor. These documents included, for example, minutes from meetings with the Technical Work Group and others, minutes from focus groups, copies of informational memos to Head Start grantees on the implementation of the NRS, reports of results from field testing, and reports of fall 2003 NRS results.

To obtain information on the professionally accepted standards for test development, we reviewed the Standards for Educational and Psychological Testing, which is sponsored and published jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. That document provides the preeminent, universally accepted, guidance for the development and evaluation of high-quality, psychometrically robust assessment instruments.

Survey of NRS Lead Assessors

To obtain information on implementation issues raised by the Head Start grantees during the first year of NRS implementation, we drew a stratified random probability sample of 472 grantees or delegates from a study population of 1,820 grantees or delegates of Head Start Programs during the 2003-2004 school year. We selected our sample from six strata defined by the total number of Head Start tests administered and the number of Head Start tests administered in Spanish in the 2003-2004 school year. Ultimately, we received 376 completed questionnaires, for an overall response rate of 80 percent. The division of the population, the division of

the sample, and the division of the respondents across the six strata can be found in table 3. Each sampled grantee or delegate was subsequently weighted in the analysis to represent all the members of the population.

Table 3: Sample Disposition

Stratum number	Stratum description	Total population size	Total sample size	Number of respondents
1	At least 200 tests and at least 100 Spanish tests	180	125	98
2	Less than 200 tests and at least 100 Spanish tests	22	22	17
3	At least 200 tests and between 1 and 99 Spanish tests	327	90	80
4	Less than 200 tests and between 1 and 99 Spanish tests	575	98	77
5	At least 200 tests and no Spanish tests	171	48	39
6	Less than 200 tests and no Spanish tests	545	89	65
Total		1,820	472	376

Source: GAO.

We developed the survey questionnaire and pretested the content and format of this questionnaire five times with NRS lead assessors, either in-person or on the telephone. During these pretests, we asked the NRS assessors whether the questions were clear and unbiased and whether the terms contained in the questionnaire were accurate and precise. We made changes to the questionnaire based on the pretest results. Questionnaires were mailed to the sample of NRS lead assessors in August 2004 and follow-up calls were made to those assessors whose responses were not received within 2 weeks.

Because we followed a probability procedure based on random selections, our sample of delegates and grantees is only one of a large number of samples that we might have drawn. Because each sample could have provided different estimates, we express our confidence in the precision of our particular sample's results as 95 percent confidence intervals. These are intervals that would contain the actual population values for 95 percent of the samples we could have drawn. As a result, we are 95 percent confident that each of the confidence intervals in this report will include the true values in the study population. All percentage estimates

from our sample have margins of error (that is, widths of confidence intervals) of plus or minus 6 percentage points or less, at the 95 percent confidence level, unless otherwise noted.

In addition to sampling errors, the practical difficulties of conducting any survey may introduce other types of errors, commonly referred to as non-sampling errors. For example, differences in how a question is interpreted, the sources of information available to respondents, or the characteristics of people who do not respond can introduce unwanted variability into the survey results. We included steps in both the data collection and data analysis stage to minimize such non-sampling errors. For example, a survey specialist in combination with subject matter experts designed our questionnaire; the questionnaire was pretested with NRS assessors; data entry was verified to ensure accuracy; and another computer programmer verified the computer programs used for analysis.

A copy of the survey questionnaire, including overall responses, is included in appendix II.

Site Visits to Head Start Grantees

To obtain information on implementation issues raised by the Head Start grantees during the first year of NRS implementation, we also conducted site visits to 12 Head Start programs in 5 states (Colorado, Maryland, Massachusetts, Rhode Island, and Virginia), where we interviewed staff who conducted the assessments and, in some cases, observed them administering the NRS to children. The states and grantees chosen for site visits were judgmentally selected to include a range of enrollment sizes, types of program, rural and urban locations, and ethnic and racial populations.

The interviews were conducted using a semistructured interview guide that included questions about preparation for and logistics of administering the assessment; experiences of conducting the assessments; effects of the NRS on the children and program; reactions to the NRS results; use of the CBRs; other assessment measures in use at the program; and contextual information about the program and community. During our site visits, we spoke with the lead assessor and, in some cases, other Head Start staff, including other assessors, staff, and managers. With the exception of sites in Colorado, we conducted our site visits during May and June of 2004. We conducted our Colorado site visits during September 2004. In all cases, we asked the staff to refer to experiences during the 2003-2004 school year. We cannot generalize our site visit findings beyond

the 12 sites we visited, but we have used these data for illustrative purposes in conjunction with our survey.

Interviews with Technical Work Group

To obtain information on whether the NRS provides HSB with the quality of information it needs to meet its goals, we conducted telephone interviews with each of the 16 members of the Technical Work Group, using a semi-structured interview protocol. We asked the members about their professional backgrounds and involvement on the Technical Work Group; their understandings of the purpose of the NRS; their assessments of the completeness of the steps HSB took in developing and implementing the NRS; their assessments of the extent to which the NRS is reliable, valid, and consistent with professional standards; specific concerns about the NRS that members had raised during Technical Work Group meetings; and their opinions on how HSB should proceed with regard to the NRS. Each of the members stated that he or she could be candid in discussing these issues with GAO. We also observed two meetings of the Technical Work Group in May and October 2004.

Technical Work Group Members

Craig Ramey, Ph.D., Chairman
Distinguished Professor of Health Studies and
Director, Georgetown University Center for Health Education
School of Nursing and Health Studies
Georgetown University
Washington, D.C.

Clancy Blair, Ph.D., Co-Chairman
Assistant Professor
Human Development and Family Studies
Pennsylvania State University
University Park, Pa.

Jason L. Anthony, Ph.D., Ed.S.
Research Assistant Professor
Texas Institute for Measurement, Evaluation, and Statistics
Department of Psychology
University of Houston
Houston, Tex.

Margaret Burchinal, Ph.D.
Senior Scientist
Frank Porter Graham Child Development Institute

The University of North Carolina at Chapel Hill
Chapel Hill, N.C.

Richard Clifford, Ph.D.
Senior Scientist
Frank Porter Graham Child Development Institute
The University of North Carolina at Chapel Hill
Chapel Hill, N.C.

Linda Espinosa, Ph.D.
Associate Professor
311D Townsend Hall
College of Education
University of Missouri-Columbia
Columbia, Mo.

Nicholas Ialongo, Ph.D.
Associate Professor
Bloomberg School of Public Health
Johns Hopkins University
Baltimore, Md.

Graciela Italiano-Thomas, Ed.D.
CEO
Centro de la Familia de Utah
South Salt Lake, Utah

Jacqueline Jones, Ph.D.
Director, Initiatives in Early Childhood and Literacy Education
Educational Testing Service
Princeton, N.J.

Ann P. Kaiser, Ph.D.
Professor of Psychology and Human Development
Director, Research Program on Communication, Cognitive, and Emotional
Development
Vanderbilt University
Nashville, Tenn.

Samuel J. Meisels, Ed.D.
President
Erikson Institute
Chicago, Ill.

Fred Morrison, Ph.D.
Professor
Department of Psychology
University of Michigan
Ann Arbor, Mich.

Robert C. Pianta, Ph.D.
Professor, William Clay Parrish, Jr. Chair in Education
Curry Programs in Clinical and School Psychology
University of Virginia
Charlottesville, Va.

Kyle Snow, Ph.D.
National Institute of Child Health and Human Development
National Institutes of Health
U.S. Department of Health and Human Services
Bethesda, Md.

W. Douglas Tynan, Ph.D., ABPP
Associate Professor of Pediatrics
Alfred I. duPont Hospital for Children
Jefferson Medical College
Wilmington, Del.

Jane Wiechel, Ph.D.
Associate Superintendent
Center for Students, Families and Communities
Ohio Department of Education
Columbus, Ohio

Expert Reviews

To obtain information on whether the NRS provides HSB with the quality of information it needs to meet its goals, we contracted with individuals recommended by the National Academy of Sciences (NAS) as experts in the areas of psychometrics and the educational testing of Spanish-speaking and bilingual children. These independent experts reviewed documents provided by HSB and its contractors and provided written comments on the adequacy and appropriateness of the assessment. We also conducted follow-up telephone interviews with each of the three experts to reconcile variations in their written reviews. We developed our own conclusions based on the information provided by these experts. The three experts are listed below.

Ronald K. Hambleton, Ph.D.
Distinguished University Professor for Research and Evaluation Methods
University of Massachusetts at Amherst
School of Education
Center for Educational Assessment
Amherst, Mass.

Luis M. Laosa, Ph.D.
Principal Research Scientist, Emeritus
Educational Testing Service
Center for Education Policy and Research
Princeton, N.J.

Robert L. Linn, Ph.D.
Professor
University of Colorado
Department of Education
Boulder, Colo.

Appendix II: Survey Instrument

The survey instrument displayed here includes the population estimates for grantees overall. The confidence intervals for these estimates do not exceed plus or minus 6 percentage points.

1. Please characterize the predominant behavioral reaction of your Head Start program’s children to the following sections of the NRS, and the NRS overall, during the 2003-04 school year? *(Check one answer in each row.)*

NRS Section	Reactions of the children, such as . . .			Not sure or no basis to judge
	Eager, smiling, curious, or engaged	Compliant, but neither enthusiastic or unenthusiastic	Tired, distracted, or answers without looking at page	
Simon Says	69%	31%	.	
Art Show	56	41	2	
Peabody Picture Vocabulary Test	22	60	17	
Letter Naming	14	55	31	
Early Math	24	61	15	
The NRS Overall	27	68	5	

Please feel free to expand upon or explain any of your above answers.

2. Approximately how many of your Head Start children exhibited extremely negative behaviors during the NRS (such as crying, non-responsiveness, or refusal to complete the NRS assessment)? *(Check one answer.)*

20% None 68% Few 12% Some Many Not sure

3. Based on your knowledge of each child’s abilities, do you believe that for the majority of your Head Start children, the NRS accurately portrayed their abilities for each of the following NRS sections, and the NRS overall, in the 2003-04 school year? *(Check one answer in each row.)*

NRS Section	NRS accurately portrayed children’s abilities			If you answered “No” or “Not sure” for any NRS section, please briefly explain.
	Yes	No	Not sure	
Simon Says	92%	5%	2%	
Art Show	91	7	2	
Peabody Picture Vocabulary Test	45	45	10	
Letter Naming	56	35	8	
Early Math	57	32	11	
The NRS Overall	56	25	18	

4. Did you or your NRS assessors administer any NRS assessments in Spanish during the 2003-04 school year?
(Check one answer.)

60% Yes (Continue.)
40% No → Skip to question 6.

5. Would you disagree or agree with each of the following statements regarding your program's experience with the 2003-04 NRS? (Check one answer in each row.)

	Strongly disagree	Disagree	Agree as much as disagree	Agree	Strongly agree	No basis to judge
Our Head Start children who took both the Spanish and English versions of the NRS reacted less positively to the Spanish version.	11%	33%	17%	15%	6%	18%
The questions on the Spanish version of the NRS were culturally appropriate for our Spanish-speaking Head Start children.	10	27	21	23	3	16
On the Spanish version of the NRS, many of our Spanish-speaking Head Start children responded correctly in English, but not in Spanish.	4	20	26	26	10	14
The Spanish version of the NRS accurately portrayed the abilities of the Spanish-speaking children in our program.	8	26	29	24	3	11

Please feel free to expand upon or explain any of your above answers.

Appendix II: Survey Instrument

6. How much of a challenge, if any, did each of the following situations present for your Head Start program during the 2003-04 school year? *(Check one answer in each row.)*

Situation	Little or no challenge	Some challenge	Moderate challenge	Great challenge	Very great challenge	No basis to judge/Not applicable
Receiving materials (e.g., binders, response sheets, etc.) in a timely manner	31%	19%	14%	15%	21%	.
Condition of materials (e.g., binders, response sheets, etc.)	81	9	5	3	.	1
Allocating staff to administer the NRS assessments	35	22	18	13	11	1
Finding Spanish-speaking assessors	32	8	11	10	10	29
Finding space to administer the NRS assessments	35	26	22	12	6	.
Finding time to assess all children	16	22	24	21	18	.
Obtaining parental consent	73	12	7	1	1	7
Setting up the Computer-Based Reporting System (CBRS) or entering data	39	27	17	8	5	3

7. During the 2003-04 school year, did you make any of the following changes to your program to fulfill NRS assessment procedural requirements? *(Check one answer in each row.)*

Program changes	Yes	No	Do not know
We pulled teachers, administrators, or other staff away from their classes or daily jobs to administer the NRS or enter CBRS data.	87%	13%	.
We asked teachers, administrators, or other staff to work overtime in order to administer the NRS or enter CBRS data.	26	74	.
We employed additional staff to help administer the NRS or enter CBRS data.	21	79	.
We delayed other tests or assessments in order to administer the NRS or enter CBRS data.	23	77	.
We adjusted the timing of our normal curriculum to accommodate the NRS.	46	53	1

Appendix II: Survey Instrument

8. Would you disagree or agree with each of the following statements regarding your program's experience with the 2003-04 NRS? *(Check one answer in each row.)*

	Strongly disagree	Disagree	Agree as much as disagree	Agree	Strongly agree	No basis to judge
Children may know more than is actually reflected by the NRS.	1%	2%	13%	35%	47%	1%
The NRS provides useful information for our program.	12	18	23	36	8	3
Many of the skills assessed in the NRS are also included in some of our program's other assessment(s) of children.	1	6	10	51	32	1
Instruction in our program has changed to emphasize the areas covered in the NRS.	24	38	19	16	2	2
The NRS assessment of many Spanish-speaking children in both English and Spanish is overly time-consuming for Spanish-speaking children in our program.	.	10	10	16	25	38
The NRS gives Spanish-speaking children the opportunity to show their abilities in both languages.	3	10	18	27	6	36
The purpose of the NRS has been adequately explained to our program.	6	7	13	51	21	.
The script that NRS assessors follow provides enough flexibility to accommodate the needs of individual children.	14	26	18	37	5	1
The financial costs of administering the NRS in our program are adequately covered by funds from the federal Head Start Bureau.	24	21	12	26	5	12
The NRS accurately measures the progress of our Head Start children from Fall to Spring.	17	20	22	25	3	13
The efforts required of programs to implement the NRS will be worthwhile in the long run to Head Start.	17	16	28	23	6	10

9. If there is anything else you would like to share with us about your experiences with the NRS during the first year of implementation, including ways it might be improved, please do so below.
(You may attach additional sheets.)

**Thank you for your time and assistance.
 Please return your completed questionnaire in the envelope provided.**

Appendix III: Comments from the Department of Health and Human Services



DEPARTMENT OF HEALTH AND HUMAN SERVICES

ADMINISTRATION FOR CHILDREN AND FAMILIES
Office of the Assistant Secretary, Suite 600
370 L'Enfant Promenade, S.W.
Washington, D.C. 20447

APR 20 2005

Ms. Marnie S. Shaul
Director, Education, Workforce
and Income Security Issues
U.S. Government Accountability Office
441 G. Street, N.W.
Washington, D.C. 20548

Dear Ms. Shaul:

The Administration for Children and Families appreciates the opportunity to provide comments on recommendations in the U.S. Government Accountability Office's draft report entitled, "Head Start: Further Development Could Allow Results of New Test to be Used for Decisionmaking" (GAO-05-343).

Should you have questions regarding our comments, please contact Windy Hill, Associate Commissioner of the Head Start Bureau, Administration on Children, Youth and Families, at (202) 205-8573.

Sincerely,

A handwritten signature in black ink, appearing to read "Wade F. Horn".

Wade F. Horn, Ph.D.
Assistant Secretary
for Children and Families

Attachment

**COMMENTS OF THE ADMINISTRATION FOR CHILDREN AND FAMILIES ON
THE GOVERNMENT ACCOUNTABILITY OFFICE'S DRAFT REPORT TITLED,
"HEAD START: FURTHER DEVELOPMENT COULD ALLOW RESULTS OF NEW
TEST TO BE USED FOR DECISIONMAKING" (GAO-05-343)**

The Administration for Children and Families (ACF) appreciates the opportunity to comment on this Government Accountability Office (GAO) draft report. We appreciate the breadth of contact made in the preparation of this report.

GAO Recommendations

To help ensure that the NRS successfully and efficiently achieves its purposes, we are recommending that the HHS Assistant Secretary for ACF take steps to better monitor some aspects of NRS implementation and examine means of improving its efficiency, including steps to:

- monitor the effects of the NRS on local Head Start instructional practices;
- improve the management and accuracy of its data on the number of children eligible for and participating in the NRS; and
- work with the Technical Work Group to determine the feasibility of sampling options for administering the NRS, including documentation of their costs and benefits.

In addition, we are recommending that the Assistant Secretary for ACF reduce uncertainty about the appropriate uses of the NRS by taking additional steps to:

- determine how the NRS data will be used for the purposes of accountability and targeting training and technical assistance, and clearly communicate this information to grantees;
- use the first year of NRS results to conduct further study to ensure that the results are reliable and valid for both the English and Spanish versions and that the results are appropriate for the intended purposes; and
- compile detailed technical information on the NRS, including appropriate uses, in a single, well-organized document and make this information publicly available.

ACF Comments

ACF has widely publicized its commitment, need and intent for improvements in the implementation of the National Reporting System (NRS), including child assessment. We believe that the GAO recommendations mirror many of ACF's public statements, as well as accurately describe some of the action steps that are already in process.

The remaining GAO recommendations are also in keeping with those arising from our internal

planning with the NRS contractors, the local programs and the Technical Work Group (TWG). Additionally, the Secretary of HHS will also be receiving recommendations from the newly formed Secretary's Advisory Committee (SAC) on Head Start Accountability and Educational Performance Standards, which will begin meeting this summer.

Specific comments related to the recommendations:

- ACF has already included a scheduled deliverable within the scope of work of the NRS contractors. Additional analyses are continuing to be conducted with the first year NRS results in order to ensure that future results are reliable and valid, and in order to be confident that the results are appropriate for the interim and final intended purposes. TWG and SAC will both assist ACF in the review of these analyses.
- ACF has included tasks that will result in the NRS contractors preparing a detailed technical report to expand beyond what is already included in the recently distributed "Report to Congress on Head Start Assessment." The new work is already in progress. We will make some version of the new document available to the public when it is cleared by ACF.
- ACF will examine ways to improve management regarding NRS participation. We believe that we can achieve this through the existing Computer-Based Reporting System data collection, data management, the quality assurance site visits, and as part of our overall responsibility for program monitoring.
- Prior to the release of the GAO report, ACF had engaged the NRS contractors and TWG in the preparation of an options paper with recommendations for sampling, including not only the benefits and cost implications for each approach but also what could or must be "given up" under the implementation of each approach. TWG and SAC will have a role in reviewing these recommendations and further advising ACF and HHS, respectively.
- ACF is examining and will continue to examine changes that occur in local curriculum implementation and teaching practices through at least three primary methods: on-site federal reviews, regular periodic contact of an assigned technical assistance liaison and the NRS quality assurance site visits.

Other Comments

- ACF would like the title as well as pertinent references throughout the document to refer to the NRS rather than "the test." The child assessment alone is not synonymous with NRS.

**Appendix III: Comments from the Department
of Health and Human Services**

- Though mentioned, ACF believes that the Year One Quality Assurance Study lacked attention in this report.
- Page 4, first full paragraph, and page 23, third paragraph - GAO states that HSB has asserted the validity and reliability of NRS measures because NRS borrows certain materials from existing tests that have met the validity and reliability criteria, but the agency has not shown NRS itself to be valid or reliable over time. Reliability and concurrent and predictive validity of the Head Start NRS measures were calculated using the Family and Child Experiences Survey (FACES) and other data on Head Start children. These results were included in the package of materials provided to GAO.
- Ongoing analyses are being conducted to further demonstrate the reliability and validity of the NRS assessment data. For example, analyses comparing matched FACES data with NRS data are being conducted to validate the assessment parallel data collected by locally trained NRS assessors with those collected by trained, experienced, professional FACES data collectors. Preliminary analyses indicate that little difference is found between the two data sets.
- Most of the subtests in the NRS battery have been used extensively in the Head Start FACES study, in the National Head Start Impact Study or in the Head Start Quality Research Center intervention studies involving more than 10,000 Head Start children, as well as in other major studies of low-income preschoolers. These measures have been used in the National Institute of Child Health and Human Development studies, the “Mother & Child Supplement” to the National Longitudinal Survey of Youth” and in the “Child Development Supplement” to the Panel Study of Income Dynamics. The results of these assessments have proved to be highly stable from cohort to cohort, not only in terms of the level of achievement with which children enter or leave the Head Start program, but also in terms of their growth trajectories.
- Analysis of longitudinal data from the Head Start FACES study has shown that vocabulary and letter-recognition assessments given in Head Start can account for nearly half of the variance in children’s tested reading skills at the end of kindergarten, and 66 percent of the variance when tested in general knowledge at the end of kindergarten. Also, scores gained from vocabulary and letter-recognition assessments account for almost one-third of the variance in kindergarten reading skills and over one-quarter of the variance in kindergarten general knowledge.
- Page 9, Figure 2 - ACF would like to see the report contain both a narrative and a timeline on NRS for the year 2004, not just for 2002 and 2003 as is currently in the report. The activities of the GAO occurred during 2004, as did the first full year of ACF’s implementation of NRS.
- Page 11, first paragraph - GAO indicates that a true “pilot,” rather than the summer field test of NRS, would take about a year to complete. ACF believes that by further

examination of the Year I data, we will have data even beyond the scope of a one-year pilot effort.

The GAO report also states that HSB did not conduct a “full pilot test.” The Head Start Bureau (HSB) conducted a field test of the NRS child assessment in the spring of 2003 with a national probability sample of 36 Head Start programs, including two migrant programs and two American Indian programs, resulting in a field test sample of over 1,430 kindergarten-eligible English- and Spanish-speaking children. The results of the field test showed that the measures were appropriate for the Head Start population, capturing a range of ability levels in the assessments domains. Year I implementation results will add significantly to this information and what we know about the properties of the assessment over time.

- Page 21, first paragraph - Though GAO has included a footnote to explain, “...actions taken by the Head Start Bureau’s contractors are attributed to the Head Start Bureau itself,” this note appears on this page long after readers can attribute actions to HSB. Since the report is written without disclosing what actions were taken or advised by whom, ACF would like the footnote to be moved to the beginning of the report or described in the opening narrative.
- Page 26, third paragraph - GAO uses a figure of 13 percent to describe the number of children who speak neither English nor Spanish. Aggregate Program Information Report data indicate that programs reported 95 percent of the children enrolled last year spoke either English or Spanish, leaving 5 percent who speak other languages. The number of children in NRS who spoke a language other than English or Spanish at home, as reported in the Computer-Based Reporting System, was approximately 4 percent or 19,000 in the fall of 2003.

HSB has two other concerns with the report. Our responses to these two are rather lengthy to help clarify them:

1. Page 17, first paragraph - The program office is concerned with the following statement in the report “... some grantees have changed instruction to emphasize areas covered in the test.” The manner in which it is stated implies that this can only be negative and that it can only be attributable to NRS in any program in which it occurs. On the contrary, we believe this illustrates a powerful positive change, inasmuch as Head Start’s heavy emphasis on instructional and curricular changes pre-date the implementation of NRS by several years. We explain our concern in detail.

As this country’s largest and only federally funded, comprehensive early childhood program, we have learned a great deal from research-based practices that enhance young children’s learning and development. Unless we ensure that programs are providing meaningful and challenging learning experiences through ongoing observation and assessment of children’s progress as required by the Program Performance Standards, participation will have little value for children. Therefore, we are not surprised to learn that local programs reported to GAO that they are making changes in their curriculum and in their teaching practices. We believe that NRS may be

giving them additional data upon which they are making such local decisions, rather than NRS serving as the sole source of such information upon which to base change decisions. We have, through various methods, specifically cautioned programs not to take actions of this nature. We believe that most programs are not using NRS Year I reporting in inappropriate ways.

The GAO report acknowledges in a small way that prior work has occurred in this area, yet GAO does not acknowledge that the prior work, rather than NRS alone, may be producing changes in curriculum and instruction. Prior to the NRS, the Head Start Child Outcomes Framework (Framework) defined the comprehensive nature of child development and early childhood education in Head Start by including the domains of: language development, literacy, mathematics, science, creative arts, social and emotional development, approaches to learning, and physical development. This focus across all domains must remain within the local curriculum and within the local ongoing assessment.

Additionally, the Head Start Program Performance Standards require that all of these areas of development be supported through age-appropriate curriculum delivered through classroom or home-based programming with the integral involvement of parents. Therefore, the focus across all domains must remain within the local curriculum and within the local ongoing assessment.

ACF has been offering and continues to offer training, technical assistance and other resources to help programs look more closely at their local implementation and to make necessary changes. Additionally, some programs have made and others are actively engaged in making these types of changes as a result of either their required program self-assessment or local aggregation of child outcome data, and/or as a result of noncompliance or deficiencies identified and reported in the process of triennial monitoring. We recognize and applaud programs that are actively engaged in making appropriate changes in the areas of curriculum, ongoing assessment of child progress and early childhood instruction across domains.

Another example of our work that is influencing changes in local programs is the Head Start Leaders Guide to Positive Child Outcomes. This resource is based on the requirements of the Head Start Program Performance Standards and the Framework. This important document has been the basis of Head Start training, providing staff with specific strategies to strengthen curriculum and to foster children's progress in each of the identified domains. These strategies assist program staff in strengthening curriculum planning and implementation regardless of the specific curriculum used in individual programs.

Both ACF's regulations and resource materials provide examples of educational quality based on:

- intentional teaching;
- outcomes-oriented learning experiences;
- child engagement; and
- challenging learning opportunities for small groups of children and for individual children.

2. Page 7, second paragraph - The GAO report states of non-NRS assessments, "The assessments occur 3 times each year and generally involve observing the children during normal classroom activities." This statement, though perhaps stated by one or more local programs, inaccurately describes grantee actions as related to two existing Head Start requirements. The first is the long-standing requirement for ongoing observations and ongoing assessment of each child's progress. Therefore, observing or assessing progress only three times a year would be a significant area of noncompliance, and more likely, a deficiency in that program. The statement on page seven further represents a misunderstanding and, therefore, an inappropriate implementation of the existing requirement. Three times per year each agency is required to aggregate, report and examine data from its locally designed and locally administered ongoing assessment of child progress. This is different from assessing children three times a year.

Head Start standards do not allow for "assessing three times per year"; rather, teachers must observe and record examples of children's development and learning on an ongoing basis throughout the year. Management requirements have programs review aggregate data from the assessment at three points in time during the year--the beginning, midpoint and the end. The information is reviewed program-wide, in aggregate, to assess children's status and progress on a wide range of areas identified in the Framework. This information is used to continue to plan the educational program for children as well as to inform the overall program assessment and planning process.

We are aware that NRS is providing an additional way for programs to look at children's progress over the course of a Head Start year. This may be contributing to a renewed focus on becoming more intentional and more deliberate regarding the early childhood educational services in local Head Start programs--the learning content, intentional teaching, and children's school readiness in the areas of both the Framework and the 1998 Congressionally mandated child outcomes.

As we look more closely at this type of change in local programs, we hope that we will be able to conclude that NRS is not currently the "cause" of the more intentional focus on school readiness, but rather that necessary changes are the result of a number of other factors, including:

- The 1998 Congressional mandate, specifying additional Program Performance Standards in language, literacy and numeracy/early mathematics and the subsequent Framework;
- The increased qualifications of teachers and the significant number with degrees;
- The increased focus on intentional teaching strategies shared through training based on research;
- The appropriate use of local outcomes data (not the NRS data);
- The appropriate use of the required program self-assessment;
- Information from research, including the finding that children's pre-school vocabulary is the best predictor of school success, and
- Individual agency and grantee responses to findings from federal, on-site and triennial monitoring of compliance with all applicable laws and regulations.

HSB's emphasis on instructional change clearly pre-dates NRS, which was launched in 2002.

**Appendix III: Comments from the Department
of Health and Human Services**

As stated earlier, separate from and prior to NRS, the Framework defined the comprehensive nature of child development and early childhood education in Head Start. Additionally, the Head Start Program Performance Standards require that areas of development be supported through age-appropriate curriculum delivered through classroom or home-based programming with the integral involvement of parents.

It is important to recognize that both the Head Start Program Performance Standards, which were initially issued in 1972 and revised in 1996, and the Framework issued in 2000, all pre-date NRS.

The 1998 reauthorization of the Head Start Act (The Act) requires the Secretary of HHS to establish "education performance standards to ensure the school readiness of children participating in Head Start," including assurances that children develop phonemic, print and numeracy/early mathematics skills; understand and use language to communicate, understand and use increasingly complex and varied vocabulary; develop and demonstrate an appreciation of books; and for English language learners, progress toward acquisition of the English language. The Act also required that the Head Start teacher qualifications be raised because of evidence that links classroom and teaching quality to the skills, knowledge and formal education of teachers.

Therefore, the Act, the Head Start Leaders Guide to Positive Child Outcomes, the Framework and the Program Performance Standards, as well as professional development experiences such as Mentor Coaching, all hold programs and local staff accountable for use of specific strategies to strengthen curriculum content, learning outcomes and intentional teaching, and to foster children's progress in each child development domain of the comprehensive Head Start program.

Ensuring developmentally appropriate programming provides a meaningful basis for observing and assessing children's progress and promoting and individualizing learning and development. NRS is providing an additional form of assessment reporting and an additional and renewed focus on local programs becoming more intentional and more deliberate regarding curriculum content, intentional teaching and children's school readiness, and is not the sole source or a source to replace existing requirements for local Head Start agencies.

ACF looks forward to additional recommendations as we move toward the use of NRS data and as we inform grantees and others about the use of the NRS data as another tool for accountability and providing training and technical assistance.

Appendix IV: GAO Contacts and Staff Acknowledgments

GAO Contacts

Betty Ward-Zukerman (202) 512-2732, wardzukermanb@gao.gov
Heather McCallum Hahn (202) 512-2890, mccallumh@gao.gov

Staff Acknowledgments

Ramona Burton, Scott Heacock, Kathryn Rooney, Carolyn Boyce, Curtis Groves, Stu Kaufman, Joan Vogel, and Sid Schwartz made significant contributions to this report.

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are \$2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548