

Contract No.: 233-02-0086 (03)  
MPR Reference No.: 6094-710

**MATHEMATICA**  
Policy Research, Inc.

**Implementation of the  
Head Start National  
Reporting System:  
Spring 2006**

*Final Report*

*February 2008*

*Cheri A. Vogel  
Renée Nogales  
Nikki Aikens  
Louisa Tarullo*

Submitted to:

Office of Planning, Research, and Evaluation  
Administration for Children and Families  
U.S. Department of Health and Human Services  
370 L'Enfant Promenade  
Washington, DC 20447

Project Officer:

Jennifer Brooks

Submitted by:

Mathematica Policy Research, Inc.  
P.O. Box 2393  
Princeton, NJ 08543-2393  
Telephone: (609) 799-3535  
Facsimile: (609) 799-0005

Project Director:

Louisa Tarullo

**This page has been intentionally left blank for double-sided copying.**

## ACKNOWLEDGMENTS

---

**W**e would like to thank the many people who contributed to the successful completion of this study and report. Jennifer Brooks, Administration for Children and Families, provided invaluable leadership and vision to shape this effort. However, this report would have been impossible without the support and participation of the 35 Head Start program directors and staff who so generously spent their time to talk with us.

Numerous staff at MPR also played critical roles in the research effort. Susan Sprachman led the training of our talented group of observers/site visitors. The site visitors included Barbara Schiff Kennen, Brian Roff, David Eden, Betsy Santos, Charlotte Hansen, Emily Sama Martin, Nii Addy, Rob Buschmann, Margo Salem, and Mina Dadgar. Barbara Carlson, our sampling statistician, assisted by Yuhong Zheng, randomly selected programs, centers and children for site visits and assessment observation. Barbara Schiff Kennen and Jennifer McGovern contacted all programs to ensure their participation. Patricia del Grosso coded the site visit summaries, and managed the project finances. Megan Hague Angus assisted with data analysis. William Garrett patiently formatted the document, assisted by Alfreda Holmes. Barbara Geehan was our editor, with help from Daryl Hall. Diane Paulsell, who originally conceptualized this implementation study and led its previous reports, provided valuable insights during the review process. Sally Atkins-Burnett served as our quality assurance reviewer. We would also like to thank John Love, the original project director, for his ongoing support.

We also extend our thanks to staff from Juárez and Associates, Rita Rico and Jatzin Alvarado, who were exceptional site visitors.

The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services.

**This page has been intentionally left blank for double-sided copying.**

# CONTENTS

---

Chapter	Page
<b>ACKNOWLEDGMENTS</b> .....	<b>iii</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>xv</b>
<b>I INTRODUCTION</b> .....	<b>1</b>
DEVELOPMENT OF THE HEAD START NATIONAL REPORTING SYSTEM.....	2
The NRS Assessment.....	3
The Computer-Based Reporting System.....	4
NRS QUALITY ASSURANCE AND SYSTEM DEVELOPMENT PROJECT .....	5
SITE VISITS .....	6
Sample Selection .....	6
The Sample of Head Start Programs and Children Observed.....	9
Developing Site Visit Protocols.....	10
Qualifications and Training of Site Visitors.....	12
Site Visit Activities.....	13
ANALYTIC METHODS.....	13
ROADMAP TO THE REPORT .....	14
<b>II ADMINISTERING THE CHILD ASSESSMENT</b> .....	<b>17</b>
APPROACH TO EVALUATING THE QUALITY OF THE NRS ASSESSMENTS.....	18

Chapter	Page
II <i>(continued)</i>	
MEETING THE CERTIFICATION STANDARD.....	19
INTER-RATER RELIABILITY .....	22
ERRORS IN PROCEDURES, ADMINISTRATION, AND SCORING .....	23
Administering the Set-Up and Warm-Up Sections.....	25
Administering and Scoring the PreLAS Simon Says .....	25
Administering and Scoring the PreLAS Art Show .....	26
Administering and Scoring the Peabody Picture Vocabulary Test (PPVT-III, Adapted) .....	27
Administering and Scoring the Letter Naming Task.....	28
Administering and Scoring Early Math Skills.....	29
IMPLEMENTATION OF THE SPANISH-LANGUAGE VERSION .....	30
Meeting the Certification Standard .....	31
Errors in Procedures, Administration, and Scoring.....	32
ASSESSORS' EXPERIENCES ADMINISTERING THE CHILD ASSESSMENTS.....	32
Children's Responses to the Assessment Process.....	34
Experiences of Assessors Who Were Also Children's Teachers.....	39
Experiences Assessing Children with Limited English Language Skills.....	41
Experiences Assessing Children with Disabilities.....	43
SUMMARY.....	45
<b>III LOCAL APPROACHES TO TRAINING NRS ASSESSORS.....</b>	<b>49</b>
EXPERIENCES WITH THE SUMMER 2005 REGIONAL "TRAINING-OF- TRAINERS" CONFERENCE .....	50
APPROACHES TO ASSESSOR TRAINING IN SPRING 2006.....	52
Approaches to Refresher Training (English).....	52
Approaches to Refresher Training (Spanish).....	55
Approaches to Training New Assessors .....	56
LOCAL PROGRAMS' RESPONSES TO TRAINING RESOURCES AND MATERIALS.....	57
Satellite Broadcasts and Webcasts .....	58

<b>Chapter</b>	<b>Page</b>
III ( <i>continued</i> )	
Delivery of Training Materials .....	59
Spring 2006 Training Video .....	59
Assessor's Guide .....	61
Technical Assistance.....	62
SUMMARY.....	62
<b>IV    LOCAL APPROACHES TO IMPLEMENTING THE NRS .....</b>	<b>65</b>
COORDINATING THE CHILD ASSESSMENTS .....	65
Staffing for NRS Training, Coordination, and Oversight.....	66
Locations Where Assessments Were Conducted.....	66
Scheduling the Assessments.....	67
Tracking Progress in Completing the Assessments.....	68
APPROACHES TO ASSIGNING STAFF TO ADMINISTER THE ASSESSMENTS.....	69
Programs' Rationales for Deciding Whether or Not to Assign Teachers to Conduct Assessments.....	70
COMMUNICATING WITH PARENTS AND POLICY COUNCILS ABOUT THE NRS .....	71
Concerns Expressed by Parents and Policy Councils .....	71
Parent Refusals .....	72
COSTS OF IMPLEMENTING THE NRS .....	73
USING THE COMPUTER-BASED REPORTING SYSTEM.....	74
Assigning and Training Staff to Enter the Data.....	74
Access to Technology for Using the CBRS .....	75
Experiences Using the Enhanced CBRS Features Introduced in Year 2.....	75
Accessing Technical Assistance .....	76
Suggestions for Improving the CBRS.....	76
SUMMARY.....	77

Chapter	Page
<b>V USING THE NRS FOR LOCAL PROGRAM IMPROVEMENT EFFORTS .....</b>	<b>79</b>
REACTIONS OF LOCAL PROGRAM STAFF TO THE 2004–2005 GROWTH REPORT .....	80
Program Perspectives on Growth Score Results .....	81
Usefulness of the NRS Reports for Local Programs .....	81
Recommendations for Improving the NRS Report .....	82
HOW PROGRAMS HAVE USED THE GROWTH REPORT .....	84
Sharing NRS Reports with Local Program Staff and Stakeholders.....	84
Making Changes in Staff Development.....	85
Tracking Results for Individual Children .....	85
AVAILABILITY AND USE OF LOCAL ASSESSMENT RESULTS .....	86
Comparing NRS Reports with Other Assessment Results.....	88
CHANGES MADE TO CLASSROOM PRACTICES.....	89
FUTURE PLANS FOR USING THE NRS REPORTS.....	90
SUMMARY.....	91
<b>VI PERSPECTIVES OF LOCAL HEAD START STAFF ON THE NATIONAL REPORTING SYSTEM.....</b>	<b>93</b>
CONTRIBUTIONS OF THE NRS .....	93
CONCERNS ABOUT THE NATIONAL REPORTING SYSTEM .....	94
Purpose of the NRS and Use of Assessment Results .....	95
Whether NRS Results Will Accurately Portray Program Performance .....	96
Concerns Specific to the Spanish Assessment .....	98
Staff Time and Resources Dedicated to the NRS.....	100
SUGGESTIONS FOR IMPROVING THE NRS.....	100
Communication and Planning .....	101
Training and Guidance on Administering the Assessment .....	102
Expanding the Assessment Battery.....	103
Improving the Spanish-Language Version.....	103
Improving Assessment Procedures.....	104



<b>Chapter</b>	<b>Page</b>
VI ( <i>continued</i> )	
SUMMARY.....	104
<b>VII   IMPLICATIONS FOR SYSTEM IMPROVEMENT.....</b>	<b>107</b>
COMMUNICATION .....	107
ACCESS TO AND USEFULNESS OF ASSESSMENT RESULTS FOR LOCAL PROGRAMS.....	109
SUPPORT FOR ADMINISTERING THE ASSESSMENT .....	111
GUIDANCE ON ASSESSING CHILDREN IN SPANISH.....	112
GUIDANCE ON ASSESSING CHILDREN WITH DISABILITIES .....	113
CHANGES TO THE ASSESSMENT BATTERY .....	114
SUMMARY.....	114
<b>REFERENCES .....</b>	<b>115</b>
APPENDIX A: NRS ASSESSMENT OBSERVATION FORMS .....	A.1
APPENDIX B: SPRING 2006 NRS INTERVIEW AND FOCUS GROUP GUIDES.....	B.1
APPENDIX C: CHAPTER II SUPPLEMENTAL TABLES: WEIGHTED QUESTION BY QUESTION ASSESSMENT ERRORS FOR ENGLISH AND SPANISH ASSESSMENTS.....	C.1
APPENDIX D: HEAD START NATIONAL REPORTING SYSTEM 2004–2005 PROGRAM REPORT.....	D.1

**This page has been intentionally left blank for double-sided copying.**

## T A B L E S

---

<b>Table</b>		<b>Page</b>
I.1	CHARACTERISTICS OF THE SAMPLE OF HEAD START AGENCIES .....	8
I.2	REASONS WHY ASSESSMENT OBSERVATIONS WERE NOT COMPLETED IN SPRING 2006 .....	11
I.3	CODES USED TO ANALYZE QUALITATIVE DATA COLLECTED DURING SITE VISITS, BY CATEGORY .....	15
II.1	DISTRIBUTION OF CERTIFICATION SCORES ACROSS OBSERVED ENGLISH ASSESSMENTS.....	21
II.2	DISTRIBUTION OF CERTIFICATION SCORES ACROSS OBSERVED SPANISH ASSESSMENTS.....	32
V.1	DEVELOPMENTAL SCREENERS AND LOCAL ASSESSMENTS.....	87

**This page has been intentionally left blank for double-sided copying.**

## FIGURES

---

Figure		Page
III.1	OBSERVED ERRORS FOR PROGRAMS THAT OFFERED AND DID NOT OFFER A REFRESHER TRAINING (ENGLISH) .....	54

**This page has been intentionally left blank for double-sided copying.**

## EXECUTIVE SUMMARY

---

In fall 2003, the Office of Head Start began implementing the Head Start National Reporting System (NRS), an ambitious initiative to assess systematically the early literacy, language, and numeracy skills of all 4- and 5-year-olds enrolled in Head Start.<sup>1</sup> Required by a directive from the Office of the President as part of the administration's *Good Start, Grow Smart* initiative, the NRS aims to collect information on a standard set of child outcomes from all Head Start programs in a consistent manner. Head Start has a decade-long history of concentrating on child outcome measures within its performance expectations. Specific national program performance measures were developed in 1995 and, in 1996, the Head Start Program Performance Standards strengthened requirements for the ongoing screening and assessment of children throughout their Head Start participation. In 1998, after the Head Start reauthorization, all programs were required to include child outcomes in their self-assessment process by 2003. In addition to using child assessment to measure program performance, the Office of Head Start supports large scale projects that assess children's performance and experiences within Head Start using representative samples. These studies include the Head Start Impact Study and the Family and Child Experiences Survey (FACES).

Analysis of NRS data will enable the Office of Head Start to determine how children progress on a limited set of outcomes during the year preceding kindergarten. These data will provide the Office of Head Start with information that can be used to develop targeted technical assistance and to enhance its ability to report for federal accountability efforts. In addition, the Office of Head Start has suggested that the NRS can be useful for local program improvement efforts, providing an additional source of information to complement local surveys, needs assessments, and child assessments.

---

<sup>1</sup> The Improving Head Start Act, signed in December 2007, discontinued the NRS. This report focuses on NRS implementation in the spring of 2006 and was completed before the discontinuation of the NRS.

The NRS includes a 15-minute child assessment battery, a system for training staff from all Head Start grantees to administer the assessment, and a computer-based reporting system that programs use to enter the completion status of assessments and report information on the characteristics of participating Head Start programs, teachers, and children. The Office of Head Start now provides each program with a summary report of average results for all children in the program who were assessed, available within a month after each fall and spring administration and in final form several months after the final completion date of administration.<sup>2</sup> Reference tables are also created to allow programs to compare their scores to national averages, to regional averages, and to programs similar to theirs based on a number of characteristics, such as the percentage of children who are English Language Learners or the program auspice.

In July 2003, the Administration on Children, Youth and Families (ACYF) contracted with Mathematica Policy Research, Inc. (MPR) and its subcontractor, Juárez and Associates (J&A), to conduct the Head Start NRS Quality Assurance and System Development Project. The project had two components—an implementation study to assess the quality and other aspects of the first year of NRS implementation (training, child assessment, data entry, and program perspectives) and support for system development activities that could enhance the quality and usefulness of the NRS. A final report on the Year 1 Quality Assurance Study was submitted to the Office of Head Start in December 2004 (Paulsell et al. 2004). ACYF decided to extend the Quality Assurance Study into a second year and in July 2004 again contracted with MPR and J&A to continue the study with the same two project components.<sup>3</sup>

This report documents Head Start programs' experiences with the NRS during the third year of implementation. The report is based on information collected through visits to a nationally representative sample of 35 Head Start programs in spring 2006. During these visits, MPR and J&A observed a sample of child assessments and interviewed staff about their experiences implementing the NRS, particularly in spring 2006.

## **DEVELOPMENT OF THE HEAD START NATIONAL REPORTING SYSTEM**

Head Start has long emphasized continuous program improvement and outcomes-oriented accountability. With an extensive history of conducting research and program evaluations, Head Start began developing specific program performance measures in 1995, in part to be responsive to the Government Performance and Results Act (GPRA) requirements. In 1996, the Family and Child Experiences Survey (FACES) was launched to collect data on the performance indicators. The 1996 revisions of the Head Start Program

---

<sup>2</sup> Beginning with the spring 2006 data wave, programs received interim reports based on their local data in July, and full reports, with the final national averages included, in October. Note that our site visits occurred prior to this change in reporting policy.

<sup>3</sup> In July 2005, oversight of the contracts for the Head Start NRS Quality Assurance and System Development Project was transferred from the Office of Head Start to the Office of Planning, Research, and Evaluation (OPRE), Administration for Children and Families, U.S. Department of Health and Human Services.



---

Performance Standards instituted new requirements for the screening and assessment of children for the purpose of improving teaching and learning.

Following the 1998 reauthorization of Head Start, the Office of Head Start required all programs to include child outcomes in their self-assessment process by 2003. In August 2000, the Office of Head Start issued an information memorandum that laid out the steps programs must take to meet this requirement. These results-based standards and performance measures were presented in the context of a “Head Start Child Outcomes Framework,” comprising eight general developmental domains, including several—specifically, language development, literacy, and mathematics—that were targeted in the assessment requirements of the NRS (ACYF 2003). The child assessments required of all programs encompass 13 legislatively mandated indicators in language, literacy, and mathematics. In implementing these assessments, all Head Start grantees have been charged with (1) improving the objectivity of their assessments, (2) analyzing the data over time in order to understand the nature and patterns of children’s progress, and (3) incorporating the results into continuous program improvement efforts. To meet these requirements, Head Start programs were permitted to select their own assessment instruments, as long as their instruments measured progress in the required developmental domains. Programs currently use a wide range of assessment strategies and tools to measure children’s progress.

While a child outcomes approach is not new to Head Start, a national assessment system implemented consistently for all 4- and 5-year-olds is. The NRS, initiated in April 2002 when the Bush administration announced the *Good Start, Grow Smart* early childhood initiative, is a key element of the “Strengthening Head Start” component of this initiative. It creates a national assessment and reporting system out of the congressionally mandated “standards of learning,” thus carrying out the president’s directive to develop “a strategy to ensure that, for the first time, every Head Start center assesses the standards of learning in early literacy, language, and numeracy skills.” As the Assistant Secretary for Children and Families noted, “The President’s *Good Start, Grow Smart* initiative challenges us to improve the operational effectiveness of Head Start programs by developing a systematic, nationwide approach to assessing every child’s school readiness” (Horn 2003).

To aid in developing the NRS, the Office of Head Start contracted with Westat, Inc., and its subcontractor Xtria, LLC, in August 2002 and convened a Technical Work Group (TWG) of 16 experts in child development, child assessment, measurement, and program evaluation. Beginning in December 2002, the TWG met three times to advise the Office of Head Start on the design of the NRS and the selection of instruments to be included in the child assessment. In addition, the Office of Head Start convened several discussion sessions, focus groups, and workshops with Head Start program staff, early childhood researchers, and assessment experts, to discuss plans for the NRS.

### **The NRS Assessment**

In April and May 2003, Westat field-tested an initial child assessment battery with 1,434 children in 36 programs. Based on the analysis of data from the field test, and

considering input from the TWG and others, ACYF finalized a 15-minute NRS assessment battery for fall 2003 that contained four components:

1. ***English Language Screener.*** This component screens children for comprehension of spoken English to identify those whose English is insufficient to participate in the full assessment. It is composed of two subtests from the Oral Language Development Scale (OLDS) of the Preschool Language Assessment Scale (PreLAS) 2000 (Duncan and DeAvila 1998). The first set of items uses the “Simon Says” game to request that children follow simple commands, such as “touch your ear” and “point to the door.” In the second set of items, children are asked to name or describe the function of objects in pictures. A Spanish-language version of the OLDS is also used with Spanish-speaking children.
2. ***Vocabulary.*** Adapted from the third edition of the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn 1997), this section includes 24 items assessing children’s receptive vocabulary that represent a range of difficulty.
3. ***Letter Naming.*** A test developed by Westat for the Head Start Quality Research Centers Consortium (Zill 2003a), this section presents all 26 pairs of upper- and lower-case letters of the alphabet in three groupings (with 30 letters in the Spanish version). Children are asked to identify the letters they know by name.
4. ***Early Math Skills.*** Adapted from the mathematics assessment used in the Early Childhood Longitudinal Study—Kindergarten cohort (ECLS-K), this section includes items on number understanding; shape recognition; relative size judgments and measures; and simple word problems involving reading graphs, counting, or basic addition and subtraction (Zill 2003b).

A Spanish-language version of the child assessment was also developed. During the first year of NRS implementation, all children whose home language was identified as Spanish were assessed in both English and Spanish, provided they passed the language screener for each version of the assessment. All children took the English-language assessment first. Spanish-speaking children who did not attain the threshold on the English language screener were assessed in Spanish only. During the second year of implementation, children whose home language was identified as Spanish were still assessed in both languages as long as they passed the language screener for each version. However, beginning in fall 2004, in response to feedback from the TWG, the Office of Head Start staff, and local Head Start staff who participated in the Year 1 Quality Assurance Study, children whose home language was identified as Spanish took the Spanish-language assessment first, followed by the English version.

Since first implementing the NRS in fall 2003, Westat has made minor changes to the NRS battery, based on recommendations from the TWG and findings from MPR’s reports on the Head Start NRS Year 1 Quality Assurance Study. These changes have included

---

shortening introductions, changing some words in the vocabulary section, simplifying the directions on the letter naming section, improving some pictures in the math section, simplifying question wording in the math section, and simplifying the hand gestures required for some items. In addition, Westat added clarifications to training materials to allow some minor deviations from the script; these changes were made to help children feel more comfortable during the assessment.

### **The Computer-Based Reporting System**

The Office of Head Start implemented the Computer-Based Reporting System (CBRS) to collect background information on Head Start programs and children, to facilitate the identification of eligible children, and to track completed assessments. The CBRS is a web-based system where Head Start program staff members enter all relevant information. Included in the program-level data are contact information for the grantee, delegate agencies, centers, and program start and end dates. Classroom-level information includes the type of class (such as part-day or full-day), total enrollment, and number of classroom staff. Information on teacher qualifications and experience is also collected. For each eligible child, staff members enter the date of birth, classroom entry date, years in Head Start, disability status, language spoken at home, level of English proficiency, ethnicity, race, and assessment completion status. The CBRS is used to assign identification numbers and can print out class rosters for use in tracking assessments, as well as assessment completion reports.

Beginning in the second year of NRS implementation, Xtria, the subcontractor responsible for developing the CBRS, expanded its reporting capabilities and system functions. Local programs can now generate assessment completion reports at the program or center level, perform data searches, and view and operate the CBRS in Spanish. They can also take advantage of both a data copy feature—to reduce data entry duplication for certain fields that remain the same from the previous program year—and a data import feature—to import data from the program’s management information system. Centers can also enter or make changes to multiple children at one time. In addition, grantees can produce NRS data reports from the CBRS about their delegate agencies.

### **NRS QUALITY ASSURANCE AND SYSTEM DEVELOPMENT PROJECT**

As noted, the NRS is the first implementation of a nationwide assessment of 4- and 5-year-olds enrolled in Head Start. Since the scale of this initiative is unprecedented, the Office of Head Start contracted with MPR and J&A to assess the extent to which locally trained Head Start program staff across the country could administer a standardized assessment with reasonable accuracy and fidelity to the assessment protocol. In addition, because this new policy of assessing all kindergarten-eligible children had generated some controversy within the Head Start community, the Office of Head Start sought to learn more about local program perspectives on the assessment process.

Results of the Year 1 Quality Assurance Study demonstrated that, while there was room for improvement, Head Start staff members were able to administer the NRS assessment

with a fairly good degree of fidelity to the protocol (Paulsell et al. 2004). Moreover, analyses of site visit interview data and assessment observations yielded a number of helpful suggestions for improving the assessment battery and process. The Office of Head Start decided to extend the study into a second year for several reasons. First, they wanted to determine whether the level of fidelity documented through the Year 1 study would be sustained over time. Second, they wanted to determine whether comparable levels of fidelity would be observed in different samples of programs (and consequently requested that MPR draw different samples in fall and spring of the second year, for a total of 70 programs). Third, the Office of Head Start wanted to collect ongoing information about programs' experiences and perspectives on NRS implementation, including their perspectives on changes made to the assessment process from Year 1 to Year 2. An extension of the study to collect data in spring 2006 was added later.

In addition to conducting the Quality Assurance Study, MPR and J&A were charged with recommending system development strategies for enhancing the quality and usefulness of the NRS. Below we describe these two primary components of the project:

1. **Quality Assurance Study.** To observe and collect information on various facets of the NRS implementation, MPR and J&A staff made site visits to a nationally representative sample of 35 Head Start programs in fall 2003, spring 2004, spring 2005, and spring 2006. During these visits, staff observed the assessments of a random sample of children and interviewed program staff members about their experiences implementing the NRS.
2. **System Development Support.** MPR and J&A supported the Office of Head Start in assessing all aspects of the ongoing implementation of the NRS with the ultimate goal of enhancing the methods of NRS child assessment, data management, and training; and developing ways to help programs understand how they can use NRS information to improve the quality and outcomes of Head Start services for all children. Information used to formulate recommendations for improvements comes from the NRS Quality Assurance Study, program practitioners' views, input from expert consultants, Office of Head Start staff, other support contractors, and other federal agencies that often partner with ACYF/ACF on issues related to children's programs and research.

The purpose of this report is to provide updated findings from the Quality Assurance Study—focusing primarily on NRS implementation in spring 2006—and to make recommendations for system improvement.

## **QUALITY OF ASSESSMENT ADMINISTRATION**

As in earlier rounds of the NRS Quality Assurance Study, most of the English language assessments observed in spring 2006 met or exceeded the standard of quality used in certifying assessors. The average certification score was 92, and 83 percent of observed assessments exceeded the certification standard of 85 points. The inter-rater reliability of

---

assessment scale scores remained high, and the quality of the Spanish-language assessments observed (65 in 11 sites) was also high, on average (certification score of 97).

Errors due to coaching, inappropriate gestures, administering non-neutral encouragement, straying from the script, and scoring errors were higher in spring 2006 than in spring 2005; only errors for inserting articles such as “a” and “the” decreased. The increase in errors may reflect changes made to the assessment in spring 2005 (particularly those that simplified gesturing) and perhaps less attention to local refresher training that would underscore such changes in administration.

Although quality was high overall, some areas of the assessment remained difficult for staff. Modifications to the assessment easel or additional guidance and training might be helpful in these areas. The most difficult areas were also problematic in Years 1 and 2, including:

- Setting up the testing in a quiet area free of distractions
- Avoiding coaching, particularly in the Simon Says section
- Gesture errors on certain Art Show and Early Math items
- Scoring the counting item (E20) correctly in the Early Math section

Performance in these difficult areas was somewhat poorer than in spring 2005. Scoring errors on the counting item, however, remained essentially unchanged—it was scored incorrectly on roughly one in four assessments. Moreover, while the Letter Naming section yielded low error rates overall, with the exception of script errors in the introduction and transitions in between letter plates, it remains a section with which many programs continue to struggle. They worry that this part of the assessment is not valid, since children may very well know some or all of the alphabet letters, but the structure of this test administration might not enable children to demonstrate what they know. Allowing assessors to point to each letter and say, “What’s this?” or “What’s the name of this one?” could help ensure that children do not inadvertently omit a letter that they lost track of while scanning the plate.

About half of sample programs reported that most children reacted positively to the child assessment. Assessors noted that some children enjoyed the one-on-one time with the assessor and liked showing off what they knew, as well as the opportunity to “play some games.” Children’s behavior was much less of a concern in the spring than it had been in the fall, a finding similar to Years 1 and 2. The major behavioral issue was that children became bored or restless during the PPVT and/or the Letter Naming task and needed a lot of redirection. Nearly half of the programs thought that the NRS assessment is too long.

Most programs (24 out of 35, or 68 percent) administered the child assessment to English Language Learners. Spanish was the most common other language spoken, followed by Chinese and Arabic. Assessors reported that most children could pass the English language screener in the spring, even if they had not passed in the fall. Some children who are classified as having Spanish as a primary language frequently are more

fluent in English; similarly, some children lose their Spanish skills over the course of the school year, and perform better on the English version in the spring. Bilingual assessors, as has been the case in previous rounds of site visits, were critical about two overarching components of the NRS assessment. First, they felt that certain Spanish speakers are at a disadvantage due to the wide variety of dialects spoken by children and families who use colloquial words not considered to be correct on the NRS assessment. Second, these assessors asserted that children are tested on four additional alphabet letters (CH, LL, Ñ, and RR) that are no longer used in Spanish. However, further investigation has shown that only RR has been discontinued in current Spanish usage, and it is therefore being eliminated from the assessment.

Nearly all programs (33 out of 35 in the sample) assessed children with disabilities, using a wide range of accommodations for the child assessments. In some situations, assessors decided to administer the NRS assessment to a child who could reasonably have been exempted. Most staff members who administered assessments to children with disabilities felt comfortable with the process and with the accommodations made. However, several programs would like the NRS training materials to address explicitly the issues of conducting the NRS assessment to children with special needs and making appropriate accommodations for these testing situations. In each round of the QA study, a number of Head Start programs have expressed interest in these same issues.

## **LOCAL APPROACHES TO NRS TRAINING**

To ensure consistent administration of the NRS assessment, Head Start programs need to follow standardized procedures when training local staff. Since the NRS was first implemented, the Office of Head Start has provided local programs with standardized training materials and guidelines to conduct training.

In summer 2005, few programs sent a representative to one of the “training-of-trainers” events offered for programs that have had a change in their NRS lead. Several programs, however, would like to see refresher trainings for NRS lead trainers, and to be able to send more than one staff member to a regional conference to build local capacity for NRS implementation. They would prefer training sessions that are available closer to their programs.

In spring 2006, all but three of the sample programs provided refresher training. However, only about one quarter adhered to the training guidelines. The other programs did not include all of the required training elements—viewing and discussing the training video, distributing the Assessor’s Guide, and role-playing the assessment—and thus most training sessions were shorter than the recommended four and a half hours. About half of the programs that needed to provide refresher training for Spanish assessors reported offering a formal group session; bilingual assessors at most other programs relied on self-study to prepare for the spring 2006 assessments. Some reasons why most programs did not adhere to the training protocol include late arrival of training and assessment materials, limited time to complete assessments by the end of the program year, and staff members’ perceptions that experienced NRS assessors did not need refresher training.

Eighteen percent of the sample programs trained new assessors in spring 2006. Although all of these programs provided separate training for new assessors and most programs certified them during a practice assessment, none of the programs followed the required training protocol for new assessors. Instead of providing a full-day training, most programs conducted the training in two to three hours. In rare cases, a small number of new assessors did not get formally certified on the NRS after attending a training session.

As in past rounds of site visits, the responses of local Head Start staff to the NRS materials and information provided by the Office of Head Start were mixed. At least one staff member at most programs watched one of the NRS broadcasts or webcasts, but programs' views of their usefulness were uneven. Problems continued with timely delivery of training and assessment materials, with nearly half of programs reporting a delay in receipt, and few programs took advantage of technical assistance offered through the help line or other means. While staff from two-thirds of programs found the video to be the best or one of the best components of local NRS training, other staff found the video too long and repetitive and preferred the role-playing exercises or the Assessor's Guide to help prepare for the spring assessment. Lack of thorough refresher training in many programs, coupled with a tendency not to seek technical assistance, may have contributed to some of the errors in administration and scoring described in Chapter II. This point is important because analysis of error rates between programs that did and did not provide refresher training showed that assessors in programs without refresher training committed significantly more errors in administration than assessors from programs that did hold them.

## **LOCAL APPROACHES TO IMPLEMENTATION**

Overall, the 35 Head Start programs we visited in spring 2006 took an approach to NRS implementation similar to that of programs visited in previous rounds of site visits. Program directors assigned a lead NRS trainer the responsibility of overseeing implementation, including assigning, training, and certifying assessors; scheduling and tracking the completion of assessments; overseeing quality assurance activities; and submitting score sheets by the deadline set by the Office of Head Start. In more than a third of programs, the lead trainer was also responsible for maintaining the CBRS.

Most programs maintained the same basic staffing structure for the NRS in the spring that they had instituted in the fall. When programs did make changes, most were due to staff turnover within the program rather than a rethinking of the program's approach to the NRS. On average, the programs trained 17 assessors. As in the sample of programs we visited in spring 2005, most programs did not rely exclusively on teaching staff to conduct the assessments. Program staff saw many advantages to having teachers administer the assessment, including that the children were comfortable and familiar with teachers; children responded better to teachers; and teachers had better knowledge of children's behavior, child development, and child assessments. Learning more about children's knowledge in specific skill areas could inform teachers' classroom practice. However, many programs expressed concerns about teacher burden, reduction in instructional time, potential for coaching or bias in administration, and the cost of hiring substitutes to cover for teachers while they were assessing children.

Nearly all programs communicated with parents and Policy Councils about the NRS. While the means of communication varied, programs made an effort to inform parents of the assessment at the start of the program year. Three-quarters of programs sought to obtain parents' written consent for the NRS assessment, far more than in previous rounds of site visits. Most programs had few or no parent refusals. Concerns among parent and Policy Council members were similar to those expressed in previous rounds of visits. They included parent requests to see their child's individual results; concerns about bias of specific items; and questions about the purpose of the NRS, how the results would be used, and how the results would affect the Head Start program.

Few Head Start directors estimated the monetary and in-kind costs of NRS implementation. However, a number of directors identified significant costs, such as staff time, travel costs, costs for paying substitute teachers, and overtime for staff working on the NRS.

### **USING THE COMPUTER-BASED REPORTING SYSTEM**

As in past rounds of site visits, programs reported that the CBRS was easy to learn and use. Programs had adequate numbers of computers and Internet connections for accessing the system. Many programs used and liked the enhanced features of the CBRS introduced in Year 2, especially the data copy, data import, and expanded report features. Some programs reported accessing the CBRS help line at least once and most of these reported receiving prompt and helpful support from the help line staff. Data managers made some suggestions for further enhancements to the CBRS, including reducing the cost of the data import feature, providing more support for using the data copy feature, and adding other user-friendly features.

### **USING THE NRS FOR LOCAL PROGRAM IMPROVEMENT EFFORTS**

The majority of Head Start staff who participated in the site visits found the 2004–2005 Growth Report easy to understand, although some said it was somewhat confusing, particularly for non-technical audiences. Staff made several recommendations to improve the report's content (for example, provide information on how subgroups of children performed) and format (for example, distribute a version that yields better photocopies). Staff in about two-thirds of the sample programs agreed that center or classroom-level reporting would be more useful in targeting practices and resources; about a quarter of programs requested individual child-level data. A majority of the sample programs shared NRS results with staff and key stakeholders, such as managers, specialists, and teachers; Policy Councils; boards of directors; and, to a lesser degree, parents. Some programs shared details on how they have used NRS data to modify classroom practices, such as spending more time on alphabet knowledge, literacy development, and, to a lesser degree, counting and other early math skills. Because the NRS reports did not provide center-, classroom- or child-level information, 11 percent of the programs reported tracking the item responses of individual children using the raw data on the Scantron answer sheets to make decisions on what areas needed improvement to better prepare children for kindergarten.



With regard to using the NRS results in the future, more than half of the sample programs said they planned to use the reports to some degree. In many cases, staff did not provide much detail on these plans, but they intend to continue incorporating NRS outcomes as one source of information for program-planning efforts. About 20 percent of programs either remained undecided about whether they would use the results or did not plan to use them at all. However, 11 percent of programs planned to observe trends over time, now that multiple years of data are available.

### **PERSPECTIVES OF LOCAL HEAD START STAFF ON THE NATIONAL REPORTING SYSTEM**

When asked to list the contributions the NRS had made to their programs, staff members in sample programs were able to attribute a number of positive effects of the NRS. Primarily, the contributions that staff listed were that it (1) helped to improve classroom practice and inform teacher training, (2) proved Head Start is effective, (3) raised program accountability by providing national comparisons, and (4) validated local assessment results. In a few programs, a contribution of the NRS was to raise staff morale. In two programs, staff indicated that a contribution of the NRS was that it helps to get children accustomed to a testing environment.

Local program staff raised several concerns about the NRS and its implications for future directions Head Start may take; most of these concerns also had been raised in previous rounds of site visits. The concern expressed most often by local staff members was about how the NRS results would be used at the national level. Many staff said that they still did not have a clear understanding of the purpose of the NRS and its implications for local programs where children did not perform well on the assessment. Local staff also expressed concern about whether the NRS results accurately reflect program performance, the amount of staff time and financial resources dedicated to the NRS, and whether it was valid to compare fall and spring assessments.

To improve NRS implementation, programs suggested that the Office of Head Start share more information about how the results would be used, send training materials and outcome reports to programs sooner, provide more information about the assessment's development and validity, and consult more with the Head Start community about future changes to the assessment. Many programs requested written materials about the NRS for parents.

Regarding the assessment battery, programs were split on whether new domains should be added. Many staff thought that at least one new domain—particularly social-emotional development—should be added, but at the same time they did not want the assessment to become longer. (The social-emotional component of the NRS was implemented for the first time in fall 2006, after our site visits.) Program staff continued to recommend that the NRS be combined with the local assessments. They also suggested changes to improve the Spanish-language version and to modify specific assessment procedures and items.

## IMPLICATIONS OF FINDINGS FROM THE QUALITY ASSURANCE STUDY

The value of the NRS for local program purposes is dependent upon how well local Head Start programs implement it, whether they accept it as a valid and reliable assessment, and whether they can use the information effectively for program improvement. Accuracy of administration, validity and reliability are also important to the other OHS goals for the NRS: to develop targeted technical assistance and to enhance the ability to report for federal accountability efforts. On the basis of findings from site visits to nationally representative samples of programs over the first three years of NRS implementation and our understanding of concerns raised by the Office of Head Start, expert consultants including Head Start program directors, and others in the field, we make the following suggestions for system improvement of the NRS. These suggestions are of several types: (1) ideas for increasing communication with local Head Start programs to improve knowledge and understanding of the purposes of the system, (2) ways to improve access to and usefulness of NRS results for local programs, (3) options for supporting programs in administering the assessment (including guidance on assessing children in Spanish and children with disabilities), and (4) ideas for changes to the assessment battery. Over the course of five rounds of visits to representative samples of Head Start programs, we have seen evidence that feedback on problems has led to improvements in the system. However, we have also found evidence that some problems persist and deserve further attention.

### Communication

Several concerns raised by local program staff during three years of the Quality Assurance Study appear to be fueled by uncertainty and lack of information. For example, not knowing the specifics of how the NRS results will be used at the national level has led to speculation about whether it will become a “high-stakes” test, helping to make decisions on funding or employment at the program, center, or classroom level. Increased communication about the Office of Head Start’s plans could address some of these concerns. Suggestions for improving communication follow:

- Provide more information about how the Office of Head Start will use the NRS assessment results.
- Provide a technical report or background materials about how the NRS assessment battery was developed and the validity and reliability of each task, including predictive validity for later school performance. Analyses of NRS test data could be done to investigate this issue, leading either to changes in items or reassurance that children’s skills are not being unfairly measured by test items.
- Explore options for increasing consultation with Head Start practitioners on future changes to the NRS.
- Provide written materials about the NRS for parents. During site visits, some staff reported feeling unsure about how to report NRS results to parents and Policy Councils or respond to their questions about the NRS.

---

### **Access to and Usefulness of Assessment Results for Local Programs**

Each round of data collection reflects more programs that are attempting to make use of the NRS results to improve their classroom practice, but they continue to manifest a need for guidance in how to accomplish this goal. As in spring 2005, program staff suggested several ways to make the reports more useful to them and requested help in interpreting the results and implementing appropriate changes to their programs:

- Send reports sooner after each round of data collection, preferably in time to use them in planning for the next program year.
- Consider providing results disaggregated to the center or classroom level.
- Provide programs with more guidance on how to use the NRS results for program improvement, including building more linkages with local assessment results.
- Provide access to resources to support local program improvement efforts undertaken in response to assessment results.

### **Support for Administering the Assessment**

Timely shipment of training and assessment materials and modest changes in the NRS training protocols have the potential to help staff feel more confident and conduct the assessments more smoothly and accurately. Specifically, recommended changes include the following:

- Ensure that programs receive materials and scheduling information on time.
- Ensure that adequate refresher training is carried out to reduce errors in administration. We conducted analyses that demonstrated a link between offering assessor trainings and assessor performance on the assessment.
- Consider requiring recertification of assessors during refresher training.
- Provide more guidance on how to interact appropriately with children during the assessment.

### **Guidance on Assessing Children in Spanish**

Assessors and trainers continued to have some concerns about administering the Spanish version of the NRS and expectations for growth in Spanish-language skills. Some possible approaches to addressing these concerns include the following:

- Allow for more regional differences in Spanish terms used in the assessment.
- Clarify expectations for growth in Spanish skills.

### **Guidance on Assessing Children with Disabilities**

All but two programs in the sample assessed some children with identified disabilities during spring 2006. Although the majority of programs felt comfortable with the accommodations they made, some requested more guidance on how to assess children with disabilities. More guidance would be helpful in the following areas:

- Clarify when a child should not be assessed because of a disability.
- Provide more examples of appropriate accommodations.

### **Improving the Assessment Instrument**

Programs acknowledge that improvements have been made to the assessment instrument over time, but continue to raise several concerns:

- A large majority of sample programs (86 percent) reported some concerns with the receptive vocabulary (PPVT) section of the test that they fear may affect overall validity, or may differentially affect validity for some subgroups, like urban versus rural children.
- The primary concerns in the Letter Naming section were due to the layout of the plates and the method of administering them.
- The form or content of the Early Math section was the source of concern for close to half of programs; for example, many programs felt that the concepts such as subtraction and reading graphs were too advanced for this age group.

In summary, based on our visits to a representative sample of Head Start programs, we suggest implications for system improvement in two major areas: (1) helping program staff better understand the purposes, interpretation, and potential uses of the NRS through improved communication and guidance (including a special focus on children with disabilities and children assessed in Spanish) and (2) making improvements in the training, administration guidelines, format, and content of the battery itself. Through the Quality Assurance Study, MPR will continue to work with the Office of Head Start, expert consultants including Head Start program staff, and the implementation contractors to foster improvements to the entire assessment system.

# CHAPTER I

## INTRODUCTION

---

In fall 2003, the Office of Head Start began implementing the Head Start National Reporting System (NRS), an ambitious initiative to assess systematically the early literacy, language, and numeracy skills of all 4- and 5-year-olds enrolled in Head Start.<sup>1</sup> Required by a directive from the Office of the President as part of the administration's *Good Start, Grow Smart* initiative, the NRS aims to collect information on a standard set of child outcomes from all Head Start programs in a consistent manner. Analysis of these data will enable the Head Start Bureau to determine how children progress on a limited set of outcomes during the year preceding kindergarten. The data will also provide the Office of Head Start with information that can be used to develop targeted technical assistance and to enhance its ability to report for federal accountability efforts. In addition, the Office of Head Start has suggested that the NRS can be useful for local program improvement efforts, providing an additional source of information to complement local surveys, needs assessments, and child assessments.

The NRS includes a 15-minute child assessment battery, a system for training staff from all Head Start grantees to administer the assessment, and a computer-based reporting system that programs use to enter the completion status of assessments and report information on the characteristics of participating Head Start programs, teachers, and children. The Office of Head Start provides each program with a summary report of average results for all children in the program who were assessed.<sup>2</sup> Reference tables are also created to allow programs to compare their scores to national averages, to regional averages, and to programs similar to theirs based on a number of characteristics, such as the percentage of children who are English Language Learners or the program auspice.

---

<sup>1</sup> The Improving Head Start Act, signed in December, 2007, discontinued the NRS. This report focuses on NRS implementation in the spring of 2006, and was completed prior to the discontinuation of the NRS.

<sup>2</sup> Beginning with the spring 2006 data wave, programs received interim reports based on their local data in July, and full reports, with the final national averages included, in October. Note that our site visits occurred prior to this change in reporting policy.

In July 2003, the Administration on Children, Youth and Families (ACYF) contracted with Mathematica Policy Research, Inc. (MPR), and its subcontractor Juárez and Associates (J&A), to conduct the Head Start NRS Quality Assurance and System Development Project. The project had two components—an implementation study to assess the quality and other aspects of the first year of NRS implementation (training, child assessment, data entry, and program perspectives) and support for system development activities that could enhance the quality and usefulness of the NRS. A final report on the Year 1 Quality Assurance Study was submitted to the Office of Head Start in December 2004 (Paulsell et al. 2004). ACYF decided to extend the Quality Assurance Study into a second year and in July 2004 again contracted with MPR and J&A to continue the study with the same two project components.<sup>3</sup>

This report documents Head Start programs' experiences with the NRS during the third year of implementation. The report is based on information collected through visits to a nationally representative sample of 35 Head Start programs in spring 2006. During these visits, MPR and J&A observed a sample of child assessments and interviewed staff about their experiences implementing the NRS, particularly in spring 2006. In the rest of this introductory chapter, we describe the development of the NRS, the Head Start NRS Quality Assurance and System Development Project, the sample of programs and children selected for the spring 2006 site visits and assessment observations, the methods used to collect data on NRS implementation, and methods used to analyze these data.

## **DEVELOPMENT OF THE HEAD START NATIONAL REPORTING SYSTEM**

Head Start has long emphasized continuous program improvement and outcomes-oriented accountability. With an extensive history of conducting research and program evaluations, Head Start began developing specific program performance measures in 1995, in part to be responsive to the Government Performance and Results Act (GPRA) requirements. In 1996, the Family and Child Experiences Survey (FACES) was launched to collect data on the performance indicators. The 1996 revisions of the Head Start Program Performance Standards instituted new requirements for the screening and assessment of children for the purpose of improving teaching and learning.

Following the 1998 reauthorization of Head Start, the Office of Head Start required all programs to include child outcomes in their self-assessment process by 2003. In August 2000, the Office of Head Start issued an information memorandum that laid out the steps programs must take to meet this requirement. These results-based standards and performance measures were presented in the context of a “Head Start Child Outcomes Framework,” comprising eight general developmental domains, including several—for instance, language development, literacy, and mathematics—that were targeted in the assessment requirements of the NRS (Administration on Children, Youth and Families

---

<sup>3</sup> In July 2005, oversight of the contracts for the Head Start NRS Quality Assurance and System Development Project was transferred from the Office of Head Start to the Office of Planning, Research, and Evaluation (OPRE), Administration for Children and Families, U.S. Department of Health and Human Services.

---

2003). The child assessments required of all programs encompass 13 legislatively mandated indicators in language, literacy, and mathematics. In implementing these assessments, all Head Start grantees have been charged with (1) improving the objectivity of their assessments, (2) analyzing the data over time in order to understand the nature and patterns of children's progress, and (3) incorporating the results into continuous program improvement efforts. To meet these requirements, Head Start programs were permitted to select their own assessment instruments, as long as their instruments measured progress in the required developmental domains. Programs currently use a wide range of assessment strategies and tools to measure children's progress.

While a child outcomes approach is not new to Head Start, a national assessment system implemented consistently for all 4- and 5-year-olds is. The NRS, initiated in April 2002 when the Bush administration announced the *Good Start, Grow Smart* early childhood initiative, is a key element of the "Strengthening Head Start" component of this initiative. It creates a national assessment and reporting system out of the congressionally mandated "standards of learning," thus carrying out the President's directive to develop "a strategy to ensure that, for the first time, every Head Start center assesses the standards of learning in early literacy, language, and numeracy skills." As the Assistant Secretary for Children and Families noted, "The President's *Good Start, Grow Smart* initiative challenges us to improve the operational effectiveness of Head Start programs by developing a systematic, nationwide approach to assessing every child's school readiness" (Horn 2003).

### **The NRS Assessment**

To aid in developing the NRS, the Head Start Bureau contracted with Westat, Inc., and its subcontractor Xtria, LLC, in August 2002 and convened a Technical Work Group (TWG) of 16 experts in child development, child assessment, measurement, and program evaluation. Beginning in December 2002, the TWG met three times to advise the Office of Head Start on the design of the NRS and the selection of instruments to be included in the child assessment. In addition, the Office of Head Start convened several discussion sessions, focus groups, and workshops with Head Start program staff, early childhood researchers, and assessment experts, to discuss plans for the NRS.

In April and May 2003, Westat field-tested an initial child assessment battery with 1,434 children in 36 programs. Based on the analysis of data from the field test, and considering input from the TWG and others, ACYF finalized a 15-minute NRS assessment battery for fall 2003 that contained four components:

1. ***English Language Screener.*** This component serves as an English-language screener to identify children whose English is insufficient to participate in the full assessment. It is composed of two subtests from the Oral Language Development Scale (OLDS) of the Preschool Language Assessment Scale (PreLAS) 2000 (Duncan and DeAvila 1998). The first set of items uses the "Simon Says" game to request that children follow simple commands, such as "touch your ear" and "point to the door." In the second set of items, children

are asked to name or describe the function of objects in pictures. A Spanish-language version of the OLDS is also used with Spanish-speaking children.

2. **Vocabulary.** Adapted from the third edition of the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn 1997), this section includes 24 items assessing children's receptive vocabulary that represent a range of difficulty.
3. **Letter Naming.** A test developed by Westat for the Head Start Quality Research Centers Consortium (Zill 2003b), this section presents all 26 pairs of upper- and lower-case letters of the alphabet in three groupings (with 30 letters in the Spanish version). Children are asked to identify the letters they know by name.
4. **Early Math Skills.** Adapted from the mathematics assessment used in the Early Childhood Longitudinal Study—Kindergarten cohort (ECLS-K), this section includes items on number understanding; shape recognition; relative size judgments and measures; and simple word problems involving reading graphs, counting, or basic addition and subtraction (Zill 2003a).

A Spanish-language version of the child assessment was also developed. All children whose home language was identified as Spanish were assessed in both English and Spanish, provided they passed the language screener for each version of the assessment. During the first year of implementation, all children took the English-language assessment first. Spanish-speaking children who did not attain the threshold on the English language screener were assessed in Spanish only. During the second year of implementation, children whose home language was identified as Spanish were still assessed in both languages as long as they passed the language screener for each version. However, beginning in fall 2004, in response to feedback from the TWG, the Office of Head Start staff, and local Head Start staff who participated in the Year 1 Quality Assurance Study, children whose home language was identified as Spanish took the Spanish-language assessment first, followed by the English version.

Since first implementing the NRS in fall 2003, Westat has made minor changes to the NRS battery, based on recommendations from the TWG and findings from MPR's reports on the Head Start NRS Year 1 Quality Assurance Study. These changes have included shortening introductions, changing some words in the vocabulary section, simplifying the directions on the letter naming section, improving some pictures in the math section, simplifying question wording in the math section, and simplifying the hand gestures required for some items. In addition, Westat added clarifications to training materials to allow some minor deviations from the script; these enable assessors to help children feel more comfortable during the assessment.

### **The Computer-Based Reporting System**

The Office of Head Start implemented the Computer-Based Reporting System (CBRS) to collect background information on Head Start programs and children, to facilitate the



---

identification of eligible children, and to track completed assessments. The CBRS is a web-based system where Head Start program staff members enter all relevant information. Included in the program-level data are contact information for the grantee, delegate agencies, centers, and program start and end dates. Classroom-level information includes the type of class (such as part-day or full-day), total enrollment, and number of classroom staff. Information on teacher qualifications and experience is also collected. For each eligible child, staff enters the date of birth, classroom entry date, years in Head Start, disability status, language spoken at home, level of English proficiency, ethnicity, race, and assessment completion status. The CBRS is used to assign identification numbers for both assessors and children and can print out class rosters for use in tracking assessments, as well as assessment completion reports.

Beginning in the second year of NRS implementation, Xtria, the subcontractor responsible for developing the CBRS, expanded its reporting capabilities and system functions. Local programs can now generate assessment completion reports at the program or center level, perform data searches, and view and operate the CBRS in Spanish. They can also take advantage of both a data copy feature—to reduce data entry duplication for certain fields that remain the same from the previous program year—and a data import feature—to import data from the program’s management information system. Centers can also enter or make changes to multiple children at one time. In addition, grantees can produce NRS data reports from the CBRS about their delegate agencies.

### **NRS QUALITY ASSURANCE AND SYSTEM DEVELOPMENT PROJECT**

As noted, the NRS is the first implementation of a nationwide assessment of 4- and 5-year-olds enrolled in Head Start. Since the scale of this initiative is unprecedented, the Office of Head Start contracted with MPR and J&A to assess the extent to which locally trained Head Start program staff across the country could administer a standardized assessment with reasonable accuracy and fidelity to the assessment protocol. In addition, because this new policy of assessing all kindergarten-eligible children had generated some controversy within the Head Start community, the Office of Head Start sought to learn more about local program perspectives on the assessment process.

Results of the Year 1 Quality Assurance Study demonstrated that, while there was room for improvement, Head Start staff members were able to administer the NRS assessment with a fairly good degree of fidelity to the protocol (Paulsell et al. 2004). Moreover, analyses of site visit interview data and assessment observations yielded a number of helpful suggestions for improving the assessment battery and process. The Office of Head Start decided to extend the study into a second year for several reasons. First, they wanted to determine whether the level of fidelity documented through the Year 1 study would be sustained over time. Second, they wanted to determine whether comparable levels of fidelity would be observed in different samples of programs (and consequently requested that MPR draw different samples in fall and spring of the second year, for a total of 70 programs). Third, the Office of Head Start wanted to collect ongoing information about programs’ experiences and perspectives on NRS implementation, including their perspectives on

changes made to the assessment process from Year 1 to Year 2. An extension of the study to collect data in spring 2006 was added later.

In addition to conducting the Quality Assurance Study, MPR and J&A were charged with recommending system development strategies for enhancing the quality and usefulness of the NRS. Below, we describe these two primary components of the project:

1. ***Quality Assurance Study.*** To observe and collect information on various facets of the third year of NRS implementation, MPR and J&A staff made site visits to a nationally representative sample of 35 Head Start programs in fall 2003, spring 2004, spring 2005, and spring 2006. During these visits, staff observed the assessments of a random sample of children and interviewed program staff members about their experiences implementing the NRS.
2. ***System Development Support.*** MPR and J&A supported the Office of Head Start in assessing all aspects of the ongoing implementation of the NRS with the ultimate goal of enhancing the methods of NRS child assessment, data management, and training; and developing ways to help programs understand how they can use NRS information to improve the quality and outcomes of Head Start services for all children. Information used to formulate recommendations for improvements comes from the NRS Quality Assurance Study, program practitioners' views, input from expert consultants, Office of Head Start staff, other support contractors, and other federal agencies that often partner with ACYF/ACF on issues related to children's programs and research.

## **SITE VISITS**

This report is based on site visits to a nationally representative sample of 35 Head Start agencies conducted in spring 2006. Here, we describe our methods for selecting the sample of agencies and children to be observed, as well as the training and qualifications of site visitors.

### **Sample Selection**

This section describes the procedures we used for selecting a sample of Head Start programs to visit and the children whose assessments we would observe. The goal of our sampling strategy was to select a nationally representative sample of 35 Head Start agencies, and a random and reasonably representative sample of, on average, 10 children per program (ranging from 5 in smaller programs to 15 in larger programs) whose assessments would be observed during the site visit. We were constrained in our selection of children by the need to conduct each site visit within about two days. Thus, we needed to limit the number of Head Start centers where children would be observed. Below, we describe in detail our procedures for selecting a nationally representative sample of Head Start programs, along with the procedures for selecting centers within programs (or home visitors for those

---

programs that offered home-based services) and children within centers (or from home visitors' caseloads).

**Programs.** To select a representative sample of 35 Head Start programs in spring 2006, we used the most current (2004–2005) Head Start Program Information Report (PIR) as the basis for a sampling frame.<sup>4</sup> Programs in the PIR that were eligible for this study were those that operated in the 48 contiguous United States, the District of Columbia, and Puerto Rico, and those that provided services to children ages 4 and 5.<sup>5</sup> We used as the primary sampling unit the PIR reporting unit—that is, grantees and delegate agencies (referred to herein as “programs”).

We used a sequential technique based on a procedure developed by Chromy (1979) and available in SAS (SurveySelect). This procedure offers all the advantages of the systematic sampling approach but eliminates the risk of systematic, list-order bias by making independent selections within each of the zones associated with systematic sampling, while controlling the selection opportunities for units crossing zone boundaries. No explicit stratification was used, and agencies were selected with equal probability. We implicitly stratified (sorted) the frame of Head Start agencies by Census region. Within a region, we sorted by racial and ethnic distribution, metropolitan status, and enrollment size.<sup>6</sup>

In preparation for the site visits conducted in spring 2006, we initially selected twice the number of programs that we were targeting (70 rather than 35), then formed sequential pairs of programs. Next, we randomly selected one program from each pair to be “released” for the sample. If a program was unable to participate in the study or was judged to be ineligible (for example, if it was about to be de-funded), we released the paired agency. Seven programs in the sample were replaced by the agency paired with them. Reasons for nonparticipation of the originally sampled programs included ineligibility and refusal to participate. (See Table I.1 for the characteristics of sample Head Start agencies.)

---

<sup>4</sup> The PIR is an annual report submitted by every Head Start grantee and delegate agency. It provides detailed information on the children and families served, staffing, and programs operated by the agency.

<sup>5</sup> We excluded programs in Alaska, Hawaii, and the U.S. territories other than Puerto Rico, as well as programs that were only Early Head Start and those serving no 4- and 5-year olds, according to the 2004–2005 PIR.

<sup>6</sup> “Census region” included the four regions—Northeast, South, Midwest, and West—plus Puerto Rico as a fifth category. The racial/ethnic variable had three categories: (1) more than 40 percent Hispanic, (2) 40 percent or fewer Hispanic and more than 40 percent African American, and (3) other. Metropolitan status had three categories: (1) metro, (2) non-metro, and (3) Puerto Rico. Enrollment size referred to the number of 4- and 5-year olds in 2004–2005, and had two categories: (1) 200 or fewer, and (2) more than 200.

**Table I.1. Characteristics of the Sample of Head Start Agencies**

Characteristics	Number of Agencies
<b>ACF Region</b>	
Region I	1
Region II	4
Region III	5
Region IV	5
Region V	6
Region VI	6
Region VII	2
Region VIII	2
Region IX	1
Region X	1
Region XI (Tribal Programs)	2
<b>Total</b>	<b>35</b>
<b>Head Start Grantee Status</b>	
Grantee That Operates Programs Directly and Does Not Have Delegate Agencies	27
Grantee That Operates Programs Directly and Has Delegate Agencies	1
Delegate Agency	7
<b>Total</b>	<b>35</b>
<b>Head Start Program Option</b>	
Center-Based Only	21
Combination of Center-Based and Home-Based	13
Locally Designed Option <sup>1</sup>	1
<b>Total</b>	<b>35</b>
<b>Type of Agency</b>	
Community Action Agency	15
School System	4
Public or Private Nonprofit	11
Tribal Government or Consortium	2
Government Agency	3
<b>Total</b>	<b>35</b>
<b>Metropolitan Status of Agency's Service Area</b>	
Metropolitan Area	22
Non-Metropolitan Area	11
Puerto Rico	2
<b>Total</b>	<b>35</b>
<b>Number of Head Start Centers the Agency Operates</b>	
1 Center	5
2 to 5 Centers	8
6 to 10 Centers	11
11 to 20 Centers	7
21 or More Centers	4
<b>Total</b>	<b>35</b>
<b>Number of Enrolled 4- and 5-Year-Olds</b>	
1 to 50 Children	4
51 to 100 Children	4
101 to 200 Children	6
201 to 300 Children	6
301 to 400 Children	7
401 to 500 Children	3
501 to 600 Children	2
601 to 700 Children	2
More than 700 Children	1
<b>Total</b>	<b>35</b>

Source: Head Start Program Information Report (PIR) on sampled programs, from program year 2004–2005.

<sup>1</sup> Often refers to serving families through Family Child Care homes.

---

**Centers Within Programs.** Many of the Head Start programs provided services at two or more centers. Given time and resource constraints for conducting the observation visits, we sorted Head Start centers within programs into groups of two or three centers that could be visited by a single observer during a site visit, and then selected one group at random with equal probability, using a simple random sample. We attempted to make the groups of centers comparable in size before sampling. For some programs that had one very large center (with many classrooms) and one smaller center, we subdivided the large center into two or more groups. If a program operated a home-based option, we sorted home visitors into groups and then randomly selected one group. All the programs in our sample that operated a home-based option also operated centers; for these programs, we selected one group of centers and one group of home visitors.

**Children Within Centers.** Once we selected groups of centers and/or home visitors, we requested caseload rosters for those centers and home visitors, and then selected a random sample of children to observe. By selecting children for observation in advance of the site visits, we were more likely to observe a representative sample, rather than observing those children who might happen to be available during the site visit or those who might be easy to assess.

We used Chromy's procedure to select children within the selected centers or home visitor caseloads. Before selecting the sample, we excluded children who, based on their date of birth, were not likely to be eligible for kindergarten the next year. To achieve a more representative sample, we sorted the children by classroom, disability status, and language (those who spoke Spanish at home), to the extent that programs were able to provide this information before making our selection. No explicit stratification was used, and children were selected with equal probability. In most of the agencies, our goal was to observe 10 assessments. In the five largest agencies (based on the 2004–2005 PIR), we aimed for 15 assessments, and in the five smallest agencies, for 5 assessments. Because we expected some children to be absent on the day of the assessment, we initially selected twice the number of children needed. We then used Chromy's procedure again to select half of these children to be part of the original sample release. The children not selected to be in the original release were randomly sorted and used as additional releases when necessary. We observed the Spanish assessments of all children who spoke Spanish at home (but not their English assessments).

### **The Sample of Head Start Programs and Children Observed**

The sample of 35 Head Start programs that participated in the spring 2006 site visits includes a diverse set of programs (Table I.1). According to the most recent PIR, the mean enrollment of 4- and 5-year-olds (those eligible for the NRS assessment) was 275, ranging from 14 to 707 across the sample programs. All selected programs provided primarily center-based services; the number of centers operated by the programs ranged from 1 to 52. Thirteen programs also offered a home-based option to some families, with the number of funded home-based slots ranging from 2 to 140 across programs.

Comparing the NRS 2006 child sample with PIR data on all programs, the sample appears representative of the racial/ethnic composition of the population served. For

example, in the 2005-2006 PIR, the average percentage of Hispanic children among all enrolled children is 27.6 percent; an average of 19 percent indicate their home language is Spanish. The 2006 NRS child sample is 26 percent Hispanic, and 19 percent Spanish-speaking at home. The NRS 2006 sample demonstrates an increase in children in Spanish-speaking families over the 2005 sample, from 11 percent in 2005 to 19 percent in 2006.<sup>2</sup>

In spring 2006, our primary sample of 337 children included 36 children with an identified disability and 148 children whose home language was Spanish.<sup>3</sup> When children from the primary sample could not be observed, we released additional children from the replacement sample, bringing the total number of children in the sample to 526. Of the 526, we completed 360 observations of child assessments; 252 of these were from the primary sample, and 108 were from the replacement sample.<sup>4</sup> Thirty-one children were ineligible to participate (not going to kindergarten the following year) and 135 were eligible non-completes (for example, some children were absent the day of the site visit, and others had been assessed prior to the observation visit; Table I.2). Because the assessment observations were, in general, scheduled for only one day, absent children were replaced immediately from the replacement sample.

The sample of children observed in spring 2006 included 36 with an identified disability and 148 whose home language was Spanish; all Spanish-speaking children were observed during the Spanish-language version of the assessment. The leading reason that site visitors were not able to observe children in the main sample was absences on the day of the visit.

### **Developing Site Visit Protocols**

We developed two sets of instruments for collecting information during the site visits: (1) a structured observation tool for observing the child assessments (Appendix A), and (2) semi-structured interview and focus group guides (Appendix B).

---

<sup>2</sup> With any random sample, the distribution of certain characteristics can be different from the population due to chance, unless it is controlled for explicitly in the design using stratification, which we did not do for Hispanic ethnicity or Spanish home language.

<sup>3</sup> Our target number of observations was 350, or an average of 10 observations per program. Depending on program size, we observed between 5 and 15 assessments per site.

<sup>4</sup> In a handful of sites, site visitors were able to observe one, two, or three assessments more than the required number. Two children were observed in both English and Spanish.

**Table I.2. Reasons Why Assessment Observations Were Not Completed in Spring 2006**

Reason	Primary Sample	Replacement Sample	Total
<b>Ineligible for the NRS</b>			
Family Dropped Out of the Program	22	6	28
Child Is Not Kindergarten Eligible	1	1	2
Child Has Disability	1	0	1
Total Ineligible	24	7	31
<b>Eligible Non-Completes</b>			
Child Absent or Unavailable	69	40	110
Parent Refused Permission for the Assessment	3	2	5
Site Refusal (Sampled children not assessed during visit)	4	4	8
Child Assessed Prior to Site Visit	2	7	9
Child Could Not Be Assessed Due to Behavior Problem	0	1	1
Spanish Assessor Not Available During the Site Visit	1	2	3
Total Eligible Non-Completes	79	56	135
<b>Grand Total</b>	<b>103</b>	<b>63</b>	<b>166</b>

The Child Assessment Observation Form is based on the Child Assessment Certification Form developed by Westat for certifying NRS assessors.<sup>5</sup> For each item in the NRS child assessment (including the warm-up section and the practice items), the observation form requires the site visitor to record (1) the child's response to the question; (2) the types of errors, if any, the assessor made in administering the item; (3) comments on any problems observed; and (4) information on any procedural errors not covered by Westat's certification process. Types of administration errors include straying from the script (using unscripted words or paraphrasing), coaching the child, providing non-neutral encouragement, employing incorrect or imprecise pointing or hand-sweeping motions, and using "a" and "the" or pronouncing words incorrectly in administering PPVT items. Procedural errors not included on the certification form include such errors as not setting up for the assessment before the child's arrival and not using probes correctly.

The interview guides for program directors and lead NRS trainers included questions about:

1. Scope of the program's NRS assessment activities
2. Preparation for, and staffing of, the NRS assessments
3. Approach to local assessor training
4. Conducting and monitoring the assessments
5. Children's responses to the assessments

<sup>5</sup> We developed two versions of the Child Assessment Observation Form—one for observing the English-language version of the NRS assessment and one for observing the Spanish-language version.

6. Problems or issues that arose, and how they were addressed
7. Experience with the CBRS
8. Communicating with parents about the NRS
9. Costs of implementing the NRS
10. Program plans to use the assessment results
11. Usefulness of the NRS growth report on 2004 results
12. Concerns and suggestions for improving the NRS

The focus group guide for staff members who conduct the assessments was designed to obtain descriptions of, and comments about, assessor training and certification along with assessors' experiences conducting the NRS assessment. In addition, in the wrap-up to the focus group, we asked for feedback on the assessment instrument and process and solicited suggestions for improving the NRS.

We also developed a site report template for site visitors to complete after each visit. This template identifies and organizes the information that site visitors were expected to collect during each visit, ensuring that consistent information was reported across the visits and facilitating analysis of the site visit data.

### **Qualifications and Training of Site Visitors**

A team of 15 site visitors conducted the spring 2006 site visits. This team had extensive experience in conducting site visits and observations in a wide range of early childhood program settings. Team members possessed a diverse set of skills, including being certified to conduct NRS child cognitive assessments, and skilled in conducting interviews and focus groups with program staff, and providing quality assurance and assess program implementation. Five of the site visitors were fluent in Spanish. Many staff members who conducted spring 2006 site visits had also conducted visits in fall 2004 or spring 2005 and some had attended a regional NRS training conference organized by Westat and Xtria in either Atlanta or Denver in summer 2003 or in Reston, Virginia, or San Francisco in summer 2004.

To ensure that all site visitors would be well qualified to carry out the site visit activities—observing and scoring the NRS assessment administered by Head Start staff, conducting a focus group with local assessors, and interviewing senior program staff—MPR conducted a two-day training for all site visitors in March 2006, prior to the spring 2006 site visits. The training included practice using all of the observation and interview tools, as well as coding and discussion of videotaped administrations of the NRS child assessment. Immediately after training, all site visitors independently reviewed a videotaped administration of the NRS assessment and completed a Child Assessment Observation Form about it. Senior project staff reviewed scoring and error coding to ensure that assessors were reliable.



---

## Site Visit Activities

Spring 2006 site visits were conducted between April 18 and May 25. Most visits were completed by a single site visitor; however, observers new to the NRS accompanied an experienced observer on one visit before conducting any visits on their own. Each visit included three primary activities: (1) assessment observations, (2) individual interviews with selected staff (Head Start director and lead NRS trainer), and (3) a focus group with local NRS assessors.

Site visitors conducted most visits in two days, with approximately one day devoted to observing assessments and one day to conducting the interviews and the focus group. In a few sites, an additional day was needed because selected centers were geographically dispersed or because 15 assessment observations were conducted.

## ANALYTIC METHODS

The data obtained during site visits were collected by a large team of site visitors and had to be processed and analyzed in a short period of time. To ensure the quality of the data, we subjected all observation forms and narrative reports to a quality control review as soon as possible after each visit. We then data-entered assessors' scores, observers' scores, and error codes. Narrative reports were compiled in a database using Atlas.ti qualitative analysis software and coded (Scientific Software Development 1997). We constructed a separate database using Microsoft Access to track completion of assessment observations and record the reason when an observation could not be completed.

To analyze the data collected on observed assessments, we computed descriptive statistics, such as frequencies and means, on various types of errors in assessment administration (for example, deviating from the assessor script, coaching, providing non-neutral encouragement, or incorrect gesturing) and aggregated them by program, assessor, and assessment scale. We also computed a certification score for each assessment following the method employed for the Child Assessment Certification Form and estimated an intraclass correlation coefficient (ICC) for each assessment subscale, to assess inter-rater reliability across Head Start assessors and MPR site visitors.

Because the NRS is being implemented in all Head Start agencies throughout the country, we wanted our descriptions of the experiences of 35 programs to be generalizable to all programs. To accomplish this, we used statistical properties of the sampling process to create weights for the observation data. These weights essentially allow us to project the number of programs or children who would have responded in a particular way, given what we know about our sample. We analyzed the data using the weights.<sup>6</sup>

---

<sup>6</sup> To construct the weights for programs, we first calculated the probability of selection of each. If we released both programs in a pair due to ineligibility or noncooperation of the main release within the pair, we account for it at this stage. The program-level sampling weight then, is the inverse of this probability. We also constructed a nonresponse adjustment within pair for those programs that did not cooperate. Within programs we selected centers by calculating the probability of selection of each center (grouped by location) and home visitor. Similarly, the inverse of these probabilities are the center- and home visitor-level sampling weights. To

*(continued)*

After reviewing a set of initial site visit reports, senior members of the project team developed a coding scheme according to key themes and topics covered during the site visits (Table I.3). Two experienced project team members then coded all the narrative reports according to this scheme using Atlas.ti software. To ensure reliability across coders, both team members coded an initial set of reports and compared the codes. In addition, to check reliability as coding progressed, one team member reviewed a subsample of the coded reports.

In findings reported from the qualitative data, our unit of analysis is the program, and thus all programs are given equal weight, regardless of the number of respondents interviewed. In the text and tables, we present the percentages of sample programs that reported a particular approach to implementing the NRS or a particular response to a question. While in some programs we interviewed more than one lead trainer, and in most programs we talked with more than one assessor, we report all findings as a percentage of programs rather than as a percentage of all staff who participated in the site visit interviews and focus groups.<sup>7</sup> Throughout the report, we note instances in which different staff members within the sample programs disagreed about a particular topic, although this was a rare occurrence.

## **ROADMAP TO THE REPORT**

We now turn to describing the results of our assessment of NRS implementation in spring 2006. In Chapter II, we describe the quality of the child assessments we observed, along with staff experiences in administering the assessments. Chapter III reports the approaches that programs took in training and certifying staff to conduct the assessments. Chapter IV describes programs' approaches to implementing the NRS, including coordinating and staffing the assessments and communicating with parents. Chapter IV also reports what program staffs told us about their costs for NRS implementation and their experiences using the CBRS. In Chapter V, we describe ways in which programs have used the NRS results for program improvement, how they plan to use their NRS results along with the results of local assessments, their future plans for using the results, and their reactions to the NRS growth reports on outcomes for their programs. In Chapter VI, we summarize what we learned from the on-site interviews and focus groups with program staff regarding the contribution of the NRS to their programs, their concerns about the NRS, and their suggestions for improvement. Based on findings from the spring 2006 site visits, in Chapter VII we synthesize recommendations for NRS system improvement.

---

*(continued)*

select children, we calculated the probability of selection of center- and home-based children, again using the inverses as the sampling weights at the child level. The last step was to adjust for child-level nonresponse and ineligibility. Each selected child's final disposition code was classified as (1) complete, (2) eligible noncomplete, or (3) ineligible. We multiplied the three weighting components together to get a cumulative sampling weight at the child level. We formed nonresponse weighting cells based on census region, proportion black or Hispanic population, and metropolitan status, and used the inverse of the weighted response rate within cell as the nonresponse adjustment factor. We multiplied this factor by the cumulative sampling weight to get the final child-level weight.

<sup>7</sup> In other words, we combined all responses from each program and counted them as one response, then calculated the percentages at the program level, unless otherwise noted in the tables.

**Table I.3. Codes Used to Analyze Qualitative Data Collected During Site Visits, by Category**

Protocol Source
Head Start Director Lead NRS Trainer Assessor Focus Group Site Visit Summary Assessment Observation Summary
Program Characteristics
Center-Based Only/Some Home-Based ACF Region Metro/Non-Metro Enrollment of 1 to 200 Children/Enrollment of More than 200 Children
Computer-Based Reporting System
CBRS Staffing Program's Experience Using the CBRS
Implementing the NRS Assessment
Assigning Staff to Administer the Assessment Assigning Spanish-Language Assessors Assessors' Experiences Conducting the Assessments Children's Responses to the Assessment Process Assessing Children with Disabilities Assessing English-Language Learners (non-Spanish) Experiences of Spanish-Language Assessors Experiences of Teacher-Assessors Communication with Parents Costs of NRS Implementation Implementation Errors Scheduling Assessments Conducting Follow-Up Observations of Assessors Tracking Progress in Completing Assessments Update on Program Operations
Training
Approach to Local Training Adequacy of Local Training National Training Materials and Guidance Suggestions for Local Training Suggestions for National Training Materials Most Useful Training Components Training Needs Other Training Issues
Lessons and Implications
Local Assessment Process Reactions to the 2004–2005 NRS Growth Report How NRS Data Are Used Use of NRS Results with Those of Local Assessments Changes to Staff Training Changes to Classroom/Curriculum Local Tracking of Individual Scores Future Plans to Use the NRS Suggestions for Improving the NRS Suggestions for Improving the NRS Reports Contributions of the NRS Contributions About the NRS Other Issues Raised

**This page has been intentionally left blank for double-sided copying.**

## CHAPTER II

### ADMINISTERING THE CHILD ASSESSMENT

---

As in the past, evaluating the implementation of the NRS child assessment was the central focus of our spring 2006 site visits. Identifying patterns of errors in administration and scoring is a crucial step, not only for understanding the quality of assessment administration but also for improving the assessment, since some errors may indicate a need for more training and guidance. Also, items that are difficult to administer may need to be modified or removed from the assessment battery. In the time between previous data collection rounds, information obtained during site visits has led to such changes as adding acceptable alternative responses, simplifying gestures and clarifying instructions, and modifying questions administered in the battery.

In this chapter, we describe the overall quality of assessment administration in spring 2006 and make some comparisons to results from earlier rounds. As in previous reports for the Quality Assurance Study, we based our analyses on the extent to which the administration of the assessments we observed met certification standards, on measures of inter-rater reliability, and on the types and frequency of administration and scoring errors. Unlike previous NRS rounds, however, we present weighted, rather than unweighted, percentage estimates of error data in the main chapters, unless otherwise noted. We examined these issues for our observations of both the English- and Spanish-language versions of the assessment.<sup>1</sup> Finally, we describe assessors' experiences conducting the assessments based on interviews and focus groups held with assessors, lead NRS trainers, and Head Start program directors.

We conducted tests of statistical significance for the spring 2005 to spring 2006 changes for the following key measures: total errors of each major type (script errors, non-neutral

---

<sup>1</sup> However, it is important to note that our observation sample for the Spanish assessment is just 70 observations (65 of which completed all five sections), and thus should not be used to generalize about Spanish assessments nationally.

encouragement, coaching, incorrect gestures, pronunciation errors, and scoring errors), along with the mean certification score, and we note significant differences below.

## APPROACH TO EVALUATING THE QUALITY OF THE NRS ASSESSMENTS

During our site visits, as described in Chapter I, we examined the quality of assessment administration by conducting structured observations of a sample of assessments (364 assessments administered by over 110 different assessors across 35 sample programs<sup>2</sup>). During these observations, trained site visitors scored children's responses along with the assessors and coded errors in administration using the Assessment Observation Form that MPR developed (described in Chapter I and reproduced in Appendix A). We designed this form to replicate the Child Assessment Certification Form developed by Westat for the purpose of certifying local Head Start staff to administer the assessment. The form documented three types of test administration errors:

1. **Scoring Errors.** Site visitors recorded a score for each item on the Assessment Observation Form and obtained a copy of the assessor's scores. We compared scores to identify scoring errors; we treated our site visitors as though they were trainers and accepted their scores as correct. See Chapter I for details on site visitor training.)

2. **Administration Errors.** The Assessment Observation Form recorded the frequency of five types of administration errors, by item: (1) straying from the assessment script, (2) coaching, (3) non-neutral encouragement, (4) incorrect pointing and hand-sweeping gestures, and (5) pronouncing words incorrectly and/or inserting articles ("a" and "the") in the vocabulary section (see Box). These types of errors are also counted in the certification score.

3. **Procedural Errors.** The Assessment Observation Form also captured procedural errors not included on the Certification Form. For example, it captured any errors in pre-assessment activities, such as not entering the child's

### Types of Administration Errors

**Straying from the Script:** Any deviation from following the script verbatim.

**Coaching:** Encouraging the child to change his or her answer, such as repeating questions when the child answers incorrectly; hinting with hand gestures, eye movements, or words; placing a finger on the correct plate; or making comments such as "You know the answer," or "Do you want to try again?"

**Non-neutral Encouragement:** Giving praise for correct answers or telling the child he or she answered incorrectly, such as "That's right!" "You are so smart!" or "That's wrong."

**Other Errors:** Pronouncing words incorrectly or inserting articles ("a" and "the") in the vocabulary section; making incorrect hand gestures, such as pointing instead of circling, circling incorrectly, forgetting to point, or unscripted pointing; and failing to slow the child down when necessary.

<sup>2</sup> Two programs were from Puerto Rico, and consequently only appear in our analysis of Spanish error data.

---

and assessor's ID numbers, not recording the date of the assessment, failing to record information in the language screener sections, not setting up the easel or preparing materials before the child was present, and not having the sheet of paper ready for the two Simon Says items that require it. It also included other procedural errors, such as not providing appropriate prompts if the child stopped identifying letters in the Letter Naming task. We present frequencies of these procedural errors in text boxes later in this chapter.

In addition to observing a sample of child assessments during the site visits, we conducted focus groups with assessors (including those we observed and some we did not) to learn about their experiences administering the NRS child assessments. We also conducted interviews with Head Start directors and staff who trained assessors (see the focus group and interview protocols in Appendix B). We supplemented our findings from the assessment observations with information from these interviews and focus groups to provide additional insights into possible reasons for the errors we observed. Chapter III, which discusses implementation of the spring 2006 refresher training, provides further insight into possible sources of errors.

The next three sections describe the quality of the sample of 294 English assessments we observed.<sup>3,4</sup> First, we report on how the assessors would have been rated, had our observation been their certification. Second, we report on how accurately assessors scored the assessments, based on an analysis of inter-rater reliability. Third, we present details of the procedural, administration, and scoring errors observed for each segment of the assessment. In the fourth section, we describe assessors' performance on the English assessments and we report similar information for the Spanish-language version of the NRS assessment for the 70 Spanish assessments observed. In the fifth section, we describe assessors' views on children's reactions to the NRS and challenges in administering the child assessments.

## MEETING THE CERTIFICATION STANDARD

In consultation with the Office of Head Start, Westat developed the procedures and standards for certifying assessors to administer the NRS child assessment. To determine whether a trainee meets the standard, a certified trainer observes the trainee administering the assessment with a child (or, if necessary, an adult playing the role of the child) and scores the assessment along with the trainee. The trainer records the frequency of scoring errors and of the four types of administration errors on the Child Assessment Certification Form.

---

<sup>3</sup> The primary focus of this report is the quality of observed NRS assessments in spring 2006. Nevertheless, we make some comparisons between assessment quality in fall 2003, spring 2004, fall 2004, spring 2005, and spring 2006 to indicate whether various aspects of assessment quality may have changed from the first year of the NRS. For detailed data on the implementation of previous rounds of the assessment, see Paulsell et al. (2004, 2005, 2006).

<sup>4</sup> As discussed in Chapter I, we present weighted estimates of assessment errors in this chapter. Appendix C also provides weighted estimate tables of question-by-question assessment errors for both English and Spanish assessments.

We determined errors in scoring by comparing trainer and trainee scores; accepting trainers' scores as the correct ones. After the observation, we sum total administration and scoring errors by type. We calculated the number of errors for each of five error categories (straying from the script, coaching, non-neutral encouragement and other administration errors,<sup>5</sup> and scoring errors).

Depending on the total number of errors in each of the five error categories, we assigned a rating of 1 to 5 (see Box). For example, a rating of 5 represents 0 to 2 errors and a rating of 1 represents 12 or more errors. Next we multiply each of the five ratings by 4 and sum these results to produce a certification score between 20 and 100. To become certified, trainees must obtain a score above 85.

**Certification Ratings**

<b>5:</b>	0–2 errors
<b>4:</b>	3–5 errors
<b>3:</b>	6–8 errors
<b>2:</b>	9–11 errors
<b>1:</b>	12 or more errors

During our site visits, MPR and J&A staff observed Head Start assessors as though they were “certifying” them. The certification analysis indicates that in spring 2006 programs did a reasonably good job of administering the assessments correctly and consistently. Across the 293 English assessments with full data, the mean certification score was 92, indicating that, on average, administration of the assessments exceeded the certification standard by seven points.<sup>6</sup> Furthermore, 83 percent of observed assessments achieved a certification score of above 85, which was the minimum score required for certification (Table II.1). Of the 17 percent below a score of 85, nearly 5 percent achieved a score of 84, only one point below the passing score.<sup>7</sup> Thus, about 88 percent of all observed English assessments either met the standard or achieved a score one point below the standard. More than half (57 percent) fell in the 96-to-100 range. Across the 33 programs with full data, 6 programs had a mean certification score for observed assessments below 85.<sup>8</sup> Of the 94 English assessors we observed with complete data, 16 had a mean certification score of less than 85. Among these 16, mean scores ranged from 53 to 84; half the assessors had a mean score ranging from 80 to 84 points.<sup>9</sup>

<sup>5</sup> Other administration errors include incorrect gesturing, incorrect pronunciation, and inserting articles before the PPVT items.

<sup>6</sup> By “full data,” we mean that we have data on the entire NRS assessment. All or part of the scoring data were missing for one case, so certification scores and average scoring errors could not be calculated. In this case, the child did not pass the language screener, and thus only scores for Sections A and B were available. In results presented below, we use the slightly smaller sample of 293 to examine the certification score and the level of scoring errors but use the larger sample of 294 to examine other types of errors.

<sup>7</sup> Due to the structure of the scoring formula, possible assessment scores result in point intervals that are divisible by 4—such as 80, 84, 88, 92, and so forth. An assessment with a certification score of 84 would have received a rating of 4 (3–5 errors) on four of the error categories and a rating of 5 (0–2 errors) on one category. Therefore, the total number of errors made on this assessment would have been in the range of 12 to 22.

<sup>8</sup> The full sample for the spring 2006 Quality Assurance Study was 35 programs. Data from two monolingual Spanish programs in Puerto Rico, however, are only included in the discussion of Spanish certification scores later in this chapter.

<sup>9</sup> It is important to note that the total number of potential errors on certification forms is quite high – 90 potential scoring errors and 224 administration or procedural errors.



**Table II.1. Distribution of Certification Scores Across Observed English Assessments**

Certification Score	Number of Assessments	Percentage of Assessments	Cumulative Percentage of Assessments
44	3	0.6	0.6
48	1	0.4	1.0
52	1	0.2	1.2
56	1	0.2	1.4
60	0	0.0	0.0
64	4	0.9	2.3
68	2	0.3	2.6
72	7	1.6	4.2
76	6	3.4	7.5
80	13	5.0	12.5
84	22	4.9	17.4
88	30	9.1	26.5
92	53	16.9	43.4
96	80	30.5	73.9
100	70	26.1	100
<b>Total</b>	<b>293</b>	<b>100</b>	

Source: Spring 2006 Observations of English NRS child assessments

Note: Possible assessment scores result in point intervals that are divisible by 4

N = 293 (one child did not pass the language screener). Percentages are weighted estimates.

Overall, the certification scores for the spring 2006 assessments were significantly lower than certification scores on assessments observed in spring 2005.<sup>10</sup> The mean spring 2006 weighted certification score was 92 in spring 2006 compared with 96 in spring 2005. More than half (57 percent) of spring 2006 assessments had certification scores of 96 or 100; however, in spring 2005 fully 80 percent of assessments fell within this point range. Moreover, 8 percent of spring 2006 assessments had mean scores below 80, but only 2 percent of spring 2005 assessments were below 80 points.

Although not a criterion for certification, the time taken to complete the assessment is important to the Head Start programs, since it is related both to operational and to cost issues. The full NRS assessment (sections A to E) was designed to be administered in 15 minutes. During spring 2006 observations, site visitors recorded the start and stop times of the assessments on the Assessment Observation Form. The average duration across all observations in which the child passed the language screener—and thus completed the entire assessment—was 15 minutes. Assessments ranged from 8 to 40 minutes; 93 percent lasted

<sup>10</sup> The difference in the mean certification score between spring 2005 and spring 2006 is statistically significant at  $p < .01$ .

between 10 and 20 minutes. The average duration of observed assessments was similar to previous rounds of site visits, which were: 14.8 minutes in spring 2004, 10.4 minutes in fall 2004, and 15.3 minutes in spring 2005. Full assessments conducted by classroom teachers lasted 2.6 minutes longer than those conducted by other staff (the difference in minutes, 16.8 versus 14.2, is statistically significant,  $p < .01$ ). One possible explanation is that children feel more comfortable around a familiar adult and are more likely to try to engage them in conversation (requiring more redirection).<sup>11</sup> Alternatively, teachers may be more knowledgeable about the amount of wait time needed for children to respond.

### INTER-RATER RELIABILITY

We examined the reliability of total scores for each section of the NRS by calculating an intraclass correlation coefficient (ICC) for each scale (Language Screener, Vocabulary, Letter Naming, and Early Math Skills).<sup>12</sup> The ICC estimates the proportion of total variance of scores that is due to variance of the scores across children (rather than the variance across the Head Start assessor and MPR observer scores). Estimates of reliability using the ICC indicate that the inter-rater reliability of all the subscale scores between Head Start assessors and MPR observers is high (see Box). Findings are very similar to previous rounds of the Quality Assurance Study.

Intraclass Correlation Coefficients for NRS Scales	
Language Screener	0.93
Vocabulary	0.97
Letter Naming	0.99
Early Math Skills	0.98
Counting (Item E20)	0.91
N = 293 English assessment observations	

<sup>11</sup> Teacher-assessors have reported in focus groups since the beginning of the Quality Assurance Study that they often need to redirect those children from their classrooms back to the NRS assessment because they want to chat instead about various topics. As discussed in Chapter IV, there was no significant difference between teachers and non-teachers in the likelihood of committing a coaching error.

<sup>12</sup> We calculated an ICC for the counting item (E20) separately because it is scored differently compared to the rest of the Early Math Skills items. Items E1 through E19 are scored as correct or incorrect, while the score for item E20 is the highest number that the child counted correctly.

## ERRORS IN PROCEDURES, ADMINISTRATION, AND SCORING

During the assessment observations, site visitors coded errors made in administration for each item in the battery, including the assessment items, the introduction sections preceding each segment, and the practice items. Approximately seven percent of

	Fall 2003	Spring 2004	Fall 2004	Spring 2005	Spring 2006
Straying from the script	3.1	1.5	1.3	1.1	2.1
Coaching	0.6	0.6	1.5	0.8	1.1
Non-neutral encouragement	0.4	0.6	0.5	0.3	0.9
Incorrect hand gestures	1.9	2.6	2.7	1.9	3.1
Inserting articles	0.1	0.1	0.1	<0.01	0.1
Scoring errors	5.6	2.3	2.9	2.1	2.2
N =	346	300	286	316	294

Sample sizes for the scoring errors are slightly lower than for other errors because these were calculated using only full assessments (sections A-E completed). These sizes were 346 (fall 2003), 297 (spring 2004), 279 (fall 2004), 305 (spring 2005), and 293 (spring 2006). All but fall 2003 are weighted estimates.

assessors whom we observed had administered NRS assessments for the first time in spring 2006. As with previous rounds of the NRS, the overall error rate on English assessments was low in spring 2006, although several types of errors increased since the spring 2005 round. Assessors had difficulty administering and scoring some items, but these difficulties tended to cluster non-randomly, in specific sections or items of the assessment, suggesting the need for changes in these areas, which we will detail below. In spring 2006, assessors made a small number of errors, on average, in scoring, giving non-neutral encouragement, coaching, and straying from the script; gesturing errors were the most common (see Box next page). The average number of errors in spring 2006 was generally similar to both fall 2004 and spring 2005. Errors from the spring 2006 round of assessments differ from the initial round in fall 2003. Script and scoring errors have decreased from the initial roll-out, while instances of coaching, giving non-neutral encouragement, and using incorrect gestures have increased (although they remain quite small). Again, it is important to note that the total number of potential errors on certification forms is quite high – 90 potential scoring errors and 224 administration or procedural errors.

For this round of the Quality Assurance Study, we examined the differences in mean error rates between spring 2006 and spring 2005 more closely. In both rounds of spring assessments, assessors, on average, scored more than 97 percent of responses correctly (differences between rounds were not statistically significant). Scoring errors sometimes occurred when assessors had difficulty seeing where the child was pointing or when the child's response was ambiguous. Errors also occurred on some items that required a combination of careful reading of the script, gesturing, observing the child's response, and, finally, recording a score. The counting item in Early Math continues to be the item that assessors score incorrectly most often; on more than one quarter of the assessments, this item was scored incorrectly (27 percent in spring 2005 and 26 percent in spring 2006). Scoring errors have decreased by more than one-half since the NRS began in fall 2003, most likely due to increased experience and comfort with facilitating the assessment while

simultaneously recording the child's answers (see Box). However, while scoring errors have remained constant since spring 2005, most types of administration errors increased between spring 2005 and spring 2006. Straying-from-script errors rose from 1.1 errors per assessment in spring 2005 to 2.1 errors in spring 2006, a difference that is statistically significant ( $p < .01$ ). Although the difference is significant, straying-from-script errors are still relatively low occurrence events. Observers reported that few assessors strayed from the script in substantial ways.<sup>13</sup>

Similarly, coaching errors, instances of non-neutral encouragement, and gesturing errors also increased over time. Each spring 2006 assessment contained, on average, 0.3 more coaching errors, 0.6 more non-neutral encouragement errors, and 1.2 more gesturing errors than spring 2005—all of these differences were statistically significant ( $p < .01$ ). Gesturing errors increased the most, despite changes made in the gestures and the instructions to simplify this aspect of assessment administration.<sup>14</sup> Coaching errors continued to be concentrated in the Simon Says section and usually involved nonverbal actions, such as looking under the table too soon, rather than verbal cues.<sup>15</sup> Assessments in spring 2006 also contained more non-neutral encouragement errors, relative to spring 2005. At times, several assessors within a program consistently made the same kinds of errors—a pattern which could suggest that trainers shared misinformation during the local NRS trainings (see Chapter III). In the rest of this section, we describe the frequencies and types of errors in each assessment section (see Appendix C, Table C.1, for item-by-item weighted results) and describe the frequency of some procedural errors noted on the Assessment Observation form but not included on the Certification Form (see text boxes throughout this section). All percentages presented in the rest of this section are weighted estimates. It is important to emphasize, as mentioned earlier, that overall error rates were relatively low. However, in order to provide information that can be useful for improving the training and materials (including the assessment itself)

**Percentage of Assessments with  
Observed Set-Up Errors**

Did not set up assessment in a quiet area	19
Did not fill out score sheet cover before assessment started	15
Did not fill out the date correctly	7
Did not place easel correctly or placed in an awkward spot	6
Did not set up area with materials before child arrived	4
Did not fill out child's ID correctly	3
Did not seat child within sight of a door for Simon Says	2
Did not have paper ready for Simon Says	1

N = 294 English assessment observations;  
weighted estimates

<sup>13</sup> In rare cases, these errors seemed to be due to nervousness about being watched by a site visitor; in others, script errors may have arisen from experienced assessors feeling very comfortable with the script and the assessment process, which resulted in taking some liberties with the wording.

<sup>14</sup> Starting in spring 2005, the types of gestures were reduced to two: pointing and circling. In addition, the instructions were made more consistent (for example, the assessor was always told to point for similar questions) and simplified (for example, assessors were directed to “circle the book with your finger” rather than to “point to the book by circling it with your finger”).

<sup>15</sup> Assessors are directed not to look under the table too early, so that the child is not given hints regarding what to do. Some assessors have difficulty timing this item, as they must look under the table to see if the child lifted his or her foot, as directed.

---

provided to programs in the future, we focus our discussion in the rest of this section on those few errors site visitors observed.

### **Administering the Set-Up and Warm-Up Sections**

The most common set-up errors observed were: (1) not setting up the child assessment in a sufficiently quiet area, (2) failure to fill out the score sheet cover page before the assessment started, and (3) failure to fill out the date correctly (see Box). While the majority of assessments we observed took place in an appropriately quiet area, site visitors noted distractions during some assessments that might have interfered with the child's ability to respond to the assessment items. These instances frequently included noise when assessments took place in offices and hallways; one site visitor observed a few assessments interrupted by the assessor's cell phone. (Other examples that program staff described during interviews and focus groups are discussed further in Chapter IV.) Overall, procedural errors during set up increased as compared to spring 2005 when, for example, only 10 percent of observed assessments were not set up in a quiet area, 3 percent did not have the score sheet completed before starting, and less than one percent did not have easels placed correctly.

The warm-up section is an introductory statement designed to make the child feel comfortable with the assessment; the script for this section has remained the same since the NRS began in 2003. Across all English observations, eight percent of assessors strayed from the warm-up script by misreading the script, omitting certain parts or choosing their own words to introduce the assessment to the child (Appendix C, Table C.1). Three percent of assessors strayed from the script in the warm-up section in spring 2005, less than half the rate observed in the current round.

### **Administering and Scoring the PreLAS Simon Says**

The PreLAS Simon Says task, based on the well-known childhood game, serves along with the PreLAS Art Show as a language screener to determine whether the child has adequate English proficiency to be assessed in English.<sup>16</sup> The task consists of an introduction, 2 practice items, and 10 assessment items; it is the same instrument used since the Year 1 assessments.

Errors in scoring, straying from the script, and coaching were most common in this section. On item A9, "Simon says point to the middle of the paper," scoring errors reached 11 percent, but they were no more than 4 percent on other items. Some assessors had difficulty seeing where the child was pointing. Others scored the item as correct when the child pointed to the middle of the easel, while others may have given a child credit if he or she pointed anywhere other than the edge of the paper, but not the precise middle or in the general proximity of the center of the page. Scoring errors for A9 were somewhat lower in

---

<sup>16</sup> In the Spanish version, a comparable task (*Tío Simón Dice*) serves as part of a language screener to determine whether the child should be assessed in Spanish.

spring 2005 (seven percent). Most script errors occurred in the introductory text (seven and eight percent), perhaps as assessors tried to make the child feel comfortable with the upcoming activities. In preparation for a subsequent Simon Says task, assessors are instructed to place a piece of paper near the child and say, “we’re going to use this in a minute.” Some assessors inserted “this paper” or “this piece of paper” into the sentence, which contradicts guidelines provided in the Assessor’s Guide.<sup>17</sup>

**Percentage of Assessments with Procedural Errors on the Language Screener**

Did not complete language screener section before continuing with the assessment	4
Did not choose correct assessment path	2
Did not choose correct bubble to indicate assessment path	1
N = 294 English assessment observations; weighted estimates	

Site visitors also noted several procedural/coaching errors not captured by the certification error codes (see Box next page). The two most common errors involved items in which assessors must look under the table to see if the child performed the command correctly. One-quarter of assessors either looked under the table too early, which could hint what the child should be doing, or did not look at all to verify if the child indeed performed the task for the item “Simon Says lift one foot.” Similarly, 15 percent of assessors demonstrated the same procedural error for the item, “Simon Says put your feet together.” After the item “Simon says pick up the paper,” nine percent of assessors either hinted by looking at, touching, or moving the paper, and three percent of assessors did not guide the child to put the paper down if needed after “Simon Says turn the paper over.”

Furthermore, a small percentage of procedural errors emerged during the practice section. On the first practice item, “Simon says look up,” assessors are instructed to model the item by looking up. For the second practice item, “Simon says look down,” assessors are trained *not* to model the action. Assessors failed to model the first item in five percent of assessments and incorrectly modeled the second item in eight percent of them. However, this represents a marked improvement in error rates for the second item as compared with spring 2005 (20 percent). The NRS training materials were revised for spring 2006 to further emphasize that assessors should not model the second Simon Says practice item, which may account for this decline in coaching.

### **Administering and Scoring the PreLAS Art Show**

The PreLAS Art Show task, the expressive vocabulary part of the Language Screener, consists of 2 practice items and 10 assessment items. In the English version, seven items require the child to name objects, such as a bee, a frog, and a pig. Three items require the

<sup>17</sup> The Assessor’s Guide is a manual updated annually that reviews assessment procedures and contains guidance on administering the NRS assessments. It is intended to be used at refresher trainings and distributed to each NRS assessor (see Chapter III).

child to tell the function of objects—a book, a cup, and a knife.<sup>18</sup> Overall, assessors made few errors of any kind on the object-naming items; seven percent made a gesture error on the bee item when assessors must be careful to point to the bee and not to the flower. Two of the three object-function items (B5 and B7) had much higher error rates. For those items, 16 and 13 percent of assessors incorrectly gestured by pointing to rather than circling the book and cup. Moreover, 34 and 33 percent of the assessors failed to point to the pictures a second time when they asked, “What can you do with it?” The prevalence of several of these gesturing errors increased substantially from spring 2005, when 19 and 20 percent of assessors failed to point to the items. For example, almost twice as many assessors in spring 2006 failed to point to items B6, B8, and B10 when asking “What can you do with it?,” suggesting that this technique should be stressed more in training materials.

Scoring errors ranged from one to six percent across these items, suggesting that assessors occasionally had difficulty determining whether a response should be accepted as correct for the function of a cup and a knife. The level of scoring errors was comparable to the previous spring. Nine percent of assessors strayed from the script.

At the conclusion of the Art Show section, assessors must add up the incorrect responses from the Simon Says and Art Show tasks. If the total number of incorrect responses exceeds 14, and the child does not speak English as a first language, the assessor is instructed to end the assessment.<sup>19</sup> Assessors did not complete the part of the form showing the results of the language screener before moving on to the vocabulary task for about four percent of assessments (see Box). This is a 50 percent decrease in this type of procedural error from the spring 2005 round, during which assessors did not complete this part of the form on eight percent of assessments. Two percent of assessors did not choose the correct path on the screener by either failing to continue or to stop the assessment appropriately.

### **Administering and Scoring the Peabody Picture Vocabulary Test (PPVT-III, Adapted)**

The vocabulary task includes an introduction, 4 practice items, and 24 assessment items. Items consist of plates with four pictures each. The child is instructed to point to one of the quadrants on each plate, which illustrate both things and actions, with a directive such as “point to fountain” and “point to delivering.” Some of the specific vocabulary words used in previous assessment rounds were replaced with words at a similar level of difficulty in spring 2006. As was the case in all previous rounds, assessors generally made few administration errors of any kind in the PPVT section, although administration errors had occurred even less often on the spring 2005 assessments. Most administration errors observed were script errors in the introduction and four practice items (Appendix C, Table C.1). Assessors strayed from the introductory script on four percent of observed assessments (compared to

---

<sup>18</sup> In the Spanish version, the Art Show (Exposición de Arte) contains no items about the function of objects; all items require children to identify the objects only.

<sup>19</sup> Starting in fall 2004, children whose home language is Spanish are assessed first in Spanish, then in English. Children whose home language is neither English nor Spanish are not assessed further if they do not pass the English language screener.

one percent in spring 2005) and on zero to three percent of assessments across the practice items (compared to zero to one percent in spring 2005). On two practice items, three percent of assessors incorrectly inserted an article (for example, “point to *the* ball” instead of “point to ball”). These kinds of errors rarely occurred during the scored (non-practice) portion of the PPVT. Vocabulary scoring errors, however, decreased over time, ranging from zero to four percent, compared with one to seven percent in spring 2005.

Assessors also made a small number of procedural errors on at least one practice item during the vocabulary section. On two percent of assessments, assessors did not read the script and model at least one practice item if a child gave an incorrect answer. On a little more than two percent of assessments, the assessor did not keep repeating one of the practice items until the child gave the correct answer. Similarly, on four percent of child assessments, assessors made errors with probes, such as not helping the child master pointing or not encouraging the child to look at all four pictures before giving an answer.

**Percentage of Assessments with  
Errors in Probing on the  
Letter Naming Task**

	Spring 2005	Spring 2006
First Plate	8	9
Second Plate	7	9
Third Plate	3	9

N = 293 English assessment observations (2006); 305 English assessments (2005); weighted estimates

### Administering and Scoring the Letter Naming Task

In the Letter Naming task, the child is presented with three panels or plates of upper- and lower-case letters (with eight to nine per plate) and asked to point to and name all the letters that he or she knows.<sup>20</sup> In spring 2006, as in spring 2005, assessors made few administration errors, such as coaching and giving non-neutral encouragement (Appendix C, Table C.1). Scoring errors were also comparable for the two spring assessment rounds. However, assessors made more straying from the script errors in the introductory section to each letter plate than in spring 2005 (7 to 14 percent versus 0 to 2 percent).<sup>21</sup>

During the Letter Naming task, assessors are instructed to provide prompts in response to various incorrect answers, such as the child’s giving the sound of the letter, giving a non-English name for the letter, or saying “zero” for the letter “O.” Nine percent of assessors failed to provide these probes appropriately on one of the three letter plates, which reflects an increase from spring 2005 (see Box). Additional errors included a few instances when

<sup>20</sup> The Spanish version contains four panels because four additional letters—CH, LL, Ñ, and RR—are included.

<sup>21</sup> The purpose of the introduction is to describe the letter naming task to children, explaining that they should point to the letters as they name them. In Year 3, test developers eliminated one sentence on the first letter plate: “Go slowly, and tell me the name of each one,” and simplified the transitions on the other plates. Some assessors may be used to some of the language from the previous version. For example, one observer described an assessor who often said “*Point to the letter and tell me the name out loud*” instead of simply “Tell me the name out loud.”



assessors did not ask the child to continue with this task,<sup>22</sup> could not help the child pace his or her pointing and identifying letters so that the assessor can score correctly, or did not score a letter correctly if it was named at least once. Observers also reported that about eight percent of the children (slightly higher than in spring 2005) commented on or reacted to the scoring.

### **Administering and Scoring Early Math Skills**

The Early Math Skills section includes an introduction and 20 items on counting, identifying numbers and shapes, comparing size, performing simple addition and subtraction, and reading graphs. One question about money that assessors criticized during the spring 2005 site visits was dropped from the NRS assessment.<sup>23</sup> (Other items were slightly different from the fall items.) As in previous rounds, the graph questions and the counting question were the major sources of errors, as well as the measurement item (E14) to some degree. In addition, script errors, coaching, and non-neutral encouragement errors increased throughout the Early Math section, for reasons that are unclear (Appendix C, Table C.1). Gesturing errors also increased, despite changes in Year 2 to make gesturing instructions more consistent.

Overall, scoring errors were relatively low and similar to error rates in spring 2005. Nonetheless, assessors continued to struggle with accurately scoring the counting task (E20), with more than one quarter of assessments scored incorrectly (27 percent in spring 2005 and 26 percent in spring 2006). Assessors continued to have difficulty scoring the item accurately when children either skipped marbles or counted the same marble twice. Procedural errors increased on this item as well. Assessors were slightly more likely not to include the required cue, “Please don’t jump around. Count the marbles in order, okay?” as well as not to encourage the child to keep counting when he or she stopped at the end of a row of marbles.

As in past rounds, incorrect gesturing was another substantial source of errors in the Early Math Skills section, although instances generally emerged among a handful of items. On item E5, assessors are instructed to sweep across the three nests while asking, “How many eggs are there altogether?” Gesturing errors increased from 2 percent in spring 2005 to 10 percent in spring 2006. On the fraction (pie slice) item (E16), gesturing errors

---

<sup>22</sup> Site visitors have observed children who name one or two letters and then stop. Assessors are instructed to encourage the child to continue naming letters by asking, “Do you know any others?”

<sup>23</sup> The nickel questions from the spring 2005 assessment (E13 and E14) had high levels of gesturing errors (12 and 13 percent). The first question (E13)—“Bobby has four nickels. His father gives him one more. How many nickels does Bobby have now?”—had errors related to gesturing (12 percent). The follow-up question (E14)—“How much money does Bobby have now?”—had the highest level of script errors of any question (16 percent of assessments had errors) and, in about 13 percent of assessments, also had gesturing errors. Site visitors reported that a few assessors simply skipped this question, because they thought it was too difficult for the children or because they did not realize some pages had more than one question. Others used the probe “How many cents is that?” even when the child had not responded “five nickels.” During the spring 2006 site visits, several program staff reported being pleased to see that the second part of this math task was eliminated from the assessment.

occurred on nine percent of assessments—more often than in spring 2005—even though the question and gestures were simplified in Year 2. Finally, assessors continued to have difficulty with administering the graph items (E18 and E19). Gesturing errors were observed in 27 and 31 percent of assessments, respectively, compared with 18 and 31 percent in spring 2005. These difficulties often resulted from assessors’ pointing to the pictures of the pets (dogs, cats, rabbits), rather than the names of the pets, as instructed. Assessors also demonstrated incorrect gesturing techniques on the number recognition items (E6, E7, and E8) as well, which increased from 2 to 4 percent of assessments in spring 2005 to 4 to 10 percent in spring 2006.

### **IMPLEMENTATION OF THE SPANISH-LANGUAGE VERSION**

Head Start children whose home language is Spanish are assessed in both English and Spanish for the NRS, provided that they pass the language screener for each version of the assessment. Starting in Year 2, NRS protocols now call for children whose home language is Spanish to be assessed on the Spanish version first, then the English version—a change from Year 1. This change was made in response to programs’ concerns that children who did not speak English well felt discouraged by the English assessment or that children had to “fail” before they could be assessed in their own language. Furthermore, the initial negative experience, staff felt, could reduce children’s scores on the Spanish-language assessment.

Sixty-nine percent (24 of 35) of programs in our sample reported that at least one NRS-eligible child spoke Spanish at home. Site visitors observed 70 administrations of the Spanish-language version (65 full assessments) across 11 programs and 18 assessors.<sup>24</sup> The sample for our study did not stratify programs according to their enrollment numbers of Spanish-speaking children. By chance, the sites selected for this round of observations had more Spanish-speaking children than the spring 2005 sample, so that the number of Spanish speakers observed was about twice as many as the previous round of the Quality Assurance Study.<sup>25</sup> Two out of 18 bilingual assessors observed were conducting NRS assessments for the first time in spring 2006.

While procedures for the language screener caused some confusion for a small number of sample programs in spring 2005, these kinds of issues were rarely encountered by site visitors in spring 2006. In fact, none of the observed assessors failed to choose the correct path after completing the PreLAS activities. However, the lead trainer from one program reported that while the program determines in which language to assess a Spanish-speaking

---

<sup>24</sup> We observed all children in our sample who spoke Spanish at home on the Spanish version of the assessment only. Spanish-speaking children were not over-sampled, but the enrollment lists were sorted by home language to ensure that some Spanish-speaking children would be selected.

<sup>25</sup> Comparing the NRS 2006 child sample with PIR data on all programs, the sample appears representative of the racial/ethnic composition of the population served. For example, in the 2005-2006 PIR, the average percentage of Hispanic children among all enrolled children is 27.6 percent; an average of 19 percent indicate their home language is Spanish. The 2006 NRS child sample is 26 percent Hispanic, and 19 percent Spanish-speaking at home. The NRS 2006 sample demonstrates an increase in children in Spanish-speaking families over the 2005 sample, from 11 percent in 2005 to 19 percent in 2006.

---

child according to parent report (as is the appropriate procedure), staff at that program will not administer the English version after the child is tested in Spanish—even if the child is fluent in English.<sup>26</sup>

Fluency of bilingual assessors, although a problem in spring 2005, was no longer an area of concern in spring 2006. Site visitors raised no concerns about the fluency levels of bilingual assessors they observed in spring 2006. They observed no incorrect recording of responses to *Nombre de las Letras* if a child gave an English name of a letter, such as (1) asking the child for the Spanish name or (2) recording it as an error, both of which go against NRS protocols. Previously, site visit observers reported that some Spanish-language assessors (mainly “heritage” speakers who learned the language in their homes, but had never studied the language formally) either were not very fluent in Spanish or knew Spanish well enough to speak effectively but not to read. They had difficulty reading or pronouncing Spanish words in the script, and, in rare cases, had difficulty understanding children’s questions in Spanish. Moreover, some assessors in spring 2005 were not aware that they could accept English responses for questions after the language screener and marked them wrong.<sup>27</sup>

### **Meeting the Certification Standard**

Across the 65 complete Spanish assessments observed (5 had missing data or the child did not pass the screener), the mean certification score was 97, 12 points above the minimum score of 85 required for certification. This mean represents a seven-point increase over the certification scores on the Spanish assessments in spring 2005, and is five points higher than the mean certification scores for English assessments in spring 2006.<sup>28</sup> Ninety-six percent of all observed Spanish assessments were completed by assessors with a certification score higher than 85 (Table II.2). Moreover, the distribution of certification scores on the Spanish-language version was less variable than the English version (see Table II.1).

The average duration of the Spanish assessment across all observations where the child passed the language screener (and thus completed the entire assessment) was 18 minutes, ranging from 11 to 27 minutes. In contrast, the English version lasted only 15 minutes, on average. In previous rounds, the Spanish-language assessment also took longer than the English version.<sup>29</sup>

---

<sup>26</sup> It was unclear whether staff from this program was uncertain about the procedures and thus would benefit from additional guidance on this issue, or if this decision was an attempt to minimize burden on children by only requiring assessment in one language for the NRS.

<sup>27</sup> Some assessors administered assessments in both languages.

<sup>28</sup> The difference in mean certification scores for English and Spanish assessments is statistically significant ( $p < .01$ ).

<sup>29</sup> Staff have remarked during interviews and focus groups that sometimes Spanish-speakers are more shy and difficult to ‘draw out,’ which may account in part for the longer assessment times.

### Errors in Procedures, Administration, and Scoring

Overall, the Spanish assessments were conducted with relatively few errors (Appendix C, Table C.2). As was the case for the English sample, Spanish assessors made very few errors by coaching or giving non-neutral encouragement. Overall, the Spanish assessors made fewer administration and scoring errors than the English assessors.<sup>30</sup> The most frequent types of errors were made in scoring and either mispronunciation or inserting articles (see Box). The mean number of errors on Spanish assessments for straying from the script, coaching, non-neutral encouragement, and scoring errors were lower than spring 2005 errors. Gesturing errors remained the same; and other errors, including inserting articles or mispronouncing words during the vocabulary section, increased slightly.

Mean Number of Errors per Spanish Assessment	
Straying from the script	1.1
Coaching	0.3
Non-neutral encouragement	0.2
Incorrect hand gestures	1.3
Pronunciation/inserting articles	1.5
Scoring errors	2.1
N = 70 Spanish assessment observations (65 for scoring errors); weighted estimates	

**Table II.2. Distribution of Certification Scores Across Observed Spanish Assessments**

Certification Score	Number of Assessments	Percentage of Assessments	Cumulative Percentage of Assessments
72	1	0.7	0.7
76	3	2.0	2.6
80	1	1.5	4.1
88	4	3.3	7.4
92	4	4.7	12.1
96	25	32.7	44.8
100	27	55.2	100
<b>Total</b>	<b>65</b>	<b>100</b>	

Source: Spring 2006 observations of the Spanish-language version of the NRS child assessment.

N = 65 (5 had missing data or did not pass the language screener). Percentages are weighted estimates.

### ASSESSORS' EXPERIENCES ADMINISTERING THE CHILD ASSESSMENTS

During site visits, assessors discussed their experiences administering the NRS assessment, describing aspects of the process that worked well and those that were difficult

<sup>30</sup> The difference in mean total errors between English and Spanish assessors for gesturing ( $p < .01$ ), giving non-neutral encouragement ( $p < .10$ ), coaching ( $p < .01$ ), and other errors such as inserting articles ( $p < .01$ ) was statistically significant.

---

for them or for the children. Examining assessors' experiences can help identify elements of the assessment where additional training and support may be needed, as well as sections or specific items that may warrant some revisions. In each round of the Quality Assurance Study, we have found many similar comments but also some slightly different perspectives from local programs.

In the rest of this chapter, we describe, as of spring 2006,<sup>31</sup> (1) assessors' reports on children's reactions to the assessment and how they responded to children's behavior, (2) the experiences of assessors who were also the children's teachers, (3) assessors' experiences administering the NRS to children who were English Language Learners, and (4) assessors' experiences administering the NRS to children with disabilities. The data are largely derived from the focus groups with assessors held at each site, with additional material provided from the NRS lead trainer interview and the Head Start director interview.

As in previous years, assessors felt that administering the assessments was much easier in the spring than in the fall, largely because the children were more comfortable, confident, mature, and knowledgeable about the information and skills being tested. Therefore, there was less need for probing and redirecting, and there were fewer awkward pauses if the child did not know the answer. When asked about how they dealt with behavior challenges, most assessors cited examples also mentioned in previous rounds, including redirection techniques, taking breaks, giving small rewards, or using a different assessor.

---

<sup>31</sup> While program staff was instructed by site visitors to concentrate on the experiences in spring 2006—unless site visitors asked them to comment on specific changes over time—it is possible that some information on experiences pertains to previous rounds of assessments as well. When staff described a specific incident and indicated the time frame, we omitted the information from our analysis here (for example, if an assessor recounts an experience with a child with disabilities from Year 1 or 2).

## Children's Responses to the Assessment Process

About half of programs (51 percent) reported that most NRS-eligible children responded positively to doing the assessment and were cooperative with assessors (see Box next page); programs rarely reported instances of children refusing to take the assessment in the spring.<sup>32</sup> Almost one quarter of programs (23 percent) said that children enjoyed the one-on-one attention with an adult. In fact, one focus group mentioned that younger children who did not yet qualify for the NRS

became jealous when they did not get to leave the classroom and play some games with the teacher. Some groups of assessors (14 percent) observed that children often were excited to go with the assessor to play games. At the same time, staff from one of these programs noted that while many children got very excited initially, about halfway through the assessment “a light bulb goes off and they realize this isn’t a game.” Nine percent of programs said that children were eager to show how much they had learned in Head Start, especially since they would soon be entering kindergarten.

Most assessors encountered few challenges with children who were shy or unresponsive in the spring as compared to the fall round, a change that assessors also noted in Years 1 and 2. Eighty-nine percent of programs said that the assessment went much more smoothly in the spring, for several reasons (see Box). Children had mastered skills over the course of the program year, and thus were less anxious or hesitant in giving responses. They were more confident overall, not only because they were familiar with the

Children's Responses to the Assessment Process	
	Percentage of Programs
Easier in the spring	89
Most children responded positively	51
Assessment was too long	49
Enjoyed the one-on-one attention	23
Excited about participating, playing games	14
Eager to demonstrate what they knew	9
N = 35 Head Start programs; unweighted estimates	

### Examples of Problem Behaviors

- Difficulty staying focused or on task
- Became bored or distracted
- Nervousness or anxiety
- Became frustrated when they did not know how to answer items correctly
- Reluctant because would rather be playing
- Concerned whether they gave the right answer; expected affirmation of performance
- Aware of scoring right versus wrong answers

N = 35 Head Start programs

<sup>32</sup> In calculating percentages, all programs are weighted equally, regardless of their size or the number of people interviewed at each. The common denominator is always 35 programs, regardless of how many staff we interviewed who may have expressed opinions on a given topic. For example, in response to the question “What do children like about the assessment?” some staff members revealed that children enjoy the personalized attention. This does not imply that children at the other programs do *not* enjoy one-on-one attention, but rather that no one from those programs identified this as something that children particularly like about the NRS child assessment. In general, specific concerns or activities are counted if at least one respondent at a program mentions them. Disagreement among respondents within programs is noted when it occurs; data are largely from the focus groups with assessors.

---

NRS process and screenings in general, but because they were familiar with the school environment as well as interacting with teachers and other staff members. With greater confidence and maturity, children were able to sit still and focus better during the spring assessment. Still, a few programs noted that sometimes children could be more restless in the spring, due to what one lead trainer described as “schoolitis” and a strong desire to be outside playing instead.

Of the 28 programs that estimated how long an average assessment takes to administer, the number of minutes across these programs was 16 (not including outliers, such as rare cases when a shy or difficult child took more than a half hour), which comes within one minute of the estimated length by the Office of Head Start and the average length of assessments observed in this round. Nevertheless, nearly half (49 percent) of programs reported that, in general, the NRS assessment was too long and that the length may have contributed to challenging behavior from certain children (see Box). Program staff worried that some children lost focus and became bored, tired, or frustrated, especially during the Vocabulary and Letter Naming sections. Criticisms of the length of these sections is a concern that has emerged repeatedly in previous assessment rounds as well, although one lead trainer acknowledged that the length is necessary to obtain validity. Assessors often used redirection during the vocabulary section (for example, “make sure you look at all the pictures”), since some children became bored. The Letter Naming task usually went more smoothly than in the fall, because children generally knew more letters. However, this reportedly could still be an arduous task for children who could not identify many alphabet letters by the spring assessment.

During the spring 2006 site visits, about half of the sample programs reported that some children exhibited challenging behaviors, such as becoming bored or distracted, wanting to chat during the assessment, displaying nervousness (one program said that children sometimes were afraid because they thought they were getting a shot from the nurse), or not wanting to interrupt their playtime (reluctance in the fall was more likely due to shyness). Staff from two programs noted that sometimes children became anxious about whether they were giving the correct answer and asked, “Am I doing this right?” or “What if I don’t know?” Some assessors from one of these two programs pointed out that when they maintained a neutral demeanor—as they had been instructed to do—some children assumed that they made a mistake because they did not get positive affirmation, as they did in the classroom, and consequently changed their answer. Another group of assessors noted that a few children were able to figure out when they were recording correct answers, and wanted to know what they should do if a child figured out that his or her answer is wrong. Finally, assessors from two programs described a unique challenge when administering the NRS assessments to children during home visits. Parents watched the assessments, which in turn may have influenced the child’s performance. For example, assessors reported parents coaxing their children or attempting to give them the correct answer, or reacting in a noticeable way to an incorrect response that encouraged the child to look at them and watch them react to each answer given.

To address these challenges, assessors reported using a variety of strategies, including those suggested in the training materials (for example, redirecting the child by saying his or

her name; taking a break and trying again later) and others that they developed, for handling the various ways in which children reacted to the assessment process (see Box). One strategy mentioned, allowing the child to sit in the assessor's lap, is not consistent with NRS protocol, occurred only in rare circumstances. In the two programs that described challenges when conducting NRS child assessments during home visits, staff took one of two approaches. One program learned to speak with all parents beforehand and explain that while they were welcome to watch, it was critical that they not react to the child's answers. Assessors from the other program asked the case manager, who was also present during the home visit, to escort the parent out of the room so that he or she could not influence the assessment score.

Most assessors felt quite comfortable facilitating the assessments and addressing these behaviors when they emerged; they had ample experience dealing with these situations in the classroom or in other work experiences (for example, as a disabilities specialist). Still, a small number of programs described difficulties they encountered, either with managing children's behaviors or facilitating the assessment. Three programs would like to be able to have more opportunities to use neutral encouragement and to be able to use a "warmer" and more inflective tone of voice that resembles what children are used to hearing in the classroom.<sup>33</sup>

Staff from one program did not feel prepared to keep children on task and respond to their comments because they felt they have received inconsistent messages about how much unscripted conversation was permitted during the NRS assessments. Assessors from another program noted that it was more difficult to assess children who were not from their own classroom and would like specific guidance on how best to engage children they do not know. Staff from two other programs found it difficult to respond to children who could not concentrate or became bored, noting that suggestions "It's okay to guess," "We can talk about that later" and

#### Assessors' Strategies for Addressing Children's Behavior

- Redirect attention using techniques such as saying the child's name, repeating question, assuring child that the assessment will end soon and they can play
- Try assessment later that day or on a different day
- Take a quick stretch, jumping, or singing break
- Spend time getting to know child
- Use different voice inflections to emphasize calming, reassuring tone
- Give out stickers at the end of the assessment
- Allow child's teacher to sit in room next to child
- Allow child to stand during testing
- Assign "problem" children to a more stern assessor; ask director to escort very active children into testing room
- Schedule next attempt with a different assessor
- Teacher accompanies child to testing room; talks about assessment days in advance to prepare children
- Child sits on assessor's lap (rare)<sup>a</sup>

N = 35 Head Start programs

<sup>a</sup>Not consistent with NRS protocols.

<sup>33</sup> Staff members from one of these three programs were still under the impression that they must follow the easel text exactly and cannot insert unscripted neutral encouragement, which makes it difficult to respond to behavior problems. There is no guidance in training suggesting that assessors cannot use a warm tone – just that they cannot give non-neutral encouragement.



“You’re doing a good job pointing” were not adequate to motivate and refocus children back to the easel.

During the spring 2006 site visits, assessors also noted children’s difficulties with the assessment, both during certain sections and with specific items. Overall, assessors reported that children had the most difficulty with specific items from the PPVT, followed by the Letter Naming and Early Math sections. Almost one in 10 programs raised concerns with the Language-Screener sections—Simon Says (PreLAS) and Art Show (PreLAS). The rest of this section describes particular items assessors reported that children had trouble with and portions of the assessment that concerned assessors (see Box).

**Simon Says (PreLAS) and Art Show (PreLAS).**

Once again, sample programs reported that children enjoyed the Simon Says and Art Show tasks overall. They viewed these sections as fun activities for children that served as a good warm-up to the rest of the assessment, as well as an effective language screener. Still, a few programs raised concerns about the PreLAS activities. Two programs said that the command “Simon Says put one hand on top of the other” was too confusing and abstract; it would be much better to use the language “put one hand on top of the other *hand*.” Staff from another program said that the two-step task involving the piece of paper (picking it up and turning it over) could be confusing for some children since they were not instructed by the assessor—or not instructed right away as some site visitors observed—to put the paper down. As such, it could be confusing to do the next command (“put one hand on top of the other”). This problem was also noted by one site visitor while observing assessments. One focus group felt that Simon Says was too easy and may have set children up for failure in subsequent sections of the assessment.

**Percentage of Programs Where Assessors Reported Difficulty or Concerns with Specific Tasks on the NRS**

PPVT	74
Letter Naming	49
Early Math	34
Simon Says	9
Art Show	9

N = 35 Head Start programs; unweighted estimates

On the Art Show, one group of assessors observed that the colored pictures of the children on the bottom of certain pages were distracting, since some children liked to start commenting on the images. Staff from another program felt that some words on the Language Screener were biased but did not identify any specific words. Finally, one program questioned whether the Art Show should come before the PPVT section as part of the language screener, or after. Some children have “receptive” vocabulary skills that are demonstrated by pointing to pictures when items are said to them (as in the PPVT), as opposed to “expressive” vocabulary skills that are demonstrated when children identify the pictures aloud in Art Show. Thus, if they did not pass the language screener, they would not have the chance to convey any receptive vocabulary knowledge.

**PPVT-III (Adapted).** Similar to the spring 2005 site visits, the PPVT task generated the greatest overall difficulty and criticisms among assessors; assessors in nearly three-quarters of the sample programs mentioned some kind of issue with the vocabulary section.

The most prevalent issue raised (15 programs) was the cultural appropriateness of the NRS assessment and whether it was regionally biased. Some of these programs felt strongly that the assessment was biased against those who lived in rural parts of the country. For example, children from rural areas would be much less likely to have seen a fountain, or they may not be familiar with the picture of *delivering* because some families pick up mail from a post office box or from a cluster of mailboxes in a central location. Another program explained that *diving* could be used colloquially to refer to someone jumping into water feet first.

In addition, staff in nine programs said that the pictures were confusing or misleading, and in some cases the “wrong” quadrants could represent correct responses; a few programs even wondered if the NRS test developers purposefully included these words to “trick” the children or determine if assessors were coaching them. For example, several of these programs said that children who had experienced a tornado—either personally or on television—were more apt to select the picture that resembled the destruction that results from the storm and not an active funnel. Some also noted that children were confused by *knight* because they often looked for its homophone, *night*. One program said that, in its view, several children missed *liquid* because one of the possible answers, ice cream, was explained to them in the classroom as “this is ice cream, you lick it.” These staffers believed that children heard “liquid” and thought “lick it,” and automatically chose the ice cream cone. Other confusing items included *delivering*, *surprised*, and *sanding*; assessors felt that each one had an additional quadrant that could also represent the word. Seven programs observed that children had difficulty with the black-and-white pictures and suggested having color pictures or photographs instead. Staff in five programs questioned how developmentally-appropriate some vocabulary words were for this age group, such as *peeking*, *selecting*, and *globe*. Five programs said that the PPVT was too long, and as a result children often became bored or restless.

**Letter Naming.** Nearly half of the programs (49 percent) reported difficulty in administering the alphabet task, which was comparable to findings from the spring 2005 visits. Eleven programs critiqued the layout of the panels. Specifically, they reported that there were too many letters on each panel—both in the quantity of different letters and presentation of upper- and lower-case letters. As in previous rounds, several programs advocated that assessors be allowed to point to each letter and ask, “What letter is this?”, or recommended reducing the number of letters per panel. This technique would ensure that children do not skip any letters that they know, as well as helping to focus children who may become overwhelmed by so many letters on one panel. Staff from four programs suggested that the Letter Naming section seemed to be testing children on their organizational skills and/or their ability to follow instructions, and not just alphabet knowledge.

Other concerns included having “A” through “E” on the first panel (leading children to expect the letters in order and to look for “F” on the first panel, then stop when they did not see it), and that it can be frustrating for children and assessors when a child does not know any or few letters but must listen to the assessor prompt several times, “Do you know any of these?” and “Do you know any others?” (One program acknowledged that this latter concern is less prevalent in the spring since children have learned some letters over the

course of the school year, but it remained an issue for some programs.) Staff from one program noted that since they used a phonics-centric curriculum, children tended to sound out the letter instead of naming it, even if they knew the name of the letter itself.<sup>34</sup> Finally, one focus group remarked that scoring the Letter Naming section would have been easier if the answer sheet mirrored the layout of the panels rather than a vertical list of letters.

**Early Math Skills.** Far fewer sample programs raised issues with the Early Math section in spring 2006 than in spring 2005 (34 percent versus 74 percent, respectively), which may reflect an effort to improve the math questions over time. For example, a few programs noted that the wording of the nickel question had improved; framing the question as “five nickels” (a counting task) instead of “25 cents” as it appeared in spring 2005 (a task about monetary value) was a vast improvement. However, nine programs strongly felt that certain questions continued to be too advanced for this age group and out of the scope of Head Start, such as the addition and subtraction items. One program explained that while it includes addition and subtraction in its curriculum, children learn these concepts with manipulatives and cannot perform the math functions in their head while looking at a two-dimensional easel plate.

Concerns about specific items mostly centered on the pie question (E16) and the graph questions (E18 and E19). First, some programs thought that these questions were too difficult for preschool children, and/or that the instructions were too wordy. Moreover, five programs pointed out that the pie had the same number of pieces that were missing and remaining (four). Some children counted the pieces that were missing instead; they got ‘credit’ for a correct answer even though they did not understand the concept of subtraction but rather demonstrated that they could count. Staff from one program reported that the measurement question (E14) still confused children about what they were being asked because assessors were no longer permitted to point to the teddy bear. Another program noted that while the wording of the counting question (E20) had improved (from “Ready, set, go!” to “Ready? Go.”), some children were still prone to race through the task as soon as they heard the word “go.” Assessors suggested changing the instructions to read, “Start now.”

### **Experiences of Assessors Who Were Also Children’s Teachers**

A little over half of the sample programs used at least some teachers as NRS assessors. In those programs, most teachers assessed children enrolled in their own classrooms. Teacher-assessors from eight of these programs said they gleaned additional insight while assessing children from their own classrooms. Some teacher-assessors cited components that the NRS targets that their local assessments do not, such as math skills and the alphabet. Others noted that it was useful to see the progress that children make over time, and whether they had mastered certain skills and information (for example, shapes and certain letters). If they determine that children cannot name a minimum number of letters or do not

---

<sup>34</sup> Head Start programs are instructed in training materials and on the easel that in these situations they should say to the child, “That’s the sound of the letter, what’s the name of the letter?”

seem to know many vocabulary words, they can adjust their lesson plans accordingly to ensure that children are better prepared for kindergarten. One teacher noted that administering the NRS allowed her to get a better sense of how much the quieter children have learned over time. Assessors from one program used the NRS to identify which topics or areas parents can work on with their children at home to reinforce classroom activities. Only one teacher-assessor acknowledged teaching specific items from the NRS assessment, in this case the word *globe*, a practice that raises concerns about “teaching to the test.”

Four programs reported that while the spring NRS assessments did not give them any additional information, the fall assessments provided a unique opportunity to spend some one-on-one time with each child. It allowed them to determine, for example, if children could follow basic instructions or which children needed extra challenges in class if they demonstrated skills that went beyond certain components of the Head Start curriculum (for example, if a child already could add and subtract). The teacher-assessors from only three programs said that the NRS child assessments did not give them any useful insights into the children in their classrooms that they did not already ascertain from local assessments.

The extent to which children’s performance on the NRS matched teachers’ expectations of their abilities varied, as in the past. Of the 16 programs where teacher-assessors commented on the children’s performance during the focus groups, assessors in six programs said that their expectations of the children’s skills usually lined up with experiences administering the assessments. In contrast, eight programs indicated that children tended to demonstrate less knowledge on the NRS assessment than what they had demonstrated in the classroom. Many teacher-assessors explained this variation by the child’s being nervous, anxious to return to playing with friends, not understanding the instructions, or having a behavioral issue that prevented them from paying attention to the assessment tasks. One teacher-assessor wondered if the validity of the NRS could account for these kinds of discrepancies. In one program, children did better on the NRS than teachers had expected, and another program’s staff did not clearly indicate the direction of discrepancy.

A handful of programs noted that this pattern of demonstrating less knowledge than the teacher believed they possessed could be particularly frustrating for staff, as programs in previous assessment rounds have observed. One teacher-assessor recalled a child who could not get any correct answers in the Letter Naming section but two days later could name several letters in circle time. She declared, “I *know* that some of them know this” material, and her colleagues agreed. Teacher-assessors from four programs reported that the extent to which a child’s performance matched their expectations varied—some did and some did not. These assessors also noted that children could surprise them and knew more than they had expected, such as knowing multiple alphabet letters or difficult vocabulary words, such as *surprised*.

Finally, a few teachers also noted that it felt odd not to praise the child during the assessment, and that some children became concerned when they did not hear positive feedback since they were accustomed to it in the classroom. Staff from one program said that children became confused when teachers seemed so stiff and formal while administering the NRS assessment, which was very unlike the way they usually acted in the classroom.

---

## Experiences Assessing Children with Limited English Language Skills

The majority of Head Start programs (24 out of 35 in our sample) had assessed or planned to assess English Language Learners; two programs were located in Puerto Rico where Spanish is the language of instruction. Among these, 22 programs had planned to assess at least one child in Spanish or had already done so, and eight programs had assessed or expected to assess English Language Learners who did not speak Spanish. As in spring 2005, the most common home languages identified by staff—other than Spanish—were Chinese and Arabic. Other languages included Bulgarian, Farsi, French, Haitian Creole, Hmong, Korean, Misteco (an indigenous language used in Central America and Southern Mexico), Mongolian, Navajo and other Native American languages, Nigerian, Punjabi, Russian, Somali, Urdu, and Vietnamese.

By the spring 2006 assessment round, programs reported that most English Language Learners were or would be able to complete the English version of the NRS assessment. Staff from five programs either expected or encountered a few children who did not pass the language screener in the spring, although it was unclear when these children enrolled in Head Start (i.e., an English Language Learner who enrolled later in the program year may be less likely to pass the screener in the spring). Moreover, assessors from two programs observed that some Spanish speakers who could not complete the fall 2005 English assessment were able to do so in the spring but, in contrast, could no longer pass the Spanish language screener.<sup>35</sup> This phenomenon is most likely due to children who were speaking English all school year and forgetting their Spanish, or children who were never very fluent in Spanish to begin with. While most sample programs did not raise concerns regarding the effectiveness of the PreLAS activities to gauge language ability and determine when it is appropriate to continue the NRS assessment, two programs strongly felt that the minimum number of items that a child must answer correctly in order to continue to the vocabulary task should be increased. Some staff members reported that a large percentage of English Language Learners passed the screener, only to struggle throughout the rest of the assessment. In other words, these two programs questioned the validity of Simon Says and Art Show to determine if a child had the comprehension skills in English to continue with the other activities. (These remarks primarily pertain to the first time that a non-English speaker takes the NRS assessment, regardless of whether it is in the fall or spring.)

Unlike previous rounds, almost no programs administering the NRS assessments to Spanish-speaking children expressed concern about the burden placed on them to complete the two assessments, especially when completed in a single session. Nevertheless, more programs in spring 2006 noted differences between the two versions; some staff felt the Spanish version was more difficult than the English one. Bilingual assessors from a few programs said that instructions for the Spanish assessment sometimes were more formal and confusing than for the English one. For example, one focus group noted that while children

---

<sup>35</sup> Head Start programs are instructed that children whose home-language is recorded as Spanish should be administered the entire NRS assessment in Spanish, regardless of how many errors they make on Tío Simón Dice and Exposición de Arte. It was unclear if assessors from these two programs stopped the Spanish assessment, or continued and just commented on the number of errors children made on the PreLAS activities.

were asked in English, “How many eggs are there altogether?” (E5), the Spanish equivalent item in Spanish (EE5) asked, “¿Cuántas uvas en total hay ahí?” (How many grapes are there in total?). Assessors said that the “en total” was more confusing than “altogether” for Spanish-speaking children. Staff from two other programs asked why children identified the function of certain objects in English in the Art Show (“What can you do with it?”) but were not asked to perform the same task on the Spanish version, and wondered if the two assessments were intended to test different skills. One criticism that has remained consistent over time between the two versions is the fact that the Spanish assessment includes four additional alphabet letters (CH, LL, Ñ, RR). Several programs strongly urged dropping these four letters from the Spanish assessment, arguing that aside from Ñ, these letters are no longer employed in modern Spanish usage. However, further exploration of this question by the OHS has shown that only RR is actually out of common usage; it is being eliminated from future assessments.

As has been the case since the inception of the NRS, a number of assessors noted that the Spanish-language version of the assessment does not use, nor does it permit assessors to use, alternate words if enrolled children speak a dialect of Spanish. Twelve programs indicated that the assessment should be more flexible and include more acceptable answers (for example, when children are asked “What is this?” in the Exposición de Arte), and that assessors should be permitted to use different words when saying, “Point to [item]” in the vocabulary section. Due to the format of the assessment, Spanish-speakers are often penalized, even if they know a word or understand a concept, because the term that is used by their family at home or in their family’s country of origin does not appear in the NRS easel, despite it being ‘correct’ and a sensible answer (see Box). Programs also had concerns about other words, remarking that either the picture or the concept of the picture was confusing. For example, in the TVIP, item CC4 asks the child to “Señala hora” (point to hour), and the correct response is a picture of a clock. Some assessors thought that the easel should instead read, “Señala reloj” (point to clock). Other items mentioned include Exposición de Arte item BB7 (the picture of the skirt looks like a lamp shade); *mecánico* (the gas station attendant looks like a mechanic); *ambulancia* (the picture represents an old-fashioned vehicle and children may not recognize it as an ambulance); *mueble* (this translates as “piece of furniture” and matches to a sofa, which some assessors think should be *sofa* or *sillón*); and *humano* (one of the incorrect quadrants, a monkey, is confusing because the drawing makes the animal look ‘humanized’—standing, smelling a flower, and smiling).

#### Examples of Differences in Spanish Word Usage

English Meaning	Accepted as Correct, Used in Directions	Used by Some Children, Scored as Incorrect
Fork	Tenedor	Cubierto
Snake	Culebra	Víbora
Doctor	Médico	Doctor

*Tenedor* and *culebra* have emerged as problematic words for some Head Start programs during previous rounds of site visits for the Quality Assurance Study as well.

## Experiences Assessing Children with Disabilities

Most programs in the sample (33 out of 35) assessed some children with identified disabilities during spring 2006. Similar to past rounds of assessments, speech and language delays emerged as the most common type of disability encountered by assessors, followed by cognitive and developmental delays, autism, ADHD, and unspecified behavioral or social/emotional problems. Other less common disabilities included partial hearing loss or deafness, Down Syndrome, visual impairment or blindness, fine or gross motor disabilities, physical health impairments (for example, a child who used a breathing/feeding tube), cerebral palsy, global development delay, bipolar disorder, and early onset schizophrenia.<sup>36</sup>

Programs identified several types of accommodations that they used in assessing

children with special needs (see Box). Many accommodations centered on children with speech and language delays. In these cases, assessors used tactics such as repeating a question, asking the child to repeat his or her answer, or reading the script more slowly to enunciate for the child. Two programs arranged for children with speech and language disabilities to be assessed by their own teachers because they were familiar with the children's speech patterns. For visually-impaired children, one program allowed a child to stand directly over the easel and laid it flat so she could see better, and one program allowed a child to use a magnifying glass. Assessors also frequently mentioned the benefits of splitting the assessment into different sessions on the same

### Accommodations Made by Head Start Programs for Children with Disabilities

- Split assessment into multiple sessions, take breaks
- Repeat questions or ask child to repeat responses
- Give extra encouragement and redirection to children with behavior issues
- Speak slowly, go slowly, and allow more time
- Consult with teacher to get a sense of the child's needs
- Use a quiet testing environment without distractions,
- Give an autistic child an advance tour of testing room
- Allow a visually-impaired child to use a magnifying glass or to stand directly over the easel
- Cover some letters for a visually-impaired child<sup>a</sup>
- Use a sign language interpreter for deaf child<sup>a</sup>
- Child with speech disability is assessed by own teacher
- Allow child with speech impediment to point to letters while assessor named alphabet letters one at a time<sup>a</sup>
- Alter script for child with a developmental delay<sup>a</sup>
- Stop assessment if a child cannot complete it
- Schedule assessment at optimal time (e.g., not recess)

N = 33 Head Start programs (two did not assess children with disabilities in spring 2006)

<sup>a</sup>Not consistent with NRS protocols.

day or on different days, particularly for children with ADHD or behavioral issues who had difficulty concentrating. Some programs made a point of giving extra encouragement and redirection to children with behavioral issues, refocusing them often and stressing the fact that the assessment would soon end. One assessor reported giving children with autism or a

<sup>36</sup> In our sample of 294 English child assessments, 31 observations (nine percent) were of children with an IEP, according to staff report. Six out of 70 observed Spanish assessments (10 percent) were administered to children with a diagnosed disability. Percentages are weighted estimates.

comparable condition a tour of the testing room before the scheduled day of the assessment. During this time, the child could ask questions and become accustomed to the testing environment, which hopefully would allow him or her to be more focused on the day of the assessment.

In contrast, staff from eight programs reported that assessors did not need to make accommodations for children with special needs. One director explained that children with disabilities usually performed well on the NRS and that one could not tell that they had a disability. A lead trainer from another program explained that children with severe disabilities that would require special accommodations are never assessed because these children are immediately placed in the public school system where they can receive more intensive early intervention services.

There were approximately four programs reporting that their assessors made an accommodation in violation of NRS assessment protocols. In general, the NRS training materials indicate that children with a severe disability, such as deaf, blind, or nonverbal children, may be exempted from the assessment but that other disabled children should be assessed. One program used a sign interpreter to translate during the assessment for a hearing-impaired child.<sup>37</sup> An assessor from another program covered several letters on each alphabet plate to allow a visually-impaired child to see the letters more easily. Moreover, a few programs described isolated incidents in which assessors simplified the testing process for children with a developmental delay. In one case, an assessor altered the script by using fewer words to help the child focus on what she was asking. For example, instead of asking, "How many eggs are there altogether?" she asked "How many eggs altogether?" At another program, an assessor noticed that a child had difficulty pointing to and naming letters simultaneously. As a solution, she pointed to each one and asked him to name the letter.

The extent to which assessors were able to complete the NRS assessments on children with disabilities varied considerably across programs. Among the 33 programs that assessed children with special needs in spring 2006, about one third reported that they successfully conducted NRS assessments on all such children and did not encounter any challenges. One assessor, however, reported that it could be difficult assessing a child with an IEP for speech and language if the assessor was not familiar with the child's speech patterns. Three programs described rare situations in which assessments were not attempted on children with severe impairments, including blindness, deafness, a breathing/feeding tube, and cognitive delays. Nearly one quarter of programs reported that staff began assessments with one or two children but then stopped and did not try again in the future. These assessors usually stopped after the language screener, due to developmental delays, speech and language disabilities, or behavior problems (for example, hyperactivity).

In contrast, staffers from one third of programs described situations in which they opted to complete an assessment with a special-needs child, some of whom may have been

---

<sup>37</sup> Using a sign language interpreter violates the NRS assessment protocol, which does not permit interpreters of any language. Children who cannot complete the assessment in English or Spanish should not be assessed.



eligible for an exemption according to NRS guidelines. These circumstances often resulted in frustrating testing environments for children and assessors, during which staff members described finishing the NRS the best way they could. For example, one assessor recounted an awkward assessment with a hearing-impaired child in which the child got all of the answers wrong. Some assessors were uncomfortable with assessing children with autism or cognitive delays (for example, Down Syndrome). Others struggled to understand children with speech impairments. One child suffered from expressive and receptive verbal problems, resulting in both the child and assessor being unable to understand each other.

Of the 33 programs that had experience assessing children with disabilities during spring 2006, 18 programs offered feedback on whether they had received sufficient training and support for assessing children with disabilities. Assessors from five programs reported that they had adequate guidance. Staff from two programs said that the video and Assessor's Guide provided good examples of different situations that they could encounter with special-needs children. One group of assessors explained that they consulted with their lead trainer about allowable accommodations or exemptions; another program deferred to its disability specialist to determine when the NRS would be an inappropriate assessment. Similarly, one group of assessors explained that its local training prepared them well for assessing children with special needs, but this was because the lead trainer also served as the disability specialist. In their opinion, the training video alone did not provide adequate guidance.

On the other hand, 13 programs reported that they received no information or not enough information on procedures for assessing children with disabilities and the accommodations that they could or could not make. While one program stated that the video effectively addressed how to handle uncooperative children, they felt that it did not address children with disabilities. Six of the 13 programs would like additional training and support, some noting that specific guidelines and examples of allowable accommodations in different circumstances would be very useful. A group of assessors acknowledged that while they knew about the technical assistance helpline, they wanted this information to be covered in their local training materials. One lead trainer noted that she did not know what to tell her staff when they asked her about testing children with disabilities. In response, she simply told them to use their best judgment as to whether the child should be assessed.

## **SUMMARY**

As in earlier rounds of the NRS Quality Assurance Study, most of the English language assessments observed in spring 2006 met or exceeded the standard of quality used in certifying assessors. The average certification score was 92, and 83 percent of observed assessments exceeded the certification standard of 85 points. The inter-rater reliability of assessment scale scores remained quite high, and the certification score of the Spanish-language assessments observed (65 in 11 sites) was also high, on average (97 points).

Errors due to coaching, inappropriate gestures, administering non-neutral encouragement, straying from the script, and scoring errors were higher in spring 2006 than in spring 2005; only errors for inserting articles such as "a" and "the" decreased. The increase in errors may reflect changes made to the assessment in spring 2005 (particularly

those that simplified gesturing) and perhaps less attention to refresher training that would underscore such changes in administration.

Although quality was high overall, some areas of the assessment remained difficult for staff. Modifications to the assessment easel or additional guidance and training might be helpful in these areas. The most difficult areas were also problematic in Years 1 and 2, including:

- Setting up the testing in a quiet area free of distractions
- Avoiding coaching, particularly in the Simon Says section
- Gesture errors on certain Art Show and Early Math items
- Scoring the counting item (E20) correctly in the Early Math section

Performance in these difficult areas was somewhat poorer than in spring 2005. Scoring errors on the counting item, however, remained essentially unchanged—it was scored incorrectly on roughly one in four assessments. Moreover, while the Letter Naming section yielded low error rates overall, with the exception of script errors in the introduction and transitions in between letter plates, it remains a section with which many programs continue to struggle. They worry that this part of the assessment is not valid, since children may very well know some or all of the alphabet letters, but the structure of this test administration might not enable children to demonstrate what they know. Allowing assessors to point to each letter and say, “What’s this?” or “What’s the name of this one?” could help ensure that children do not inadvertently omit a letter that they lost track of while scanning the plate.

About half of sample programs reported that most children reacted positively to the child assessment. Assessors noted that some children enjoyed the one-on-one time with the assessor and liked showing off what they knew, as well as the opportunity to “play some games.” Children’s behavior was much less of a concern in the spring than it had been in the fall, a finding similar to Years 1 and 2. The major behavioral issue was that children became bored or restless during the PPVT and/or the Letter Naming task and needed a lot of redirection. Nearly half of the programs thought that the NRS assessment is too long.

Most programs (24 out of 35, or 68 percent) administered the child assessment to English Language Learners. Spanish was the most common other language spoken, followed by Chinese and Arabic. Assessors reported that most children could pass the English language screener in the spring, even if they had not passed in the fall. Some children who are classified as having Spanish as a primary language frequently are more fluent in English; similarly, some children lose their Spanish skills over the course of the school year, and perform better on the English version in the spring. Bilingual assessors, as has been the case in previous rounds of site visits, were critical about two overarching components of the NRS assessment. First, they felt that certain Spanish speakers are at a disadvantage due to the wide variety of dialects spoken by children and families who use colloquial words not considered to be correct on the NRS assessment. Second, programs felt that children should not be tested on four additional alphabet letters (CH, LL, Ñ, and RR), which they believe are not in modern usage; in fact, RR is no longer used and has been

---

eliminated from future versions of the assessment. Nearly all programs (33 out of 35 in the sample) assessed children with disabilities, using a wide range of accommodations for the child assessments. In some situations, assessors decided to administer the NRS assessment to a child who could reasonably have been exempted. Most staff members who administered assessments to children with disabilities felt comfortable with the process and with the accommodations made. However, several programs would like the NRS training materials to address in more detail the issues of conducting the NRS assessment to children with special needs and making appropriate accommodations for these testing situations. A number of Head Start programs in each round have expressed interest in these issues since the inception of the NRS.

**This page has been intentionally left blank for double-sided copying.**

## CHAPTER III

### LOCAL APPROACHES TO TRAINING NRS ASSESSORS

---

Consistent administration of the NRS assessment across all Head Start programs is crucial for maintaining the integrity of the assessment results. To promote consistent implementation, programs are given standardized procedures for training local staff on the NRS and for administering the assessment. When the NRS was first implemented in 2003, the Office of Head Start, working with contractors Westat and Xtria, developed a detailed training plan and materials, and devoted considerable resources to ensuring that they were used appropriately.

Due to the need to conduct a large number of assessments across the country, the Office of Head Start required an extensive cadre of certified assessors to implement the NRS. Recognizing the short time frame in which assessors were to be trained, the Office of Head Start used a “training-of-trainers” (ToT) model to achieve this goal. In summer 2003, prior to conducting the first round of assessments, federal officials hosted a series of 13 regional training conferences to train local program staff who had been selected by their programs to become lead NRS trainers. The lead trainers, in turn, trained those staff who would be administering the assessments at the local level. The Office of Head Start and national contractors developed standardized training procedures and materials for the NRS so that local staff in all programs across the country would receive the same training and would become certified to conduct the child assessment using consistent procedures and criteria.

Since the completion of the initial round of training conducted in fall 2003, training for NRS assessors has focused primarily on refresher training sessions at the local level for experienced assessors. These sessions are designed to reacquaint assessors with procedures for administering the assessment and to instruct them on administering new items. In addition, some programs have needed to provide separate trainings and certification for new assessors who were not trained in previous rounds. In summer 2005, prior to the third year of NRS implementation, the Office of Head Start provided three regional ToT conferences

(in Silver Spring, Maryland, Chicago, and San Francisco) for programs that had new lead NRS trainers. For the fall 2005 assessment, the Office of Head Start and its contractors prepared materials for local programs to use in refresher training for their local assessors. Programs then held refresher trainings and a limited number of new assessor trainings as needed locally in preparation for administering the spring 2006 assessment.

In this chapter, we describe the approach to training NRS assessors utilized by programs in spring 2006. Based on individual interviews with Head Start directors and NRS lead trainers, as well as focus group discussions with assessors in the 35 sample programs, we first briefly discuss the experiences of any programs that sent staff to attend one of the regional conferences in summer 2005. We then turn to describing local programs' approaches to conducting refresher and new assessor training for the English- and Spanish-language assessments. Finally, we present the perspectives of local program staff on the training, materials, and support they have received from the Office of Head Start and its contractors.<sup>1,2</sup>

#### **EXPERIENCES WITH THE SUMMER 2005 REGIONAL “TRAINING-OF-TRAINERS” CONFERENCE**

Out of the 35 sample programs, only three programs sent staff to attend one of the events that took place in Silver Spring, Maryland, Chicago, or San Francisco in summer 2005. One program intended to send its only bilingual assessor, who is not a native speaker, to one of the regional conferences. However, when they discovered that the training was conducted entirely in Spanish, she did not feel that she was fluent enough to participate. Another program considered sending an additional person to get trained firsthand in NRS procedures, but none of the locations was convenient to program staff.<sup>3</sup>

---

<sup>1</sup> As part of the spring 2005 round of the Quality Assurance Study, we observed five local refresher trainings. The selected sites included urban and rural programs, programs with only English assessors, those with both English- and Spanish-language assessors, programs that would only be training a few assessors, and programs with large groups of trainees. Across the five sites, we observed 47 experienced assessors participating in a refresher training. (See Paulsell et al. 2006.) We did not, however, observe any local trainings as part of spring 2006 data collection.

<sup>2</sup> In calculating percentages, all programs are weighted equally, regardless of their size or the number of staff interviewed. The common denominator is always 33 programs, regardless of whether a given program had an opinion on a certain topic. (Two programs located in Puerto Rico are excluded in discussions about training for the English-language version of the assessment.) For example, in response to the question “Did the initial training you received at one of the training-of-trainers regional conferences (or one of the “replacement” trainings) adequately prepare you to train local program staff to conduct the NRS assessment?,” many lead trainers did not provide an opinion on this topic because most programs did not attend a regional conference in summer 2005. If several programs did not provide information on a given topic, we present the number of applicable programs (for example, “Six programs decided...”) as opposed to a percentage of programs. In general, specific concerns or activities are counted if at least one respondent at a program mentions them. Disagreement among respondents within programs is noted when it occurs.

<sup>3</sup> Although programs considered sending assessors to this training, the OHS has made clear that it is not appropriate for this purpose; instead it is a Training of Trainers, intended for replacement NRS leads.

---

Three programs sent their lead trainer to a regional conference either to address an internal staffing need or to receive the national NRS training from the federal contractors firsthand. One Head Start director, who attended a summer 2003 ToT event and had served as the lead trainer for the first two years of NRS implementation, transferred responsibility for local training to an experienced bilingual NRS assessor. Since the director could not speak Spanish and felt inadequate to answer local assessors' questions about the Spanish version of the assessment, she decided that the bilingual assessor would be a more effective lead trainer. Another program (a delegate agency) had never sent someone from its own staff to a regional conference; instead its lead trainer had been trained and certified by a lead trainer from a nearby Head Start program. The director decided that it would be helpful to have an internal staff member who could get the groundwork training firsthand. The lead trainer at the third program attended a ToT conference because she took over for the previous education coordinator and NRS lead trainer.

All lead trainers gave positive feedback about the content and materials for the summer 2005 training events. Overall, participants found the written materials and video to be useful, noting that they appreciated learning about the differences between the English and Spanish versions, as well as being able to speak with a national trainer and ask questions in person. One lead trainer felt that she got even more out of the ToT event because she had already conducted NRS assessments, as opposed to attending as someone who had no experience with the NRS whatsoever.

Several sample programs expressed interest in having refresher trainings for lead trainers, or the opportunity to send more than one person to a regional conference. Five programs advocated refresher trainings for lead trainers. As one lead trainer explained, the initial ToT event in summer 2003 was overwhelming, despite the fact that someone else from her program attended with her. National trainers distributed a lot of information, and she worried that no one has given her feedback on how she conducts local trainings or how she administers the assessments since becoming certified. She felt that instituting a refresher training for lead trainers would help support NRS lead trainers and help ensure that "bad habits" would not get perpetuated among individual programs. In addition, two programs would like the Office of Head Start to sponsor cluster trainings (local trainings), perhaps through regional offices. As suggested by staff at these programs, local programs would be able to share strategies and discuss issues that arise during the assessments, and lead trainers could ask regional TA specialists questions to make sure they are following the NRS protocols appropriately.

A handful of other programs would like future ToT events to be held in more locations so that staff can avoid traveling a great distance, and would like to receive funding to be able to send more than one person to a regional conference. This approach would enable more people who implement the NRS to receive information firsthand and might help mitigate the effects of unexpected turnover (that is, someone who has already been to a ToT could quickly step into the lead trainer's position).

## **APPROACHES TO ASSESSOR TRAINING IN SPRING 2006**

In preparation for the spring 2006 round of assessments, the Office of Head Start again relied on programs' lead trainers to train local staff. It provided each program with copies of the spring 2006 NRS assessment battery in an easel format, score sheets, guidelines with an agenda for conducting refresher and new assessor training, a trainer's script, several role-play scenarios, new training videos in English and Spanish, and the Spring 2006 Assessor's Guide. Materials were also subsequently posted on the Office of Head Start's website.

The refresher training guidelines for the spring 2006 assessments instructed lead NRS trainers to incorporate four core elements into the local training: (1) show the video, (2) lead a discussion about the video, focusing on techniques used to address behavioral problems and on changes to the assessment, (3) distribute the Assessor's Guide and review it during training, and (4) conduct two rounds of unscripted role-plays, followed by feedback to the group. Experienced assessors were not recertified as part of refresher training. The suggested time frame for completing these activities was approximately 4.5 hours. Following refresher training on the English version, bilingual assessors were expected to watch the Spanish segment of the video, review an updated easel and role-play the assessment. This training for bilingual assessors was recommended to take an additional hour. As for English assessors, experienced bilingual assessors were not recertified as part of refresher training.

The new training video for the spring 2006 assessments described changes made to the assessment since fall 2005, guidelines for completing the answer sheet, and procedures for administering the assessment. The video also included more examples than in the fall 2005 video (but was comparable to the spring 2005 version) of how to manage children's difficult behaviors during the assessment. As mentioned above, the video included a parallel segment on the Spanish-language assessment.

### **Approaches to Refresher Training (English)**

The majority of sample programs used a lead trainer who had attended one of the past summer regional training conferences hosted by the Office of Head Start to conduct refresher trainings on the English-language assessment. In contrast, at one program, the director, who never attended a regional training, has led all refresher trainings for local assessors; a ToT-certified lead trainer from a nearby Head Start program is in charge of certifying any new assessors for this site. At another program, the director attended a summer 2003 training and continues to conduct all trainings and certifications of new assessors, but relies on an experienced NRS assessor to be in charge of refresher trainings. With few exceptions, lead trainers who attended one of the regional ToT conferences during the first three years of NRS implementation indicated that the training adequately prepared them to train their own staff and answer questions that emerged regarding NRS implementation.

As in previous rounds of the Quality Assurance Study, there was considerable variation in how programs approached refresher training on the English-language version of the assessment; however, the reported rate of compliance with the training protocol was higher overall in spring 2006 than in spring 2005. With few exceptions, most programs provided a



refresher training session on the English-language assessment. More than one quarter of sample programs followed the training guidelines provided by the Office of Head Start—showing and discussing the video, reviewing the Assessor’s Guide, and role-playing the assessment (see Box next page), compared to just 14 percent of sample programs in spring 2005. Two of these programs conducted role-play exercises with non-NRS eligible children enrolled in Head Start. Approximately one-quarter of programs included three of the four elements—showing the video and discussing the changes, along with either role-playing the assessment or reviewing the Assessor’s Guide as a group. Twenty-seven percent included two elements; most of them showed the video and did one other activity, usually a discussion of changes to the easel. Three programs opted for self-study to prepare for the NRS assessments (one assessment team met without the lead trainer). Two other programs did not offer a refresher training, and a third program only delivered a refresher training to half of its experienced assessors. Unlike spring 2005, none of the sample programs held a group training that incorporated only one training element.

**Approaches to the Spring 2006 Refresher Training  
for the English-language Assessment**

	Percentage of Programs
<b>Included all required elements</b>	
Video, discussion, guide, role-play	27
<b>Included 3 of 4 required elements</b>	
Video, discussion, role-play	12
Video, discussion, guide	12
<b>Included 2 of 4 required elements</b>	
Video, discussion	24
Video, role-play	6
Discussion, role-play	3
Self-study	9
No refresher training	9 <sup>a</sup>

N = 33 Head Start programs (two programs from Puerto Rico are only included in discussion of Spanish trainings)

<sup>a</sup>One program delivered a refresher training to only about half of those assessors who conducted NRS assessments prior to fall 2005; experienced assessors only attend one refresher training each school year, either in the fall or spring.

In spring 2006, refresher training sessions lasted for approximately 2 hours and 15 minutes on average, ranging from 45 minutes to 4 hours.<sup>4</sup> This represents a slight decline in the average length of training for the spring NRS assessment rounds, which in the past have lasted, on average, 2 hours and 45 minutes (both spring 2004 and spring 2005).

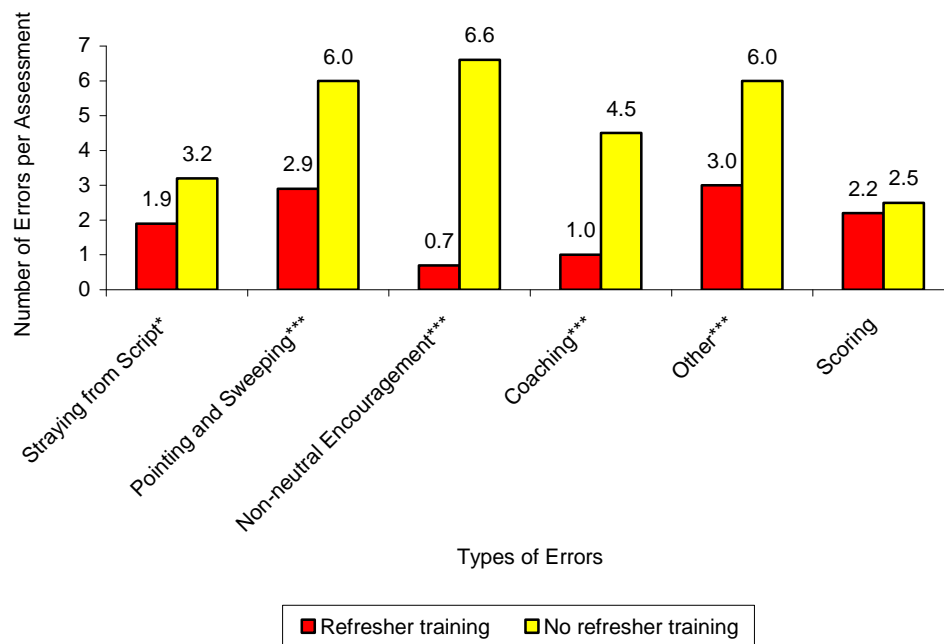
Several factors may explain some programs’ lack of compliance with the training protocols. Several lead trainers reported having to omit certain training components due to the late arrival of materials, some of which came only weeks or days before the program year ended. For example, three programs could not conduct role-playing exercises because the easels and scripts had not yet arrived; others had to compress the training schedule or use self-study as there was little time to convene multiple assessors before assessments needed to

<sup>4</sup> These figures do not include self-study sessions or one program that opted to provide one-on-one training sessions between assessors and the lead trainer that lasted 30 minutes each.

begin. Moreover, several lead trainers reported that they skipped certain elements because the assessment team consisted of experienced NRS assessors who did not think that it was necessary to sit through the suggested training model as presented by the Office of Head Start. In fact, staff in some of the programs suggested shortening or eliminating refresher training, suggested shortening the video (as discussed later in this chapter, staff in many programs found the video redundant). Concerns about staff burden may have led some programs to shorten training. As discussed in more detail in Chapter VI, the amount of staff time devoted to the NRS—including time spent in training—is a significant concern in more than one-third of the sample programs. Another factor may be the late arrival of training materials. Some programs opted to cut training short and begin assessing children as soon as possible. Another issue that may affect training plans is that there is no requirement that experienced assessors be recertified prior to each round of data collection. Programs might be more likely to conduct refresher trainings if recertification were required.

Whether programs decided to offer refresher trainings was related to the likelihood that assessors would commit administration errors. Of the 32 programs that administered the English version of the NRS assessment, two programs did not offer any refresher training for experienced assessors. Observations revealed that assessors at these two programs were much more likely to make administration errors, as Figure III.1 shows.

**Figure III.1. Observed Errors for Programs That Offered and Did Not Offer a Refresher Training (English)**



Note: "Other" errors include omitting gestures, along with inserting articles (for example, "a" and "the") and mispronouncing words during the PPVT.

N = 32 programs.

\*Statistically significant at  $p < .10$ .

\*\*\*Statistically significant at  $p < .01$ .

About twice as many gesturing errors and “other” errors (for example, omitting required gestures and mispronouncing vocabulary words) per assessment emerged among observations that occurred at programs without refresher trainings. Differences in providing non-neutral encouragement and coaching errors were even more striking. In contrast, assessors observed by site visitors were not any more likely to score items incorrectly if they did not receive NRS refresher training.

Moreover, assessors who were observed during the site visits achieved significantly lower certification scores on the English assessment (77.1 points) as compared with those from programs that offered refresher trainings (93.2 points). (Differences in mean certification scores were statistically significant at  $p < .01$ .) These findings suggest that programs that chose to forego refresher trainings may be compromising the fidelity of the administration procedures for the NRS, a trend that would likely be reversed if refresher trainings were offered consistently.<sup>5</sup>

#### Approaches to the Spring 2006 Refresher Training for the Spanish-language Assessment

	Percentage of Programs
<b>Included all required elements</b>	
Video, discussion, guide, role-play	13
<b>Included 3 of 4 required elements</b>	
Video, discussion, role-play	17
Video, discussion, guide	13
<b>Included 2 of 4 required elements</b>	
Video, role-play	6
Video, discussion	4
Discussion, role-play	4
Self-study	42
No refresher training	8 <sup>a</sup>
N = 24 Head Start programs	

<sup>a</sup>One program delivered a refresher training to only about half of those assessors who conducted NRS assessments prior to fall 2005; experienced assessors only attend one refresher training each school year, either in the fall or spring.

#### Approaches to Refresher Training (Spanish)

Nearly all of the 24 programs with bilingual assessors provided some form of refresher training for them. However, like the training for the English assessment, the degree to which programs adhered to the suggested training format varied substantially. As in previous rounds of the Quality Assurance Study, programs reported that sometimes it was difficult to identify a certified NRS trainer who could deliver training for the Spanish assessment. Six programs could not rely on the designated lead trainer since that individual was never certified as a bilingual NRS trainer. As a solution, these programs usually made

<sup>5</sup> There was a third program that required its veteran assessors to attend a refresher training either in the fall or spring; about half of the assessment team attends one or the other. We excluded this program from Figure III.1 in the main part of the text because it was impossible to know if the assessors we observed attended the refresher training in fall 2005 or spring 2006. However, when we computed error rates and certification scores, we found similar patterns with comparable levels of statistical significance. Once again, scoring errors were not statistically significant.

the Spanish video and other materials available to bilingual assessors to review on their own time, after they participated in the training for the English version.

More than half of the sample programs (54 percent) conducted a formal group training on the assessment (separate from the refresher training on the English assessment) that included various combinations of the core elements—watching the video, discussing the assessment, reviewing the Assessor’s Guide, and role-playing (see Box next page). Three programs completed all four training elements, as instructed by the Office of Head Start. Self-study emerged as another popular method for preparing to conduct the Spanish-language version of the assessment (42 percent), much more so than for the English one. Often, this consisted of distributing the Spanish video to bilingual assessors to view and encouraging them to ask questions as needed; however, not all lead trainers could verify that assessors had watched the video. One lead trainer distributed a memorandum with instructions for self-study in both languages that included watching the video, reading the easel several times out loud, and reviewing the Assessor’s Guide. Two lead trainers who served as the only bilingual assessors for their programs watched the video and read through the materials. Only one program did not provide refresher training for bilingual assessors who were initially certified at a nearby Head Start program. Two of these four assessors attended their own program’s refresher training for the English assessment; none of them went to the nearby program where they initially were trained in Spanish. Moreover, as was the case with the training on the English-language version, one program reported that it does not require experienced NRS assessors to attend refresher training in both fall and spring. Across programs, trainings on the Spanish-language version lasted, on average, slightly longer than the English one—about 2 hours and 30 minutes.

Reasons staff reported for not following the training format outlined by the Office of Head Start were similar to those mentioned regarding the English-language assessment training. For example, two programs did not receive the Assessor’s Guide in Spanish in time, and another program did not receive the child role-play scripts or easels in time (the lead trainer attempted to download the scripts from the Office of Head Start’s website, but they had not yet been posted). Two other programs decided to forego the practice sessions since they have conducted NRS assessments since 2003 and did not think that role-play was necessary. One lead trainer prepared self-study memos for each assessor because none of the training materials arrived in time to organize and reschedule a group training. Bilingual assessors from three programs had been certified elsewhere and opted for self-study for logistical reasons. In the course of reviewing interview and focus group notes, we discovered four sample programs that use bilingual assessors who never received an initial training and certification in Spanish. At a fifth program, the lead trainer conducts all Spanish assessments and ‘trained and certified’ herself on the Spanish-language version. An additional program has a few enrolled Spanish-speaking children but has never conducted NRS assessments in Spanish.

### **Approaches to Training New Assessors**

Training guidelines instruct lead NRS trainers to provide new assessors with the full-day fall 2005 training, certify new assessors on the fall assessment, and then provide them

---

with spring 2006 refresher training along with the experienced assessors. This method is intended to ensure that assessors who are new in spring 2006 receive the same training that experienced assessors received.

Six of the sample programs (18 percent) reported training and certifying staff who would be administering NRS assessments for the first time in spring 2006; an additional seventh program trained and certified two experienced assessors on the English-language version to become bilingual assessors. The number of new staff trained at these programs ranged from one to seven. All these programs reported certifying the new assessors based on an observation of a practice assessment, and half of them reported using a child for certification (either a Head Start child who was not eligible for the NRS, or a non-Head Start child going into kindergarten). As was the case in spring 2005, none of the programs, however, reported conducting a full one-day training using the fall materials as instructed in the training guidelines. Instead, most reported conducting a two- to three-hour training in addition to the refresher training. One program delivered a half-day training for three new assessors, and then required that they attend the refresher training as well; another program provided a four-hour one-on-one training with the new assessor. The initial training for a bilingual assessor at a third program seemed to consist solely of attending the refresher training with all other assessors and then getting certified using a child.

In rare cases, lead trainers seemed to neglect certifying new assessors in spring 2006. Experienced NRS assessors from one program who were getting trained for the first time on the Spanish-language version reported that they were never certified in Spanish. Instead, they watched the video, discussed the materials with the lead trainer, and role-played without a certification form being completed. Similarly, according to one assessor from another program, three assessors in that program did not see either language video and were never certified; training entailed reviewing the easel and role-play activities. As discussed previously, concerns about staff burden and the limited time available to complete assessments by the end of the program year may explain why some programs did not follow the training protocols recommended by the Office of Head Start.

### **LOCAL PROGRAMS' RESPONSES TO TRAINING RESOURCES AND MATERIALS**

As in past rounds of the Quality Assurance Study, local Head Start staff impressions of the NRS training materials and information they received were mixed. This section presents the views of local program staff on the usefulness of the NRS broadcasts and webcasts,<sup>6</sup> the spring 2006 training video, and the Assessor's Guide provided by the Office of Head Start and its contractors for supporting NRS implementation. We also describe difficulties associated with the delivery of NRS training and assessment materials, as well as programs' use of the technical assistance help line.

---

<sup>6</sup> During staff interviews, site visitors asked directors and lead trainers specifically about the March 2006 webcast. During numerous conversations, however, staff members also commented on the satellite broadcasts that took place from 2003–2005. Consequently, some of the information presented in this section includes information on experiences and perceptions of the earlier satellite broadcasts, as well as the webcast that was launched in March 2006.

### Satellite Broadcasts and Webcasts

The Office of Head Start hosted satellite broadcasts on August 4 and October 27, 2004, and February 17, 2005 to provide updates on the NRS. These covered information on upcoming rounds of training and child assessments, new CBRS features, and the availability of technical assistance. Researchers also reported on NRS outcomes and on findings from the Quality Assurance Study. Satellite broadcasts allowed the Office of Head Start the opportunity to reach all Head Start programs and, because the programs could call in or e-mail their questions, to clarify any confusion program staff might still have about NRS implementation. Transcripts were also made available after the broadcasts took place.

In spring 2006, the Office of Head Start posted a prerecorded webcast video of approximately 22 minutes in length that was available to local programs to view at any time. A transcript of the video was also available to programs. Topics covered were similar to past broadcasts and included the results of the fall 2005 NRS child assessments, the spring 2006 refresher training procedures, the spring 2005 results of the Quality Assurance Study, and modifications and updates to the assessment battery. In lieu of a live question-and-answer session, federal officials periodically posted inquiries and responses that had been submitted to the Office of Head Start online over the course of the program year. This decision came about partly in response to feedback from programs in previous study site visits, which indicated that the broadcasts were too lengthy and the fielded live questions from local staff who called in were sometimes repetitive and not very useful.

At least one staff member from nearly all sample programs reported watching at least one of the broadcasts or the webcast. Only two programs had never watched any broadcasts or the webcast or downloaded any materials for review, and in a third program the lead trainer had never watched or accessed any training materials from the website. Eight programs (23 percent) ensured that one or more individuals watch each of the sessions sponsored by the Office of Head Start, and about 40 percent either viewed the March 2006 webcast or downloaded and read the transcript. Staff from a handful of programs explained that they had not yet viewed the March 2006 webcast because they were too busy; one program was preparing for its PRISM review. Viewership varied considerably across programs. Fifty-seven percent of Head Start directors reported watching one or more of the webcasts either alone or with other program staff, although in some cases directors delegated this responsibility to the NRS lead trainer.<sup>7</sup> As in spring 2005, lead trainers were the only staff members to view the March 2006 webcast in about one third of sample programs.

Among those programs in which staff watched at least some of the broadcasts or webcasts, feedback about their usefulness was mixed. Twelve programs reported that the webcasts were useful because they provided information about the assessment and training schedules, assessment outcomes, the difference between the NRS child assessment and local assessments, or some ideas about how to use the NRS growth reports for program improvement. Three of these programs appreciated the question-and-answer portion of the

---

<sup>7</sup> This percentage includes directors who only had watched one broadcast or webcast since the NRS began (n = 4).

broadcasts. One lead trainer found the broadcasts and webcasts to be invaluable for a rural program like hers as a means to stay connected to the wider Head Start community. Nevertheless, an equal number of programs noted that they were too long, too scripted, or repetitive. A program director did acknowledge, however, that the format of the latest webcast was preferable because programs could now download and review it at a time that was convenient for their staff. Five programs expressed mixed opinions. Two of these programs noted that the earlier satellite broadcasts had been helpful, but they had not provided any new or interesting information since Year 1. Local staff from the remaining three programs reported that while they found certain sections useful (for example, the mailing schedule or information about the overall direction of the NRS) the webcasts were also too long and overly scripted.

### **Delivery of Training Materials**

Nearly half (49 percent) noted that materials arrived too late to carry out recommended training, and six programs only received part of the expected materials. For example, two programs did not receive easels in time for the training and three did not receive the Assessor's Guides in time for training. Due to the compressed time frame in which assessments needed to be completed, some programs had to make contingency plans. In certain cases, lead trainers omitted a component of the local training (for example, if no easels arrived then they could not do role-play exercises). Because there was not enough time to reschedule a group training, one lead trainer sent a memorandum to each assessor with instructions on how to prepare for the spring assessments, while another lead trainer scheduled individual 30-minute sessions with each assessor to review the video and changes to the easel. One program received materials so late that assessors could not prepare at all; the first time they saw the new easel was when they began the first spring assessment. Another program only received easels in time to conduct assessments as a result of their participation in the QA study (in which a single easel was sent to the program by MPR staff and was used to complete assessments).

Staff from many programs emphasized how difficult it was to plan for group refresher trainings—sometimes in multiple centers spread out over several miles—in the midst of competing end-of-the-year activities and deadlines. Programs frequently must schedule NRS trainings weeks or months in advance, planning that can be disrupted by delayed shipment of materials. As discussed in Chapter VI, more than one-third (37 percent) of the sample programs either complained of receiving materials late or specifically requested that materials be distributed sooner. The earliest programs received their materials was mid-April, which was only weeks before the end of the year for several programs in the representative sample. A few programs suggested that as a back-up plan, the Office of Head Start should ensure that *updated* NRS training materials be made available on its website as a last resort.

### **Spring 2006 Training Video**

As in past rounds of training, the video was the centerpiece of the spring 2006 refresher training. It demonstrated correct administration of the assessment, introduced any new

procedures (such as allowing for more unscripted interaction with the child), and demonstrated helpful techniques for managing children's behavior during the assessment.

Lead trainers showed the spring 2006 video during refresher training in 32 of the 35 sample programs. Local program staff generally had positive impressions of the video and found it helpful for training (see Box). In fact, staff from 26 programs, including 19 focus groups and 9 lead trainers, remarked that the video is one of the best components of the local NRS training and should continue to be shown. Assessors appreciated that more recent versions of the video have given more guidance on offering supplemental neutral encouragement that still falls within NRS protocols, along with suggestions on how to handle children who exhibit behavioral challenges.<sup>8</sup> Some staff also underscored the usefulness of the video's summary section and review of core components of the NRS easel.

#### **Impressions of the Spring 2006 Training Video**

- **Positive Impressions**

- Shows how to appropriately administer supplemental neutral encouragement
- Shows more types of behavior challenges than some previous videos and offers some strategies for managing them
- Shows how to complete answer sheet
- Includes a summary section of key points; highlights common administration errors
- Effective learning tool for visual learners

- **Negative Impressions**

- Too long
- Too repetitive
- Scenarios not typical of many Head Start children
- Does not address children with disabilities

At the same time, some programs did not think that the video was particularly helpful and often found that role-playing was a more useful training activity. Some staff said it was too repetitive. In particular, programs felt that the videos have shown, more or less, the same thing over time. They thought that experienced assessors did not need to see the same material each assessment round. Instead, the video should just focus on changes made to the instrument and on key points rather than demonstrating entire assessments—at least for experienced NRS assessors. While many staff appreciated the examples of strategies for assessing children whose behavior is challenging to manage, they still felt that the video could do even more to include more difficult scenarios (for example, children who are “fidgety” and do not immediately return to the task at hand when redirected by assessor).<sup>9</sup> Moreover, a handful of programs would like the video to address directly the topic of assessing children with disabilities (see Chapter II for a discussion of programs' experiences administering the NRS assessment to children with disabilities).

<sup>8</sup> However, there was at least one program in which the lead trainer continued to follow instructions from Year 1, when NRS assessors were told to adhere strictly to the easel script. Although based on a small sample, this incident suggests that a small percentage of programs may be reluctant to use additional neutral encouragement at their discretion, even though this is presented as acceptable in more recent training videos.

<sup>9</sup> The Assessor's Guide includes a three-page section entitled “Tips for Dealing with Behavioral Issues during the Assessment.”



---

The Spanish video gave bilingual assessors an opportunity to watch a complete administration of the assessment in Spanish. All but two programs that conducted the Spanish-language version of the assessment either showed the video or, in a few cases, made it available to bilingual assessors to view on their own. Staff members in one program mentioned that watching the entire video in both languages was redundant. In addition, one assessment team noted that the vocabulary and dialect used in the video differs from the way they speak in Puerto Rico, but they did not imply that they changed the language in their scripts.

### **Assessor's Guide**

The Assessor's Guide is a tool that outlines general NRS procedures, instructions for completing the score sheet, detailed discussions about each of the five sections, tips for addressing behavior challenges, and some information on assessing children with disabilities. The Office of Head Start envisioned the guide as one resource to which staff can refer if they have questions or need to verify something (assessors can also speak to lead trainers, and programs can contact the HSNRS Assessment Helpline). Like previous assessment rounds, the spring 2006 training guidelines stressed the importance of ensuring that all assessors have a guide; they also encouraged trainers to review the guide during training. Almost 70 percent of programs reported that they received an Assessor's Guide, although a handful of programs did not receive them in time for their local training. Three programs did not receive an Assessor's Guide in their NRS packets in time for the local training, and instead only received a summary sheet of changes. The lead trainer from one of these programs attempted to download a copy from the Office of Head Start's website, but the spring 2006 version was not yet available.<sup>10</sup> Seven programs did not provide any information about the guide during site visits.

Of those programs that reviewed an Assessor's Guide, most distributed copies to each assessor and reviewed the guide during training; one lead trainer kept the only copy in her office for assessors to borrow as needed. This represents an overall increase in the number of sample programs that distributed the guides to assessors and reviewed them as a group. In previous rounds of site visits, we found that many assessors did not have a copy of the Assessor's Guide or did not refer to it when they had questions about assessment procedures. The Assessor's Guide is a key reference tool and may help to reduce some of the scoring and administration errors noted in Chapter II. Several programs noted that the Assessor's Guide is a useful tool they can refer to if they have questions; one focus group said that the "what if" scenarios have been helpful. Staff from 12 programs said that the guide is a key training component that local programs should continue to receive.

---

<sup>10</sup> An English version of the guide has been made available on the website for each assessment round since fall 2005.

## Technical Assistance

Aside from the information provided in the training packets, local staff can contact one of two telephone help lines with questions about NRS implementation—one for questions about the child assessment, and one for questions about the computer-based reporting system (CBRS). Programs are informed about these telephone numbers and corresponding e-mail addresses in the trainer’s binder, in the Assessor’s Guide, in the CBRS user manual, and on the Office of Head Start’s website.

As in previous rounds, only a small number of programs reported calling the help line with questions about administering the assessment (see Chapter IV for programs seeking support on the CBRS).<sup>11</sup> Three programs asked whether a child who enrolled after the fall 2005 assessment round should be assessed in the spring; one program asked if a child who was bound for kindergarten but older than five years old should be assessed. Feedback from staff on these inquiries was positive. However, one program that called the help line about a Simon Says item did not find the technical assistance to be helpful.

Among programs that did not seek technical assistance, staff nevertheless seemed to be familiar with the toll-free telephone numbers and e-mail addresses for the Office of Head Start and its contractors. Some staff mentioned other sources of support, including the state Head Start association, a delegate’s grantee agency, and the lead trainer from a nearby program that certified the assessment team. One program director has referred parents to the Office of Head Start’s website if they have questions, which has been a useful resource. No staff members reported that they would contact their regional Head Start T/TA specialist for guidance on NRS implementation.

## SUMMARY

To ensure consistent administration of the NRS assessment, Head Start programs need to follow standardized procedures when training local staff. Since the NRS was first implemented, the Office of Head Start has provided local programs with standardized training materials and guidelines to conduct training.

As expected, in summer 2005, few programs sent a representative to one of the “training-of-trainers” events, which are designed for training replacement NRS leads. Several programs, however, would like to see refresher trainings for NRS lead trainers, and to be able to send more than one staff member to a regional conference to build local capacity for NRS implementation. Having training available closer to their programs would be beneficial.

In spring 2006, all but three of the sample programs provided refresher training. However, only about one quarter adhered to the training guidelines. Other programs did not include all of the required training elements—viewing and discussing the training video, distributing the Assessor’s Guide, and role-playing the assessment—and thus most training

---

<sup>11</sup> Ten programs that called to ask about when NRS materials would arrive are not counted in this section.

---

sessions were shorter than the recommended four and a half hours. About half of the programs that needed to provide refresher training for Spanish assessors reported offering a formal group session; bilingual assessors at most other programs relied on self-study to prepare for the spring 2006 assessments. Possible reasons why most programs did not adhere to the training protocol include late arrival of training and assessment materials, limited time to complete assessments by the end of the program year, and staff members' perceptions that experienced NRS assessors did not need refresher training.

Although noteworthy, the lack of consistent training practices did not result in a complete lack of reliability in assessment administration. However, to continue to ensure consistency of administration across programs, it may be beneficial to alter training protocols. As our analyses demonstrate, those programs without refresher training demonstrated more administration errors.

Eighteen percent of the sample programs trained new assessors in spring 2006. Although all of these programs provided separate training for new assessors and most programs certified them during a practice assessment, none of the programs followed the required training protocol for new assessors. Instead of providing a full-day training, most programs conducted the training in two to three hours. In rare cases, a small number of new assessors did not get formally certified on the NRS after attending a training session.

As in past rounds of site visits, the responses of local Head Start staff to the NRS materials and information provided by the Office of Head Start were mixed. At least one staff member at most programs watched one of the NRS broadcasts or webcasts, but programs' views of their usefulness were uneven. Problems continued with timely delivery of training and assessment materials, with nearly half of programs reporting a delay in receipt, and few programs took advantage of technical assistance offered through the help line or other means. While staff from two-thirds of programs found the video to be the best or one of the best components of local NRS training, other staff found the video too long and repetitive and preferred the role-playing exercises or the Assessor's Guide to help prepare for the spring assessment. Lack of thorough refresher training in many programs, coupled with a tendency not to seek technical assistance, may have contributed to some of the errors in administration and scoring described in Chapter II.

**This page has been intentionally left blank for double-sided copying.**

## CHAPTER IV

### LOCAL APPROACHES TO IMPLEMENTING THE NRS

---

Several dimensions of NRS implementation were left to the discretion of the local programs, resulting in different local approaches to critical components of the system. Some of the approaches adopted by local programs could have implications for the administration of the NRS. For example, different standards for assessor recruitment, coupled with differences in timing and location of assessments, could influence the quality of the completed assessments. A better understanding of these relationships might be useful to the Administration for Children and Families for developing further guidance to local programs on how to administer the assessment.

In this chapter, we describe the approaches to NRS implementation in the 35 programs we visited in spring 2006. Specifically, we describe the programs' approaches to coordinating the assessments, assigning staff to conduct the assessments, and communicating with parents and Policy Councils about the NRS. We also discuss the NRS-related costs reported by local programs, along with their experiences using the computer-based reporting system (CBRS).

#### COORDINATING THE CHILD ASSESSMENTS

Head Start directors and managers needed to decide how to coordinate NRS activities to support smooth implementation and timely completion of the assessments. This section focuses on programs' approaches to four coordination issues: (1) who would have overall responsibility for NRS implementation, (2) where assessments would be conducted, (3) when and how to schedule assessments, and (4) how programs would track progress in completing assessments.

## Staffing for NRS Training, Coordination, and Oversight

As with the samples of programs visited for the Year 2 Quality Assurance Study, program directors tended to delegate overall coordination and supervision of the assessment activities to the lead NRS trainer. Of the 35 programs we visited, only 5 program directors had lead responsibility for the coordination of NRS activities. In at least four programs, the program director played some role in coordinating and overseeing NRS activities, but other staff within these programs maintained primary responsibility for the NRS.

Lead NRS trainers had responsibility for conducting refresher and new assessor training, certifying new assessors to administer the NRS assessment, and monitoring and tracking progress in completing assessments. In more than one-third of the sample programs, lead trainers also acted as lead NRS data managers and, as such, were responsible for overseeing data entry into the CBRS. In addition, nearly all lead trainers had primary responsibility for ensuring that NRS score sheets were submitted by the deadline set by the Office of Head Start.

Most lead trainers were either education coordinators or program specialists (see Box). In contrast to spring 2005, no programs hired an outside consultant to manage NRS activities, relying instead on program staff to coordinate these activities.

Only one of the sample programs reported that it had experienced turnover in the lead NRS trainer position from spring 2005 to spring 2006. In this case, turnover occurred because the staff member appointed as lead trainer left the program. The lead NRS trainer position was vacant when we visited this program, and the program director planned to lead refresher training for the spring 2006 round of NRS assessments.

Programs' Lead Trainers	
	Percentage of Programs
Program manager <sup>a</sup>	49
Education coordinator <sup>b</sup>	34
Program director	14
Teaching staff	3
Vacant	3
N = 35 Head Start programs	
<sup>a</sup> “Program manager” includes such positions as family support coordinator, disability specialist, and program coordinator.	
<sup>b</sup> “Education coordinator” includes education coordinators, education managers, education directors, and education specialists.	

## Locations Where Assessments Were Conducted

All programs conducted the spring 2006 assessments in locations within the Head Start centers. Most lead trainers reported that assessments were conducted in spaces that could be closed off to traffic and noise, such as empty offices and classrooms, cafeterias, auditoriums, multipurpose rooms, parent involvement and resource rooms, kitchens, and bathrooms. Staff in one program reported conducting assessments in a quiet area of the Head Start classroom while other children were present.

Some programs reported difficulty identifying appropriate, quiet spaces for the assessments. For example, an assessor in one program reported conducting assessments in a

bathroom with a loud fan and heavy traffic from children often entering and exiting. Fewer than one-third of the sample programs reported conducting some or all assessments in more open spaces, such as hallways, stairwells, foyers, or even outdoors. In these programs, lead trainers indicated that they tried to conduct assessments in these locations at times when the area was usually empty and quiet. Nevertheless, conducting assessments in these locations may account for site visitors' reports that observed assessments were not always conducted in a quiet area (see Chapter II). Specifically, among English-language assessments, site visitors reported distractions including general noise (12 percent), other adults (6 percent), other children (3 percent), and other sorts of distractions like cell phones ringing (3 percent). Similarly, among Spanish-language assessments, site visitors reported distractions, including general noise (7 percent), other children (8 percent), and other distractions (1 percent). As with previous rounds of site visits, three of the sample programs in spring 2006 reported occasionally conducting assessments outside.

### Scheduling the Assessments

All but one sample program reported pulling children out of their classrooms to conduct the spring 2006 assessments, and programs used four different approaches for scheduling them (see Box). Most programs (51 percent) left the scheduling of assessments to the discretion of the assessors or the staff at each center. Second most often, the lead NRS trainer, program director, or lead NRS data manager scheduled the assessments and distributed the schedules to the assessors. In a few programs, assessors were instructed to complete all assessments

within a specific window of time, usually during a two to three-week period. Finally, in one program, scheduling varied according to the characteristics of the centers—assessments were scheduled at the centers serving the fewest children first, followed by larger centers.

Two-thirds of the sample programs reported assessing most or all children in the morning, usually between breakfast and lunch, because they tended to be more alert and energetic at that time. Staff in one of these programs reported occasionally assessing children after naptime if a child could not be assessed in the morning. The remainder of the programs reported conducting assessments at any time throughout the day.

Slightly fewer than two-thirds of the sample programs reported assessing at least one child in both English and Spanish in spring 2006. Of those that described how they scheduled assessments for children who needed to be assessed in both languages, 33 percent reported administering each version of the assessment consecutively with each child—first Spanish, then English. In contrast, the remaining two-thirds reported that staff routinely scheduled the Spanish and English versions on different days. Staff in most of these

<b>Programs' Approaches to Scheduling NRS Assessments</b>	
	Percentage of Programs
Assessor or center staff discretion	51
Scheduled by lead trainer, director, or data manager	29
Scheduled within a set window of time	20
Based on characteristics of centers	3
N = 35 Head Start programs	

programs said they administered the English assessment with each child at least one day after the child completed the Spanish version; staff thought it would be too difficult for children to be assessed twice in one day. Programs also sometimes assigned the English assessments to a separate assessor because they had limited numbers of bilingual staff members who needed to focus their attention on completing Spanish language assessments.

Several programs reported not administering the Spanish version of the assessment in the spring to children from language-minority households who were perceived by staff as being primarily conversant and proficient in English, even though the Office of Head Start has instructed all programs to administer the Spanish and English versions with all children whose home language is Spanish. Staff in these programs indicated that these children often were shy about responding in Spanish during the Spanish-language assessment, or were confused about being questioned in Spanish or asked to respond in Spanish because they were not asked to respond in Spanish in the Head Start classroom. Even Head Start programs that followed the recommended protocol for assessing Spanish-speaking children felt strongly that children who were proficient and conversant primarily in English should not be assessed in Spanish. Accordingly, fewer programs assessed children in both Spanish and English in the spring than in the fall. Finally, in one program, staff reported scheduling the English version of the assessment before the Spanish version, even though the Office of Head Start has instructed all programs to administer the Spanish version first with children whose home language is Spanish. In this program, staff said that they tried to conduct the Spanish version before the English version for most children. However, because of the short time frame they had for completing all of the assessments, staff reported that they could not always wait for the Spanish-certified assessors to schedule and complete the Spanish version before they administered the English version.

### **Tracking Progress in Completing the Assessments**

Programs used two approaches to track progress in completing assessments. Nearly all programs (97 percent) required assessors to report completed assessments to the lead NRS trainer or data manager. At some programs, assessors were asked to report progress daily or weekly. At others, assessors sent completed score sheets to the lead trainer or data manager once the group of children they were assessing had completed all assessments. Of programs that used this approach, most distributed classroom rosters to assessors, which were used to track progress. Programs reported using CBRS-generated rosters, rather than either internally-designed rosters or checklists to support their tracking efforts.

The second approach for tracking progress—used by only one program—was to have assessors or center staffs enter completion status directly into the CBRS. The lead NRS trainer then reviewed the CBRS periodically to track the number of assessments that had been completed program-wide.

Two programs did not describe a specific system used to track progress in completing the assessments. However, the lead NRS trainer at one of these programs reported informally checking in with assessors as the deadline for submission to the Office of Head Start approached.



## APPROACHES TO ASSIGNING STAFF TO ADMINISTER THE ASSESSMENTS

Programs took one of three main approaches to assigning staff to conduct the NRS assessments: (1) assigning teachers and assistant teachers only, (2) assigning non-teaching staff only, and (3) using a combination of teachers and other program staff, such as program managers and area specialists. Similar to previous samples of programs visited for this study, fewer than a third of programs used teaching staff exclusively to administer the assessments (29 percent in spring 2006, compared to 33 percent in fall 2004 and 43 percent in spring 2004) (see Box). This pattern differs from the approach adopted in the samples of programs visited in spring 2005, which relied less heavily on only teaching staff to complete assessments (11 percent). In the current round, nearly half of the programs used only non-teaching staff for the NRS assessments, while almost one quarter of sample programs used a combination of teaching and non-teaching staff to administer the assessments. For both English and Spanish-language assessors, the error data indicate that there was no significance in the likelihood that teachers versus non-teachers would commit a coaching error. This finding is particularly interesting given initial concerns of programs that teachers might be more likely, either consciously or unconsciously, to bias the results of assessment administration. In any case, programs are now moving more toward having non-teaching staff to conduct the NRS child cognitive assessments.

Of the programs that used teaching staff to conduct the assessments, most reported that teachers assessed some or all of the NRS eligible children who were enrolled in that teacher's classroom. However, at two of these programs, teaching staff also assessed children enrolled in other classrooms in order to ensure that all assessments were completed in the allotted time frame.

Two-thirds of the programs in our sample reported conducting the Spanish-language assessment in spring 2006. Programs used similar types of staff to administer the Spanish- and English-language versions (see Box), although far more programs used either teaching staff only or non-teaching staff only to conduct the Spanish assessment. As discussed in more detail below, this difference may be due in part to programs' likelihood of having fewer Spanish-language assessors than English-language assessors.

The number of trained assessors at each program varied considerably by program size (see Box next page). Small programs, defined as having 200 or fewer enrolled children (20 percent of the sample programs), had between three and eight trained assessors, with an average of five assessors; large programs with more than 200 enrolled children (80 percent of programs) had between 1 and 82 trained assessors, with an average of 20 assessors. The

	Percentage of Programs
No teachers	46
Teachers only	29
Mixed approach	23
N = 35 Head Start programs	

	Percentage of Programs
No teachers	48
Teachers only	43
Mixed approach	9
N = 23 Head Start programs	

average number across all 35 programs was 17 assessors. Some programs reported training and certifying additional staff as back-up assessors who could be used to conduct assessments, if necessary. Other programs trained and certified a large number of assessors, to reduce the amount of burden on any single staff person. For example, one program trained and certified all of its teachers with NRS-eligible children so that each teacher would be able to complete assessments in his/her own classroom, limiting the need for staff to travel for the assessments.

Average Number of Assessors	
Small programs	5
Large programs	20
All programs	17
N = 35 Head Start programs	
Small programs had 200 or fewer enrolled children; large programs had more than 200 enrolled children.	

Half of the programs that conducted the Spanish-language assessment had only one or two Spanish-language assessors. Comparatively, the number of English-language assessors at these programs ranged from 2 to 53 assessors, with an average of 13 assessors. In addition, programs reported that they often had few choices in selecting Spanish assessors because only a limited number of staff spoke Spanish. In fact, two programs reported being unable to assess Spanish-speaking children because they did not have access to staff or other individuals who spoke Spanish.

Nearly half of the programs experienced some turnover in the assessors who were certified in Years 1 and 2. Almost all changes were the result of staff turnover in the programs rather than changes in programs' approaches to staffing the NRS. Some programs also reported training additional assessors in Year 3 to increase the overall number of certified assessors. In addition, one program hired (but did not certify) a bilingual staff member to conduct the Spanish assessments, rather than rely on a Spanish-certified assessor from another program as it had in Year 2.

### **Programs' Rationales for Deciding Whether or Not to Assign Teachers to Conduct Assessments**

Programs' approaches to choosing staff to conduct the NRS assessments depended, in part, on their views about the advantages and disadvantages of using classroom teachers to assess the children. Over half of the sample programs assigned teaching staff to conduct some or all of the assessments. Many of these programs recognized some drawbacks to using teaching staff, such as the cost of hiring substitutes, the loss of instructional time, and the potential for teachers to coach children. One rural program reported that because it serves a large geographic area and many of its centers are far apart, assigning teaching staff was not a feasible approach. Using other types of staff both reduced the costs of implementation and guaranteed that the program would be able to complete the assessments by the deadline. However, many programs determined that the advantages of using teaching staff outweighed the disadvantages.

The most common reasons programs reported for using teaching staff to conduct assessments were that the children were comfortable and familiar with teachers; children

---

responded better to teachers; and teachers had better knowledge of children's behavior, child development, and child assessments. Some programs also reported that conducting the assessments gave teaching staff the opportunity to observe the performance of children directly and to learn more about children's knowledge in specific skill areas. The teachers could then use this information to inform their classroom practice.

Forty-six percent of the sample programs assigned only non-teaching staff to conduct the assessments. Programs that took this approach identified four primary reasons for not using teaching staff. The most common reason reported by programs was that teachers were busy with other responsibilities and did not have time to participate in training and to conduct the assessments. Second, programs reported that conducting the NRS assessments would constitute a significant loss of instructional time and would disrupt children's learning. Third, programs reported assigning only non-teaching staff to conduct the assessments because they thought teachers might coach or inappropriately encourage children. Finally, programs chose not to use teachers because of the added cost of hiring substitutes while teachers conducted the assessments.

### **COMMUNICATING WITH PARENTS AND POLICY COUNCILS ABOUT THE NRS**

Nearly all of the sample programs shared information with parents about the NRS—in parent newsletters, handbooks, letters home and through presentations at parent meetings, enrollment and orientation sessions, and Policy Council meetings. Fifty percent reported sharing the NRS outcome report(s) for Year 2 with their Policy Council. During focus groups, many assessors also reported informally discussing the NRS with parents.

Compared to the sample of programs visited in spring 2005, a higher percentage of programs we visited in spring 2006 required parents to provide written consent for their child's participation in the NRS. In spring 2005, two-thirds of the sample programs required written consent, compared to more than three-quarters of the programs visited in spring 2006. (This represents an ongoing increasing trend, as only 38 percent of sample programs required similar consent in fall 2004.) Of these, more than half distributed the consent forms to parents at registration or enrollment. The rest distributed consent forms to parents at the beginning of the program year; one program obtained written consent during each family's first home visit. In almost all of these programs, consent obtained at enrollment or at the beginning of the year covered both the fall and spring assessments; only one of the programs requested consent for the NRS again in the spring. Three-fourths of programs that required parental consent included the NRS assessment on a comprehensive consent form that contained a list of all assessments and screenings that would be carried out during the program year. The remaining programs requiring consent created a permission form specifically for the NRS.

### **Concerns Expressed by Parents and Policy Councils**

Although most programs (74 percent) indicated that parents and Policy Council members had some concerns and questions about the NRS, most reported that staff were able to address them adequately. According to staff in the sample programs, the question

raised most often by parents was how their child performed on the assessment. Parents of some Spanish-speaking children were concerned about their child's performance in English. Parents of children with disabilities also expressed concerns about whether the assessment was appropriate for their child and whether their child might become frustrated during testing. Other issues raised by parents and Policy Council members included concerns about how the scores would be used, perceived bias of some items, confidentiality of results, the quantity of assessments done in Head Start, and questions about the purpose of the NRS. Staff typically responded by telling parents that the NRS was a federal requirement providing information about how their program and Head Start were doing as a whole. To alleviate parent concerns, programs also reported providing copies of the program's baseline and/or growth report to parents, assuring parents that individual child results were confidential, and providing examples of items from the NRS assessment and role-playing activities from the training materials for the Policy Council.

#### **Concerns and Questions Raised by Parents and Policy Councils**

- Parents cannot see own child's results
- Specific NRS items perceived as "biased"
- Confidentiality of results
- Over-testing in Head Start
- Questions about the purpose of the NRS
- Questions about how the NRS results will be used
- Questions about how the results will affect the Head Start program

N = 35 Head Start programs

Staff in a few programs reported that they did not feel that they had adequate information to respond to concerns and questions raised by parents and the Policy Council. In four programs, staffs indicated that they experienced difficulty addressing parent concerns and questions because they themselves were uncertain about the purpose and use of the NRS. In these programs, staff typically desired more guidance and written materials from the Office of Head Start to answer parent questions.

### **Parent Refusals**

One-quarter of the programs that required written consent from parents reported at least one parent refusal in 2005–2006 or preceding years. All of these programs reported a small number of refusals (one to five per program). Programs typically indicated that these parents tended to refuse other assessments and immunizations offered by the program (that is, there was nothing unique about the NRS that prompted their refusal) or were concerned about the number of assessments taking place within the program.

Some programs with parent refusals contacted the parents to determine why the parent refused and to discuss their concerns about the NRS assessment. Staffs said that during these discussions they were sometimes able to respond to the parents' concerns and obtain their consent. For example, three programs stressed that the NRS is intended to give a snapshot of how Head Start programs are performing individually and as a whole to inform program improvement efforts, not to focus on how individual children perform on the assessment.

## COSTS OF IMPLEMENTING THE NRS

As in previous rounds of site visits, few Head Start directors in the sample programs were able to estimate a monetary cost of implementing the NRS. In fact, in 20 percent of the sample programs, directors reported that the program incurred no significant costs in implementing the NRS because it was viewed as a component of staff members' responsibilities and was part of their regular work day. In contrast, the remaining 80 percent of program directors reported significant monetary and in-kind costs to the program (see Box).

A major in-kind cost mentioned by more than half of the program directors was staff time.<sup>1</sup> Some felt that redirecting staff members—whether teachers or non-teaching staff—from their regular responsibilities to the NRS had implications for the quality of the program. Seventeen percent of program directors reported that their program had to pay for the costs of substitutes while teachers were assessing children. Similarly, a few programs cited the cost of overtime pay for staff working on the NRS, CBRS-related resources, and photocopying and postage.

One-third of the sample programs cited travel costs associated with NRS activities. Staff accrued travel costs when traveling to a central location for assessor training, and when traveling from center to center to conduct the assessments. Rural programs were especially likely to cite travel as a significant cost. For example, one rural program reported that travel costs associated with the NRS were high because the program's service area covers 10,000 square miles.

Only 23 percent of the sample programs estimated the monetary cost of NRS implementation. Of these, estimated costs for Year 3 ranged from \$800 to \$34,000. However, comparisons across programs are not useful, since some programs only cited the program's additional outlays for the assessments while others included regular staff time devoted to the NRS. For example, one program estimated the costs associated with completing the 2005–2006 NRS assessments totaled \$4,094: \$2,082 in staff time, \$2,116 in travel costs, \$556 in training costs, and \$150 in supplies.

### Costs of NRS Implementation Cited by Program Directors

	Percentage of Programs
Any cost	80
Staff time	57
Travel costs	34
Substitute teachers	17
Overtime pay	6
CBRS-related costs	6
Photocopying	3
Postage to return score sheets	3

N = 35 Head Start programs

Specific costs are calculated as a percentage of all 35 programs.

<sup>1</sup> Not all program directors listed specific costs of the NRS, and few calculated a dollar value of them.

## USING THE COMPUTER-BASED REPORTING SYSTEM<sup>2</sup>

The CBRS was created for the NRS to collect background information on Head Start programs and children, to facilitate the identification of children eligible for the NRS assessment, and to track the completion status of assessments. The CBRS is a web-based system where Head Start program staffs enter program-, classroom-, and child-level data. After programs enter these data, the CBRS assigns unique identification numbers to Head Start grantees, delegates, centers, assessors, classrooms, and eligible children. As assessments are completed, program staff members enter the completion status for each eligible child. Data are entered into the CBRS through a secure and encrypted Internet website that is designed to ensure confidentiality.

This section describes the experiences of the 35 Head Start programs visited during spring 2006 in using the CBRS. We discuss programs' (1) approaches to assigning and training staff on the CBRS, (2) technical capacity and use of enhanced CBRS features introduced during Year 2, (3) experiences obtaining program data on teachers and children, (4) approaches to ensuring data quality, (5) difficulties with the system and use of technical assistance, and (6) recommendations for improving the CBRS. Overall, programs have found the CBRS fairly easy to use and report that the technical assistance help line assists staff to efficiently resolve problems that arise.

### Assigning and Training Staff<sup>3</sup> to Enter the Data

The lead NRS trainer was also designated as the lead NRS data manager in nearly half of the sample programs (see Box). In other programs, a variety of staff members, including program specialists, program data managers, teachers, and administrative assistants or office managers, were assigned the role. Although some programs spread CBRS data entry across several staff members, most of the lead data managers had the sole responsibility for entering the data and maintaining the CBRS.

<b>Programs' Lead NRS Data Managers</b>	
	Percentage of Programs
Lead NRS trainer	46
Program specialists	29
Data managers	11
Administrative assistants/ office managers	11
Teaching staff	3
N = 35 Head Start programs	

The staff members assigned to enter data into the CBRS had remained the same since the NRS began in 80 percent of the sample programs. In many cases, when changes

<sup>2</sup> In this round, MPR did not conduct direct interviews with data managers. However, data managers typically reported programs' experiences with the CBRS via the lead trainer or program director.

<sup>3</sup> Because MPR did not conduct interviews with data managers, we did not gather much information on staff training on the CBRS. However, program directors and lead trainers typically provided cursory information on this training. For example, interviews with such staff indicated that data managers who had not attended the initial Training of Trainers generally became acquainted with the CBRS by reviewing the manual or by reviewing the materials that staff had obtained at the original regional training conference.

occurred, they were due to staff turnover. However, programs also assigned new staff to the CBRS because of a change in their staffing approach. For example, in one program, the lead NRS trainer was also serving as the lead data manager and entered all of the data herself. In Year 3, she trained a team of teachers and teacher's assistants to enter data and later provided checks on the data for the program as a whole before submitting them for processing.

### Access to Technology for Using the CBRS

All of the sample programs reported that they had an adequate number of computers available for using the CBRS. All but three programs had Internet access available for using the CBRS at the start of NRS implementation. Two programs had an Internet connection installed to accommodate the CBRS, and one upgraded its connections. In addition, a few programs purchased or planned to purchase new computers that would be used for the CBRS. In most cases, however, the new computers were part of a general program upgrade and not purchased specifically to accommodate the CBRS.

### Experiences Using the Enhanced CBRS Features Introduced in Year 2

Based on experiences and feedback in the first year of NRS implementation, including data from the Year 1 Quality Assurance Study, the CBRS was updated with several new capabilities for fall 2004. These features focused on reducing the data entry burden for local Head Start staff and making the system easier to navigate. Enhanced features that allowed programs to compare themselves to similar programs and that allowed programs to generate CBRS reports in Spanish were also introduced in Year 2. In spring 2005, 37 percent of the programs in the sample had reported using at least one of the enhanced features of the CBRS. In the spring 2006 sample, 43 percent of programs reported using at least one of these features. Most features are specifically designed to assist programs in the fall when all of the new data need to be entered. As a result, most programs had not accessed these features since the fall and might have misreported their use of the features since many months had passed. During site visit interviews, lead NRS data managers mentioned five specific enhanced features that were helpful to them: (1) data copy, (2) data import, (3) search capability by child's ID number or name, (4) expanded report, and (5) a form for entering multiple children at once.

Use of Enhanced CBRS Features	
	Percentage of Programs
<b>Any new feature</b>	43
Data copy	31
Data import	17
Search by ID number or name	14
Expanded report	11
Form for entering multiple children at once	3
N = 35 Head Start programs	

As in fall 2004 and spring 2005, data copy was the most popular enhanced feature, used by 31 percent of programs (see Box). It offered a one-time-only ability to copy information on centers, classrooms, and teachers from the previous year. The data could then be edited and updated as needed. Most programs that used this feature described it as helpful and time saving, since information such as addresses and teacher information did not need to be

re-entered. One data manager reported, however, that she was unable to copy the data when trying to use this feature.

Another enhanced feature of the CBRS was the data import feature, designed to facilitate importing demographic data on children from the local program's management information system (MIS). Head Start programs typically use software, such as the Head Start Family Information System (HSFIS) or Child Plus, from one of a few specialized MIS vendors, to maintain their records on children and families. To use the CBRS data import feature, however, the MIS software had to have the capability of exporting data in a format the CBRS could read. Xtria, the subcontractor responsible for the CBRS, worked with a number of the vendors to develop the required software. In spring 2006, six programs reported using this feature. A few other programs reported that although they were interested in using the data import feature, the MIS software vendor was offering the export capability at a price higher than they were willing to pay.

Programs also found the capability to search ID by number or name, the expanded report, and the form for entering multiple children at once to be helpful enhanced features of the CBRS. One program reported that the first feature was very useful when children had the same first or last name, while another found the last feature to be particularly helpful because it allowed the data manager to enter changes for several children at once. Only a few programs indicated that they had not yet used any of the enhanced CBRS features introduced in Year 2.

### **Accessing Technical Assistance**

Data managers in one-quarter of the sample programs reported using the CBRS technical assistance help line operated by Xtria at some point since the NRS was first implemented. Most of the programs reported accessing technical assistance only one or two times. Programs reported that many questions were answered immediately. If a problem could not be resolved immediately, it was generally resolved within 24 hours. As in the Year 1 and 2 rounds of site visits, programs generally reported that the assistance they received was prompt and helpful, and that the help line staff were polite.

Programs sought help with a range of topics on the system in the 2005–2006 program year. For example, data managers at one program had basic questions about moving from one screen of the CBRS to another. Another program experienced difficulty in rolling over child cases from Child Plus into the CBRS. Another program had concerns about children being dropped from the CBRS. Data managers explained that the CBRS was not allowing them to delete children from the system, even if the child was no longer enrolled at the program. Other requests involved questions about the program's designation as rural or urban in the CBRS, and issues with basic computer glitches.

### **Suggestions for Improving the CBRS**

For the most part, local Head Start programs continued to praise the CBRS for its clarity and ease of use. Nearly all programs described the system as user-friendly and easy to



navigate. Most data managers described the training materials as clear and the technical assistance help line as helpful and responsive. However, staff members in some programs were not entirely satisfied with the training received. For example, one data manager had limited computer knowledge and suggested having separate training sessions for less computer-savvy users at regional training events. Most programs that used the enhanced features of the CBRS introduced during Year 2 reported that the features were easy to use and an improvement over the original system introduced in fall 2003.

These enhanced features addressed many of the concerns that had been raised about the CBRS in the Year 1 Quality Assurance Study. Based on their ongoing experiences with implementing the CBRS during Year 3, several programs mentioned new concerns that suggest possible areas in which the system could be improved further. The recommendations made by programs in spring 2006 largely coincided with those made by the programs visited in spring 2005:

- ***Provide more support for the data import feature.*** Although the CBRS now offers an import capability, programs must purchase an export feature from their MIS vendor to use the new software. Some Head Start programs reported that the cost of the feature was more than they could afford; perhaps additional funding could be made available to programs to expedite use of this data feature.
- ***Provide more support for the data copy feature.*** The data copy feature offers programs a one-time opportunity to copy the data from the previous year. While many programs found this feature helpful and time-saving, some programs complained that the feature did not work properly or that they were unable to use it. Programs that were unable to use the feature may have missed the one-time opportunity to use it.
- ***Add features to make the CBRS even more user-friendly.*** Most programs described the CBRS as a user-friendly system; however, some data managers recommended further improvements, such as further easing the process of navigating between screens and using features like data copy, adding a message that alerts the user when data has been sent to Xtria, providing a feature that maintains center and teacher information within the system across years, and easing the process of locating relevant child information (without having to perform multiple searches or search individual records).

## SUMMARY

Overall, the 35 Head Start programs we visited in spring 2006 took an approach to NRS implementation similar to that of programs visited in previous rounds of site visits. Program directors assigned a lead NRS trainer the responsibility of overseeing implementation, including assigning, training, and certifying assessors; scheduling and tracking the completion of assessments; overseeing quality assurance activities; and submitting score sheets by the

deadline set by the Office of Head Start. In more than a third of programs, the lead trainer was also responsible for maintaining the CBRS.

Most programs maintained the same basic staffing structure for the NRS in the spring that they had instituted in the fall. When programs did make changes, most were due to staff turnover within the program rather than a rethinking of the program's approach to the NRS. On average, the programs trained 17 assessors. As in the sample of programs we visited in spring 2005, most programs did not rely exclusively on teaching staff to conduct the assessments. Program staff saw many advantages to having teachers administer the assessment, including that the children were comfortable and familiar with teachers; children responded better to teachers; and teachers had better knowledge of children's behavior, child development, and child assessments. Learning more about children's knowledge in specific skill areas could inform teachers' classroom practice. However, many programs expressed concerns about teacher burden, reduction in instructional time, the potential for coaching or bias in administration, and the cost of hiring substitutes to cover for teachers while they were assessing children.

Nearly all programs communicated with parents and Policy Councils about the NRS. While the means of communication varied, programs made an effort to inform parents of the assessment at the start of the program year. Three-quarters of programs sought to obtain parents' written consent for the NRS assessment, far more than in previous rounds of site visits. Most programs had few or no parent refusals. Concerns among parent and Policy Council members were similar to those expressed in previous rounds of visits. They included parent requests to see their child's individual results; concerns about bias of specific items; and questions about the purpose of the NRS, how the results would be used, and how the results would affect the Head Start program.

Few Head Start directors estimated the monetary and in-kind costs of NRS implementation. However, a number of directors identified significant costs, such as staff time, travel costs, costs for paying substitute teachers, and overtime for staff working on the NRS.

As in past rounds of site visits, programs reported that the CBRS was easy to learn and use. Programs had adequate numbers of computers and Internet connections for accessing the system. Many programs used and liked the enhanced features of the CBRS introduced in Year 2, especially the data copy, data import, and expanded report features. Some programs reported accessing the CBRS help line at least once and most of these reported receiving prompt and helpful support from the help line staff. Data managers made some suggestions for further enhancements to the CBRS, including reducing the cost of the data import feature, providing more support for using the data copy feature, and adding other user-friendly features.

## CHAPTER V

### USING THE NRS FOR LOCAL PROGRAM IMPROVEMENT EFFORTS

---

A fundamental goal of the NRS has been to support local Head Start programs in their program improvement efforts. After each round of assessments, the Office of Head Start provides grantees and delegate agencies with summary reports that present program-level results of how children performed in the four skill areas covered by the NRS: (1) English Language Screener, (2) Vocabulary, (3) Letter Recognition, and (4) Early Math Skills. Programs can then compare their average scores with the national averages for all Head Start programs, as well as with results from programs similar to their own (for example, other programs in their federal region or programs serving similar populations). These average scores can be used in concert with other data sources to help guide programs in their quality improvement efforts.

In winter 2006, the Administration for Children and Families distributed the second NRS growth reports, covering the 2004–2005 program year, along with a report on spring assessment outcomes. Unlike other NRS reports that document how children performed on either the fall or spring assessments, the growth report documents progress made during the Head Start year among children who were assessed in both fall 2004 and spring 2005. Accompanying materials offered guidelines and suggestions for reviewing the reports and using them for local program improvement efforts. Our spring 2006 site visits afforded MPR an opportunity to discuss with local Head Start staff their views on the usefulness of the report, how they have used the NRS results thus far, and how they plan to use the reports in the future. In spring 2006, we also obtained more detail about programs' use of local child assessments and how programs felt they compared with the NRS.

This chapter begins with a discussion of the reactions of program staff to the 2004–2005 growth report, including their perspectives on the report's usefulness to their local programs. Next, we describe how programs have used the growth report for program planning and any specific changes they have made to their classroom practices based on their NRS results. We include information about what assessment tools programs use for their

local assessments, and the types of information these provide. We conclude the chapter by describing how programs intend to use the NRS reports in the future.

### **REACTIONS OF LOCAL PROGRAM STAFF TO THE 2004–2005 GROWTH REPORT**

As in previous rounds of site visits for the Quality Assurance Study, a key objective was to learn about the views of local program staff on the NRS reports they had received. The most recent reports distributed by the Office of Head Start prior to our site visits presented results for the spring 2005 assessments, as well as the 2004–2005 Growth Report that included results from children who were assessed in both the fall and spring of that program year. Both reports included average results for the local program and national averages across all Head Start programs. Staff could also download reference tables from the CBRS that compared results for their program to average results for programs with similar characteristics, such as programs in their federal region, programs with similar proportions of English language learners, or programs in either rural or urban settings.

The 2004–2005 Growth Report was divided into two main sections (see a sample program report in Appendix D). The first section provided programs with the HSNRS Spring 2005 Child Assessment Report for their program. This comprised a set of tables that presented information on the skill levels of all children in the program who were assessed at the end of the 2004–2005 program year, regardless of whether or not they were assessed in fall 2004. These tables were presented separately for children assessed in English and in Spanish. For each of the four skill areas (Language Screener, Vocabulary, Letter Recognition, and Early Math Skills), the program received its mean score, an average (median) skill level from 1–6, and the percentage of children at each skill level. In addition, for each skill area, the Office of Head Start provided the Head Start national average and national skill level, based on data from all available programs during the same data round.

The second section presented “Fall-Spring Growth Charts”—bar graphs that compared the percentage of items correct in each skill area in fall and spring. For example, children may have given correct responses to 55 percent of vocabulary items in the fall, and 71 percent in the spring. “T” bars superimposed on each bar in the graphs indicated the national average for each skill area in the fall and spring, enabling programs to compare scores for their children to national averages. The report also included growth charts for different groups of children according to their language background. Separate bar graphs were presented for four groups of children: (1) all children assessed in English, (2) all children assessed in Spanish, (3) all native English speakers assessed in English, and (4) all English language learners assessed in English.

The second section of the 2004–2005 Growth Report displayed a series of four bar graphs (“Skill Level Growth Charts”) that compared the percentage of children at particular skill levels in the fall and spring for each of the four skill areas. For example, for children who are English language learners, a bar graph showed the percentage of children with limited or no understanding of English in the fall and in the spring. Likewise, another bar graph compared the percentage of children who identified 17 or more alphabet letters in the

fall and spring. As in the first section, “T” bars superimposed on each bar in the graph showed national averages.

Below, we describe the reactions of local program staff to the 2004–2005 Growth Report, including staff impressions of how well the NRS results aligned with their expectations of how the children would perform, their views on the usefulness of the NRS results to their programs, and their suggestions on how to improve the NRS reports.

### Program Perspectives on Growth Score Results

During on-site interviews and focus groups, we asked local program staff if the children’s outcomes presented in the 2004–2005 Growth Report were higher, lower, or about what they expected prior to receiving the report. Staff from 57 percent of the sample programs agreed that the results were in line with their expectations of how much children had learned during the Head Start year (see Box). Several staff members noted that the NRS Growth Scores were very validating; perceived strengths and weaknesses were confirmed. One director noted that the NRS results played an important role in validating the strengths and weaknesses of the students in their classrooms. The director noted that the NRS and the local assessment scores were “pretty much hand-in-hand” with each other. One program reported that the percentage change in average scores exceeded their expectations.

	Percentage of Programs
In line with expectations	57
Lower than expected	23
Higher than expected	3
Mixed views among staff	11
No impression reported	6
N = 35 Head Start programs	

In another 23 percent of programs, staff agreed that results were lower than expected in at least one skill area. One program was “shocked” by below-average scores, since special efforts had been made in the past year with regard to letter recognition. Staff members from a few programs were surprised by lower program scores in specific skill areas. One program director noted that the scores were a real “eye opener” for the teaching staff. The scores clearly indicated that additional instructional work was needed in the classroom. As a result, the program plans to spend time on specific skill areas, such as helping children identify letters.

In 11 percent of the sample programs, perspectives on the NRS results were mixed, with staff having different opinions on the scores. In these cases, growth scores were lower than the assessors’ expectations, but they were on par with what the directors or lead trainers had anticipated.

### Usefulness of the NRS Reports for Local Programs

The degree to which programs found the reports useful and relevant for program improvement efforts is an important factor for the Office of Head Start to consider as it moves forward with NRS system improvement efforts. The views of local program staff are

an indicator of how likely programs are to use the NRS reports as envisioned—that is, using the NRS as one of several tools to inform local decisions on strategies for improving program quality and effectiveness.

Compared to last year, where less than half of the programs remarked that the Growth Report was easy to understand, this year’s interviews revealed that nearly 75 percent of the programs viewed the reports as user-friendly. Staff members commented on how much they appreciated the bar charts and reference tables; they offered program staff a visual representation of the programs’ growth and progress. Many programs found that the data were very clear and easy to understand.

Of the 17 percent of programs that had mixed or negative views, one lead trainer noted that the report was too technical. She stated, “We are educators, not statisticians.” Several other lead trainers agreed that changes should be made to present the data in such a way that it would be quickly understood by parents and program staff.

In addition, 17 percent of programs thought that the delayed timing of receiving the growth report—midway through the 2005–2006 program year—diminished its usefulness. Several respondents noted that it would have been helpful to receive the NRS results closer to the end of the 2004–2005 program year to aid in program planning.

### Recommendations for Improving the NRS Report

Head Start programs offered several recommendations to improve the NRS Growth Report, both in terms of content and format. Most programs would like to see the results broken out beyond the program level (see Box). Sixty-three percent of programs requested the addition of either center or classroom level. Twenty-six percent of the programs would like to see results at the individual child level.

Staff from 46 percent of the sample programs offered a range of other suggestions to enhance the way in which outcomes are presented in the Growth Report (see Box next page). With regard to content, some programs would like outcomes to be reported for

#### Views on the 2004–2005 NRS Growth Report

	Percentage of Programs
Program views on format	
Clear and easy to understand	74
Difficult to understand	4
Mixed opinions on report’s format	3
No information	9
Program requested technical assistance to interpret the results	3
Report received too late	17
N = 35 Head Start programs	

#### Preferred Level of Aggregation for NRS Results

	Percentage of Programs
Program plus center or classroom	63
Program only	37
Individual child level	26
N = 35 Head Start programs	

certain subgroups of children, including by geographic differences (urban vs. rural), use of different specific curricula, part-day versus full-day, and children who are English language learners (although some separate information is already provided for the last group). One program wanted outcomes broken out by gender. Another program mentioned that it would be helpful to have data reported by teacher qualifications.

Some directors thought it would be useful if NRS reports automatically included comparisons at the state and regional levels, without requiring that they generate this information from the CBRS. One delegate would like to be able to compare its skill levels with other delegates within its grantee agency. Staff from a few programs wanted more specific guidance on ways to use the NRS results. They

requested additional technical assistance to better understand the results (for example, what is considered an adequate average score and which specific skills each section of the NRS is testing). If programs understood the intent of each item, staff felt they could better target their efforts in the classroom.

Several programs indicated that while the layout of the report was appealing, the colored text did not lend itself to photocopying. The report should either be produced in black and white, or programs should receive a master copy in color along with a black and white version that can be reproduced. Programs also suggested sending out electronic versions of the report with tables and graphs in PowerPoint, which would help staff prepare presentations. One program suggested creating a simplified version of the Growth Report to share with parents and members of the Policy Council. One program suggested that one way to present the information to “lay” audiences would be to indicate, for the various components of the assessment, the percentage correct by the program compared with corresponding national percentage (this presentation of information is already included in the Growth Report).

#### Suggestions for Presenting the NRS Results

	Percentage of Programs
Offered suggestions	46

#### Content suggestions:

- Provide comparisons with non-Head Start children
- Provide national benchmarks for performance
- Provide information on how much progress different subgroups of children made (gender, teacher qualifications, curriculum)
- Provide information on individual items
- Present data results for state and region
- Give more detailed guidance on using data
- Offer additional explanation of results

#### Formatting suggestions:

- Make report more easily photocopied
- Send electronic versions of reports
- Format graphs and tables in PowerPoint
- Produce a simplified version for parents
- Provide report in a Word file

N = 35 Head Start programs

## HOW PROGRAMS HAVE USED THE GROWTH REPORT

In the 2004–2005 Growth Report, the Head Start Bureau included some general guidelines and suggestions for incorporating NRS results into local program planning in effective, appropriate ways. Although some programs reported that they preferred using local assessments to inform decisions about their curricular and professional development needs, more than 90 percent of programs identified at least one way where they have shared or made use of the NRS results. In this section, we describe the extent to which program managers have shared NRS results with staff and key stakeholders and how programs have used NRS reports. The section following this one explores in greater detail how programs have made specific changes to classroom practices.

### Sharing NRS Reports with Local Program Staff and Stakeholders

All of the sample programs shared the NRS results with staff beyond the directors and lead trainers who were directly responsible for overseeing NRS implementation (see Box). More than three-quarters distributed results across various levels of staff, including education specialists, center managers, NRS assessors, and teachers. Sometimes the dissemination of information occurred in a linear fashion. That is, the director or lead trainer first briefed an assistant director or area managers, then program specialists, and finally teachers. Alternatively, the director or lead trainer shared the NRS results with the

How Programs Shared NRS Results	
	Percentage of Programs
Shared NRS results beyond director and/or lead trainer	100
Shared NRS results with	
Policy Council	94
Management team	83
Program staff	77
Board of directors or equivalent	45
Parents	43
Grantees	14
N = 35 Head Start programs	

education coordinators before presenting them at an all-staff meeting. Most programs shared the 2004–2005 Growth Report with stakeholders beyond local staff. Ninety-four percent of the programs presented the results to their Policy Councils, and 43 percent shared the results with parents (compared with 23 percent in spring 2005). Some programs opted to share the results with parents in the form of formal parent-teacher conferences; other programs shared the results with parents if the parent requested the score results. In addition, 45 percent presented NRS results to a board of directors or its equivalent (for example, a tribal council), up from 37 percent in spring 2005, and several programs presented them to their administrative offices, or grantees.

Local programs used a variety of methods to disseminate the NRS results to staff, parents, and other stakeholders. Program administrators typically shared the data with local staff at meetings, though a few respondents mentioned reviewing them during in-service training days. Some programs opted to share information with staff members individually (for example, an education coordinator meeting with a teacher one-on-one to talk about



---

the results). Typically, Policy Council members learned of the NRS results at regular monthly meetings.

The degree to which programs shared details of the growth report varied considerably. In some cases, programs shared copies of the report with every staff member who works with the children. In other cases, the teachers were simply notified of how the children were doing overall. Some programs opted for more informal approaches, which seemed to be more common when addressing those outside local program staff. For example, three directors explained that they very briefly told their Policy Councils about how the program performed against the national averages but did not go into detail. Another program conducted a formal presentation to the Policy Council board. Results provided to the Policy Council were then sometimes shared with other community stakeholders for public relations purposes. At least one program transformed the information into a more succinct snapshot of information that it gave out to less statistically savvy audiences, including staff, parents, and community stakeholders.

### **Making Changes in Staff Development**

Forty-six percent of programs reported using NRS results to inform staff development. Frequently, they targeted training efforts to better integrate developmentally appropriate pedagogical practices in math and literacy skills into the classroom. Some programs held in-service workshops, offering guidance to teachers on implementing activities on numeracy, including charting and graphing instruction, and literacy, including “taking a holistic approach to letters and sounds” and brainstorming teaching practices for letter naming. Two programs emphasized training teachers to build on “teachable moments” in the classroom, in one case, to expand vocabulary. In one program, training was also provided to help teachers better engage parents in instructional practices. One program had education mentors work with teachers to add increased emphasis to activities that were already ongoing.

### **Tracking Results for Individual Children**

As in past rounds of site visits for the Quality Assurance Study, some of the sample programs (11 percent) opted to collect and track item responses to the NRS assessment in an effort to identify skills that children did not perform well. These data are typically used for training purposes, or to inform classroom practices. Generally, programs take the time to track and analyze the raw data because NRS results are only provided at the program level, and they do not consider this level of reporting useful for program improvement efforts.

To track individual results, staff collected completed Scantron answer sheets and recorded item-level responses. Two programs limited their tracking to the Letter Naming section to show teachers the letters children did and did not know. Another program mentioned tallying results to get information at the classroom level.

## AVAILABILITY AND USE OF LOCAL ASSESSMENT RESULTS

Since 2000, Head Start programs have been required to carry out local assessments of child outcomes based on the Head Start Child Outcomes Framework (ACYF 2003), and to use these results to individualize work with children, as well as for aggregating the data to inform program self-assessment. The tools used for these assessments are often linked to the specific curriculum used in classrooms, but some programs have also tailored these tools for local needs or to better align with requirements of local school districts. For the first time in the current round, we asked more-specific questions about the local assessment tools used by sample programs. The results are outlined in Table V.1. As the table shows, just over half of the programs used Creative Curriculum’s assessment tool, which is completed three times a year. Programs varied more in the types of developmental screeners they used than in the types of local assessment tools. Several programs reported using a combination of developmental assessments and local screeners. Local assessments are carried out by teachers and teaching assistants.

Staff reported that they were generally very satisfied with their local assessment tools—69 percent expressed satisfaction. They highlighted flexibility, timeliness, and the ability to obtain detailed information on the progress of different subgroups as advantages of the local assessments. For example, staff praised the design of the assessments that allowed for timely provision of outcome data about individual children: “The outcomes of the local assessments are ready the day they are administered by teachers. This makes the local assessment tools very useful.” Staff liked that their local assessments provide an up-to-date picture of the progress and development of children that can be viewed immediately and shared in discussions with parents. Other staff members highlighted that results could be viewed across different levels, such as centers, classrooms, and individual children.

Some staff contrasted the usefulness of the local assessment results with the usefulness of the NRS. One program emphasized the availability of local results to improve classroom practice. Teachers can use the data to determine areas in which individual children are not performing well and organize their instruction accordingly. One teacher contrasted the program-level NRS data with local assessment data: “The NRS is really not beneficial to teachers in general. It’s more beneficial to determining how Head Start is doing their job, since we don’t really get results on [children] in a fashion that would benefit us.”

A few programs had worked to develop a local assessment that was tied to their curriculum. While the effort required was termed “tremendous,” staff felt that their work had paid off in the end. One lead trainer highlighted the fact that because their assessment is tied both to the Head Start curriculum and the public school curriculum, there is a “seamless academic pathway to the kindergarten and the elementary grades from Head Start.” Creating this alignment was the major motivation for the choice of curriculum and local assessment.

**Table V.1. Developmental Screeners and Local Assessments**

	Number of Programs <sup>a</sup>	Percent of Programs
Developmental Screener <sup>a</sup>		
Brigance	8	23
ESI	7	20
DIAL 3	4	11
Denver II	4	11
DECA	4	11
LAP-D	2	5
Battelle	2	5
Other	16	46
Local Assessment		
Creative Curriculum	18	51
HighScope COR	7	20
Locally designed	4	11
LAP	3	9
Other	3	9
Conducted by:		
Teaching staff	35	100
Reviews Results <sup>a</sup> :		
Teaching staff	27	77
Management team	29	83
Policy Council	33	94
Parents	15	43
Board	16	46
Grantees	5	14
Uses of Local Assessment Results <sup>a</sup> :		
Target staff training and TA	27	77
Inform changes to classroom practices	26	74
Individualization for children	13	37
Program planning, assessment and improvement	24	69
Satisfaction with Local Assessment:		
Satisfied	24	69
Satisfied but with some concerns	10	29
Not satisfied	0	0

Source: Head Start National Reporting System Quality Assurance and System Development Project, Year 2 Spring 2006 data collection, interviews with program directors and lead trainers.

Note: N = 35 programs. Note that not all programs provided information.

<sup>a</sup>Does not total 35 because some programs provided more than one response.

In contrast, some programs expressed concerns about the chosen local assessment tool. Twenty-nine percent of programs noted some concerns about their assessments. Some staff asserted that the assessments were carried out in such a way that there was room for subjectivity by the assessor. Staff members at another two programs argued that some of the questions on local assessments were not culturally sensitive. Others expressed concern about whether their local assessment met the needs of Spanish-speaking children in their programs.

## Comparing NRS Reports with Other Assessment Results

As part of their reviews of the 2004–2005 Growth Reports, programs reported comparing NRS outcomes to outcomes from two other assessments: (1) their local assessment results and (2) NRS results for similar programs. Half of the sample programs compared their NRS growth scores with local assessment results for the domains covered in both assessments. For the most part, staff compared the two assessments on a fairly informal level, mostly to reinforce and verify the results of the local assessment. All programs that compared their NRS results to local assessments stated that the outcomes from the two tests were similar; the comparisons did not yield surprises. Several programs elected not to compare the NRS reports to their local child outcomes, often explaining that the formats of the instruments were too dissimilar, or that they did not have much faith in the NRS.

Approximately 14 percent of programs found it useful and interesting to compare their NRS results with those of similar programs by downloading the reference tables available on the CBRS. Programs generally reviewed these reports to determine whether their programs were performing at a similar level as programs with comparable characteristics. One director said it was useful to “get confirmation about what you’re doing.” Staff reviewed reports that compared their program to other programs in their federal region, state, and those with similar demographic characteristics (for example, urbanicity, or percentage of English language learners). One program shared these comparisons with a parent committee, and another shared them at a meeting with other program directors. Of the programs that did not compare the NRS results with other similar programs, approximately a quarter stated that they did not know that feature was available to them. Another quarter said that they did not have the time to make the comparisons. An additional 25 percent stated that they did not compare because of concerns about validity of NRS results, while the remaining quarter of the programs did not provide a reason.

This year, we also asked programs if they used the NRS growth reports to make comparisons across multiple years. Just over half of the programs conducted such an analysis. Forty-four percent of the programs noted that the children appeared to be improving. One program director noted that they use the NRS reports to identify areas that need improvement, and then work to concentrate staff efforts and classroom lessons on those areas. The director stated that this practice has produced good results. They have seen a marked improvement in the focus areas. Thirty-three percent of the programs that did compare the NRS results across the years noted that the strengths and weaknesses of the programs were consistent from year to year.

Finally, this year we asked programs whether the NRS provided any new information beyond what they learned from local assessments. Seventeen percent said that the results confirmed their local assessment findings, while fourteen percent mentioned that the NRS provided more information about ELL performance than they would otherwise have obtained. For some, seeing improvement in alphabet knowledge in English was “common sense” because they are instructed in it during the year. But another staff member noted that the results helped them to realize the “geographical and cultural differences between groups of students.” Individual programs mentioned specific information that the NRS added,

including more information on math performance, letter recognition, and vocabulary. Two programs noted the usefulness of the NRS as a tool for comparing their program with national averages.

### CHANGES MADE TO CLASSROOM PRACTICES

Eighty percent of Head Start programs reported modifying their classroom practices in response to the NRS (see Box next page). For the most part, programs have placed greater emphasis on skill areas included on the NRS assessment during classroom instruction. More than half of the programs mentioned focusing more attention on cultivating language and literacy skills, such as alphabet knowledge, building children's vocabulary, and reading to children. Letter naming was an emphasis for nearly half of the programs. Often, children could associate the sound that the letter made with the letter but could not 'name' the letter. Programs increased their focus specifically on letter naming. One program had given increased attention to building children's vocabulary and language skills. Nearly half of the programs also have placed greater importance both on developing math skills and expanding their classroom instruction to include simple addition and subtraction, counting, shapes, graphs, and math vocabulary.

More than one-quarter of the sample programs have used the NRS as a guide for purchasing classroom materials. Books and other items to support literacy development were most common. Staff from another program aimed to promote geography and purchased a globe for the children to use. A few programs acquired classroom materials that centered on math instruction, focusing on math manipulatives charts, rulers and graphs.

To supplement the curricular changes taking place at their Head Start centers, 11 percent of programs asked parents to become more involved in nurturing their child's literacy and numeracy skills outside of the classroom. Additionally, one program has encouraged parents to join the Parents as Reading Partners group, which aims to enhance the children's learning environment by exposing parents and children to language. Teachers from these three programs reported talking regularly to parents about activities they could incorporate into their daily routines at home to promote skill areas, such as number recognition, counting, and letter naming.

<b>Changes in Classroom Practices as a Result of the NRS</b>	
	Percentage of Programs
Any changes to classroom practices	80
Increased focus on:	
Letter naming	47
Counting/number recognition	17
Graphing/addition/subtraction	14
General literacy skills; reading	3
Vocabulary	3
Purchased classroom materials	26
Materials purchased	
Books, literacy materials	9
Math materials, charts, graphs	9
Encouraged parents to supplement classroom learning at home	11
N = 35 Head Start programs	

## FUTURE PLANS FOR USING THE NRS REPORTS

While visiting Head Start programs in spring 2006, we asked staff if they planned to use the NRS results in the future. About 57 percent indicated that they did intend to make use of the NRS results (see Box next page). Most programs described fairly general plans, such as using NRS data for informing program planning decisions, identifying training needs, and determining which skill areas required more attention through classroom instruction. No program planned to use the NRS as the sole source of information for planning program improvement efforts. A few programs planned to compare their program's NRS scores with those of other local programs, which they had not

done in the past. Eleven percent of the programs hoped to use the NRS scores to observe overall longitudinal trends in student achievement. Three programs stated that they planned to use the NRS results as a sort of self assessment. They hoped that the results would provide them with feedback on which aspects of their instruction should continue and what areas need improvement. In that way, staff reported, the NRS acts as a "measurement tool of where the program is and where it needs to go."

Five programs indicated that they did not plan on using the NRS results in the future. One program found several troubling areas in the NRS growth reports. The program asserted that the results were not specific enough. They felt that, because data were not reported at an individual level, this reduced the usefulness of the NRS as an assessment tool. Secondly, the program felt that other important domains of a child's development were not assessed using the NRS assessment. A second program believed that the NRS was not linked to the program's curriculum. In order for the assessment to be useful, a more explicit link is needed between what is going on in the classroom and what is being assessed. Two programs asserted that they do not plan to use the growth reports in the future because the formats of the reports were not useful; they would prefer data to be presented at the classroom or center level. Finally, staff in one program noted that they would not plan to use the NRS reports in the future unless it was made clear that the NRS results are to be used for program improvement. The lead trainer of this program said that her understanding of the message being sent from the Office of Head Start was that it was important to collect the data but that it was not to be used explicitly for program improvement.

<b>How Programs Plan to Use the NRS Results</b>	
	Percentage of Programs
Intend to use NRS results	57
No intention of ever using NRS results	14
Undecided about the use of NRS results	6
Data unavailable	3
<b>Specific plans:</b>	
Program planning/curriculum changes	17
Teacher training	6
Compare program to other local programs	11
Observe trends on a year-to-year basis	11
Provide self assessment	9
N = 35 Head Start programs	

---

Two programs were unsure of how the NRS reports would be used in the future due to anticipated personnel changes. Finally, eight programs did not provide any information on future plans to make use of NRS reports.

## **SUMMARY**

The majority of Head Start staff who participated in the site visits found the 2004–2005 Growth Report easy to understand, although some said it was somewhat confusing, particularly for non-technical audiences. Staff made several recommendations to improve the report's content (for example, provide information at different levels of aggregation, such as the classroom) and format (for example, distribute a version that yields better photocopies). Staff in about two-thirds of the sample programs agreed that center or classroom-level reporting would be more useful in targeting practices and resources; about a quarter of programs requested individual child-level data. A majority of the sample programs shared NRS results with staff and key stakeholders, such as managers, specialists, and teachers; Policy Councils; boards of directors; and, to a lesser degree, parents. Some programs shared details on how they have used NRS data to modify classroom practices, such as spending more time on alphabet knowledge, literacy development, and, to a lesser degree, counting and other early math skills. Because the NRS reports did not provide center-, classroom- or child-level information, 11 percent of the programs reported tracking the item responses of individual children using the raw data on the Scantron answer sheets to make decisions on what areas needed improvement to better prepare children for kindergarten. Half of these programs collected information on letter naming only.

With regard to using the NRS results in the future, more than half of the sample programs said they planned to use the reports to some degree. In many cases, staff did not provide much detail on these plans, but they intend to continue incorporating NRS outcomes as one source of information for program-planning efforts. About 20 percent of programs either remained undecided about whether they would use the results or did not plan to use them at all. However, 11 percent of programs planned to observe trends over time, now that multiple years of data are available.

**This page has been intentionally left blank for double-sided copying.**



## CHAPTER VI

### PERSPECTIVES OF LOCAL HEAD START STAFF ON THE NATIONAL REPORTING SYSTEM

---

This chapter describes the views of local program staff about the NRS as they complete the third year of implementation. The perspectives of Head Start staff can serve as important sources of feedback for the Office of Head Start on how to support programs implementing the NRS. It can also help shape clear and targeted messages to programs about the purpose and anticipated uses of the NRS and enhance its ongoing development.

We begin this chapter by discussing the main contributions of the NRS that local program staff identified during the spring 2006 site visits. We then describe their overarching concerns and the implications these have for future directions that Head Start might take. Finally, we articulate the main suggestions of program staff for improving communication and planning, and improving specific aspects of the assessment and its procedures.

#### **CONTRIBUTIONS OF THE NRS**

When asked to list the contributions the NRS had made to their programs, staff members in sample programs were able to attribute a number of positive effects of the NRS. Primarily, the contributions that staff listed were that it (1) helped to improve classroom practice and inform teacher training, (2) proved Head Start is effective (3) raised program accountability by providing national comparisons, and (4) validated local assessment results. In a few programs, a contribution of the NRS was to raise staff morale. In two programs, staff indicated that a contribution of the NRS was that it helps to get children accustomed to a testing environment.

Nearly three-quarters of sample programs mentioned that the NRS helped them to change instructional practice in the classroom by showing areas of strength and weakness.

Some programs specifically mentioned they had increased emphasis on letter recognition and naming, introduced more-difficult vocabulary, and taught new math concepts (mainly graphs). Staff in a few programs described their surprise that children knew fewer letters than expected, while others were pleased that children performed well on such areas. One program said that the NRS confirmed that the curriculum they were using was a good one. Staff reactions to NRS reports and results for their programs are described in detail in Chapter V.

Twenty-nine percent of programs cited fall to spring growth and stated this “proved” Head Start was effective. A similar percentage felt that the NRS increased program accountability, and making comparisons to other programs was a source of motivation to perform better. Others felt that proof of effectiveness was beneficial at the national level, to show that Head Start children could learn and that the programs were working. Some also stated that the NRS helped them understand what the federal government’s priorities are and that a national assessment raised awareness about the program.

<b>Contributions of the NRS</b>	
	Percentage of Programs
Helped improve classroom practice	63
Proved Head Start works	29
Raised program accountability	29
Raised teacher morale	11
No contributions	34
N = 35 Head Start programs	

Some programs (17 percent of the sample) indicated that that the NRS provided useful information that augmented or validated local assessments. One program indicated that the NRS showed it was possible to test 4-year-olds. One program said results were useful to them at the local level to help involve parents in the program.

Finally, in 11 percent of sample programs, a contribution of the NRS was improved teacher morale. Staff said that observing improvements from fall to spring showed that they were making a difference, that the children could learn, and that Head Start teachers were professionals and not “just babysitters.”

Staff in more than one-third of programs said it had made no contributions.

## **CONCERNS ABOUT THE NATIONAL REPORTING SYSTEM**

To address effectively program concerns and doubts about the NRS and provide needed information or training, the Office of Head Start must first have a clear understanding of these concerns. During site visits, we asked respondents the open-ended question, “What are your strongest concerns about the NRS?” Responses clustered around four primary concerns: (1) lack of clarity about the purpose of the NRS and how the results will be used, (2) the financial resources and staff time devoted to the NRS, and (3) the ability of the assessment to accurately reflect children’s abilities and programs’ performance. These concerns are similar to those reported during previous rounds of site visits, except that

programs rarely mentioned interpretation of NRS results. Details of their concerns are below:

### Purpose of the NRS and Use of Assessment Results

In more than two-thirds of the sample programs, staff expressed concerns about the purpose of the NRS (see Box). As in prior reports (Paulsell et al. 2003, 2004, 2005), staff in most programs said they did not have a clear understanding of the purpose of the NRS. Not understanding the purpose of the NRS led some program staff to voice the ways they feared the data would be used at the national level and what they saw as potential consequences for programs when children did not perform well on the assessment.

Concerns About How NRS Results Will Be Used	
	Percentage of Programs
Do not understand purpose	69
Results may be used to discredit Head Start	40
NRS growth scores may affect future funding	34
Individual scores may be shared with schools	9
N = 35 Head Start programs	

Uncertainty about the purpose of the NRS also appeared to fuel other concerns. Almost without exception, programs that did not understand the purpose of the NRS speculated that the true purpose was to harm, undermine, discredit, or dismantle Head Start. Among sampled programs, 40 percent expressed concerns that the NRS was a way to show that Head Start does not work. For example, staff in one program thought that there was a “hidden agenda” behind the NRS; in another, staff thought the purpose of the NRS was to test whether teachers were really doing their jobs. In yet another staff admitted they did not understand the purpose of the NRS, “but we can guess.” They made it clear they assumed a negative purpose. In 12 percent of programs, staff expressed concerns that the NRS frustrated children and made them feel a sense of failure. A few thought that the assessment intentionally “set up” children to fail, and one assumed that the design of the assessment was bad “on purpose” so that children would perform poorly, thereby making Head Start appear ineffective.

As a natural corollary of confusion and suspicion about the purpose of the NRS, staff in one-third of programs worried that the NRS could become a high-stakes test, with financial consequences for the programs. Staff worried that NRS results might at some point be used to inform future program funding decisions and some thought that programs without sufficient improvements in performance could be closed down. Staff in a few programs also wondered whether the NRS scores would be used as part of Head Start’s program monitoring visits that take place every three years, or if it was in fact currently a monitoring instrument. Questions about “What are they [the federal government] doing with the data?” came up frequently.

One less common, but troubling concern for 9 percent of programs was how results might negatively affect individual children. In two programs, staff wondered about the possibility that programs would inappropriately share results for individual children with schools, and that poor performance could “label” children as slow learners.<sup>1</sup> Another program wondered if the NRS was the start of a system to “track” or otherwise negatively affect the later academic careers of low performing children.

### Whether NRS Results Will Accurately Portray Program Performance

Perhaps in part because of their concerns about the consequences of the NRS results for their programs, local program staff raised a number of concerns about the validity of the NRS as a tool for measuring Head Start program performance. Many of the concerns staff raised have implications for validity of the NRS (measuring what each section is supposed to measure). In this chapter we make special note of issues with items that programs felt may affect validity *differentially* across subgroups of children. For most concerns about items in which inaccuracies are likely to be consistent across children, we mention them and reference the fuller discussion in Chapter II.

Concerns About Whether the NRS Will Accurately Portray Program Performance	
	Percentage of Programs
Vocabulary items contain cultural, regional, or socioeconomic bias	80
Standardized testing not developmentally appropriate for prekindergarteners	51
Validity of assessment/item development	37
Fall to spring comparisons not valid	17
Scripting too formal, stiff, or unnatural and instructions not appropriate	17
Potential for cheating	6
N = 35 Head Start programs	

Staff in more than half of sample programs expressed concern that a standardized test such as the NRS was not an appropriate way to assess preschool-age children. Rather, they felt observation over time would be a better choice for children in this age group. Primarily, staff concerns about the nature of the NRS were that preschool children were not accustomed to sitting for assessments or listening to what they believed to be complex instructions. Staff also noted difficulties redirecting children because of uncertainty about how much flexibility there was in the rules for them to stay on script (Chapters II and III present detailed discussions of staff experiences administering the NRS and their requests for additional training). Related to the issue of standardization procedures with young children, staff had concerns that the scripting of the instructions felt “stiff,” “unnatural,” and “robotic” and that this was confusing to children who were used to interacting with adults in

<sup>1</sup> NRS results are not reported to programs at an individual child level. However, some programs copy and retain individual children’s assessment answer sheets, making sharing theoretically possible, although obviously inappropriate.

---

a much warmer manner. Staff believed that children's performance would be negatively affected in such a setting.

In more than one-third of programs staff worried that the NRS was not valid either because implementation varied too much across sites, or that they did not have enough information about how items were developed to trust that comparisons were valid. Staff in a few sample programs voiced this concern because the assessor, setting, and quality of implementation could be very different across programs. Two programs suggested that assessors in other programs might falsify children's assessment data to improve the appearance of growth. One program remarked that assessors had varying levels of education and that some were "not very skilled as assessors." Others questioned validity because of unexpectedly low growth in areas that the program emphasized, and believed the NRS did not accurately measure children's knowledge in these areas. One tribal program believed that the NRS did not accurately reflect the way Indian children learn, specifically because they may be unwilling to point or express themselves during the assessment. Chapter II describes our findings about the rate of errors in administration of the NRS in this representative sample of programs.

Other concerns had to do with logistics and procedures. These included aspects of the assessment that staff felt confused children such as instructions to say "Good" after practice items only. Staff felt that this set children up to think they were doing poorly when they did not receive the same time of positive reinforcement on the scored items.

In some instances, staff noted that areas or items of concern had improved over time (especially the Early Math section). However, a large majority of sample programs (86 percent) reported some concerns with the receptive vocabulary (PPVT) section of the test. We characterize concerns about the receptive vocabulary portion of the NRS as (1) those that programs fear may differentially affect validity, and (2) those that programs fear may affect overall validity, though not differentially affect validity.

Program staff expressed concern that there may be bias in receptive vocabulary items (biases reported were against rural, urban, or children from low-income families). For example, staff in some programs indicated that rural children might not be familiar with "fountain." Others cited regional colloquial usage of words (for example, "diving" in some settings is also used to indicate jumping feet-first into water). In 17 percent of programs staff took issue with the word "tornado" stating that children who had experience with tornados may be more likely to select the picture of destruction caused by one, rather than the funnel. Other item level concerns are described in detail in Chapter II.

Common concerns voiced by program staff did not have implications for differential validity, but appeared to greatly affect staff attitudes toward the NRS. As described in Chapter II, many programs had concerns about the vocabulary section, mainly because they felt the section was too long, items too difficult or unfamiliar to children, the pictures poorly rendered, or the distracters on the plates unfairly similar. Staff in many sample programs said that it was difficult to keep children engaged through the vocabulary section, which they attributed to its length, and the fact that the pictures were not visually interesting or

engaging. Concerns ranged from words being too difficult for children of this age group or that the distracters on the plates were also valid answers.

Other sections of the NRS concerned staff as well, including the Letter Naming section and Early Math items. The primary concerns in the Letter Naming section were due to the layout of the plates and the method of administering them. As Chapter II details, examples of layout concerns included having too many letters per plate, presenting upper and lower case letters together, and placing boxes around letters. As far as administration, some programs complained that children skipped letters that assessors were certain they knew but simply missed on the plate.

The form or content of the Early Math section was the source of concern for close to half of programs (see Chapter II). Many programs felt that the concepts such as subtraction and reading graphs were too advanced for this age group and that not offering children manipulables made it very difficult for them to formulate correct answers. Some of the concerns had to do with questions about the overall validity of measures. In particular, the pie item (E16), a word problem involving subtraction had instructions that program staff felt were too long to hold children's attention, thus depending more on attention and language processing skills than mathematical ones. Others noted for this same item that children could get the answer correct without understanding the concept because the correct answer (4) is the same for the number of pieces missing as well as pieces remaining.

Local program staff also questioned the validity of comparing NRS results and calculating growth scores. In some cases, these concerns reflected lack of information about valid assessment practices, which may be alleviated with training or information. For example, staff in 17 percent of programs felt that the NRS could not validly measure growth from fall to spring since the assessment battery was changed between fall and spring administrations. Still others expressed concern about comparing the results of the English and Spanish assessments, since there are differences in the measures used. Others were concerned that the assessments were too difficult for children just entering the program in the fall. One program started earlier than most and was concerned that children were assessed in the fall after nearly three months of Head Start exposure, which did not provide a true baseline.

Finally, some programs were concerned that the assessment was too restricted and should include other domains to capture the entirety of children's development—discussed later in this chapter. Concerns about too narrow a focus appear to be similar over time. In spring 2006, 34 percent of programs reported this as a concern, compared to 30 percent in fall and spring 2005, and 60 percent in the Year 1 study.

### **Concerns Specific to the Spanish Assessment**

In programs that conducted Spanish assessments (24 of 35 sampled programs), staff expressed a number of concerns that fell into four general areas: (1) wording of the scripted instructions and the PreLAS, (2) content of the Vocabulary section, (3) content of the Letter Naming section, and (4) differences between the Spanish and English versions of the NRS.

---

In the discussion below we note where the concerns raised have implications for the validity of the assessment in Spanish. For concerns that relate to the overall assessment (and would affect all children in the same way) we mention them briefly and refer to the fuller description in Chapter II.

Concerns about wording of the script in the Spanish assessment were similar to those mentioned about the English assessment: some staff were concerned that the language is too formal and stilted, and therefore unfamiliar to children. In particular, some assessors believed that the use of particular words or phrases in the Simon Says instructions were not meaningful to children who spoke Spanish with different dialects. An example is that the scripted word for knock (as in “knock on the table”) is *golpea*, which means hit or punch. Children become confused because they are always instructed not to hit and the word is used out of context. According to assessors the equivalent phrasing of “knock on the door” would be “*toca la puerta*,” directly translated as “touch the door.” Another example is also from Simon Says: assessors said children did not understand the request to “put one hand on top of the other” [*pon una mano sobre la otra*]. Instead, assessors said the phrase should be “*pon una mano sobre de la otra*.”

Acceptable word alternatives and the wording of the instructions were also of concern to program staff. Staff noted that words that were part of the local Spanish dialect were not acceptable answers for some of the PreLAS expressive vocabulary items. An example is “*tenedor*” [fork], where in some dialects it would be “*cubierto*.”

As in the English assessment, the Vocabulary section of the Spanish assessment was the source of many concerns. Again, the quality of the drawings was an issue, and staff noted that the black-and-white line drawings were not only visually uninteresting to children, but that the renderings were also difficult to identify even when children knew the word. A common example was that the drawing of the skirt that staff felt looked like a lampshade. Others noted that the ambulance picture was of a very old-fashioned-looking vehicle that would not be familiar to children in this day and age. The word “*mechanico*” also came up as an overly difficult item because it is presented along with a gas station attendant as a distracter item; staff argued that this is a person that some children would have experienced as someone who also fixed cars.

Another example is the word “*juala*” [cage]. According to staff in Puerto Rico, this is a general categorical term for something that holds/contains animals, and the distracters on the page were all within that category: doghouse, birdhouse, and beehive.

Staff noted other concerns specific to the Spanish assessment. Some felt it was not fair to have different items on the Spanish and English versions of the test and then use them for comparison. In addition to differences in the vocabulary section, a few staff members noted that in the English version of the PreLAS children are asked “What can you do with it?” when shown a knife and a cup, and this question was not a part of the Spanish assessment. Concerns in the Letter Naming section were nearly unanimously about the inclusion of four Spanish letters (*ch*, *ll*, *ñ*, *rr*) and a few programs also commented that they did not understand why children had to name the letters in Spanish (for example, ‘*ah*’ as

opposed to “a”), when they are only taught to say letters in English in the classroom (see Chapter II).

### **Staff Time and Resources Dedicated to the NRS**

Staff in more than half of the sample programs expressed concern about the amount of time and financial resources dedicated to the NRS (see Box). Similar concerns had been raised in previous rounds of site visits, but the concerns appear to be somewhat diminished in spring 2006. During the spring site visits, staffs in 40 percent of programs talked about the burden that conducting NRS assessments places on their staff. In spring 2005, 60 percent of programs had this concern. Directors and lead trainers noted that time dedicated to the NRS took teachers away from their classrooms.

Local program staff also voiced concerns about financial resources dedicated to the NRS, both locally and nationally. At the local level, some programs needed to pay for substitute teachers to replace teacher-assessors during testing; another had to extend the contracts for college interns to help with assessments. One staff member characterized the NRS as an example of “increasing demands with decreasing funding,” because programs do not receive additional funds to cover these costs. In several programs, staff wondered how much money was being spent on the NRS at the national level, both for development of the assessment and the evaluation. Nearly one-third of programs specifically cited the costs of printing new assessment easels for each round when only a few items typically change; they felt that sending replacement pages would a better approach. In a few programs, the staff further expressed resentment about across the board funding reductions or lack of expansion in Head Start, while other expenditures (such as on easels and the NRS in general) appeared to be unaffected.

### **SUGGESTIONS FOR IMPROVING THE NRS**

During site visits, we asked respondents the following open-ended question: “Can you suggest ways to improve the NRS?” Based on their experiences conducting NRS assessments during the past three years, staff in the sample programs offered a broad range of suggestions for improvement that cluster in five main categories: (1) communication and planning, (2) training and guidance on administering the assessment, (3) adding domains to the assessment battery, (4) improving the Spanish-language version of the assessment, and (5) improving specific assessment procedures and items.

<b>Concerns About Staff Time and Resources</b>	
	Percentage of Programs
Staff time devoted to the NRS	40
Funds spent on the NRS	34
Funds spent printing easels	31
N = 35 Head Start programs	



## Communication and Planning

Consistent with the high proportion of programs having concerns about how the NRS results would be used at the national level, staff in 69 percent of the sample programs suggested that the Office of Head Start provide clarification about the purpose of the NRS and more specific information about how the results will be used (see Box). This finding is consistent with results from the spring 2005, fall 2004 visits and the Year 1 Quality Assurance Study; where comparable numbers of programs made this suggestion in each round of site visits. Some staff said that more specific information would help to assuage fears about why the assessments were being conducted and how their programs would be affected by the results. In addition, staff in a few programs asked for more guidance on how to use the results locally, in particular how they might combine it with local assessments. About one-quarter of programs had requested this guidance in spring 2005 and fall 2004 as well.

<b>Suggestions for Improving Communication and Planning</b>	
	Percentage of Programs
Provide more information on purpose/how results will be used	69
Provide information on development, validity, and reliability	29
Send training and assessment materials sooner	20
Send outcome reports sooner	17
N = 35 Head Start programs	

Late shipment of materials has been an ongoing challenge of NRS implementation. In spring 2006, for example, some programs received their training and assessment materials only weeks—and in some cases days—before their program year ended. In one program, Quality Assurance Study site visitors had to send easels to programs in advance of their visit so as to have assessments to observe. Not surprisingly, more than one-third of the sample programs requested that they receive their materials sooner, so that they would have more lead time to prepare to conduct each round of assessments. Planning for training activities was particularly affected by this late shipment of materials. Knowing when to expect materials to arrive would help program staff plan for completing developmental screenings, NRS assessments, and local assessments throughout the year. Knowing when the NRS assessments will begin and end would help staff integrate them with other program activities.

Slightly less than one-fifth of the sample programs requested that the Office of Head Start send outcome reports for each wave of assessments and the growth report on children's progress from fall to spring much sooner after each wave of assessments. Programs received the results from spring 2005 and the 2004–2005 Growth Report in winter 2006. Many staff members said they would like to receive growth reports before the start of the next program year so they could begin using the results for program improvement. Several wanted them even earlier so that they could work with individual children (however, NRS reports are not intended to measure individual children and data are not disaggregated to that level).

To address staff concerns about the validity of the assessment, 29 percent of programs requested more written information about how the assessment was developed and who was involved, as well as information about its validity and reliability. Several suggested that the instrument be reviewed by experts to determine the validity of each item. Staff at some of these programs thought that written materials about the validity of fall-spring comparisons would help them feel more confident about the outcome and growth reports. Staff in several programs also suggested that the Office of Head Start find ways to consult more closely with the broader Head Start community as it considers modifications and additions to the assessment. In particular, a number of staff members questioned whether early childhood educators had provided input to the initial development of the NRS.

Many programs asked for written materials they could use to describe the NRS and its purpose to parents and Policy Council members. In particular, staff in some programs felt unprepared to explain to parents the overall purpose of the assessment, why parents could not receive the test results for their child—since developmental screening and local assessment results are reviewed with parents—and why children who speak Spanish at home need to be assessed in both English and Spanish. Programs had made similar requests for these materials in spring 2005 and fall 2004.

### **Training and Guidance on Administering the Assessment**

The topic of training and guidance on administering the assessment came up in the context of providing more examples of neutral encouragement and coaching, ways to redirect children, when assessments can be stopped, and how to accommodate children with disabilities. Staff wanted to know how much they could deviate from the script, especially to use regional or colloquial words. Similarly, they wanted guidance on what constitutes neutral encouragement and coaching; one said that the scripted neutral encouragement “You’re doing a good job pointing!” was stilted and artificial-sounding. Some assessors suggested a “cheat sheet” of allowable phrases that could be placed beside the answer sheet as a reference. A few requested more information on how to redirect children who were acting out or otherwise not complying with the assessment. A few wanted more information on when it was appropriate to stop and restart an assessment. Staff in about 9 percent of programs said they needed clarification and more guidance on assessing children with disabilities—including the types of accommodations they are expected to make and how to determine whether or not they should attempt to assess a child with a disability. Several suggested adding a box to the scoring sheet to indicate a child has an IEP (however, confidentiality rules may prohibit such disclosures).<sup>2</sup>

As we note in Chapter III, 27 percent of sample programs reported including all recommended components in their spring 2005 refresher training (an increase from only 14 percent in spring 2005). Apart from those reported above, staff made relatively few references to training needs, and it may be that after three years of NRS implementation, experienced assessors may not feel a need for refresher training. It is difficult to discern

---

<sup>2</sup> The CBRS includes each child’s disability status, so that this information can be linked.

whether more intensive training occurred mainly for new assessors than for experienced ones. One program suggested that the Office of Head Start eliminate the need for training by maintaining a pool of people to conduct all the assessments. Another said that teachers doing the assessments should get a stipend for the additional work it entails.

### Expanding the Assessment Battery

As in spring 2005, staff in the sample programs were split on the question of whether new domains should be added to the NRS. Staff seemed to be torn between a desire to have their programs assessed on a broader set of outcomes (to better reflect the comprehensive nature of Head Start) and their concern that a longer assessment would be more challenging for both programs and children. Staff in 37 percent of programs said that new domains should not be added—either because they did not want the assessment to become longer, other domains were already covered by the local assessment, or they did not think that other domains (such as social-emotional development or creative arts) could be evaluated using a direct assessment such as the NRS (see Box).

Expanding the Assessment Battery	
	Percentage of Programs
Do not add new domains/ lengthen assessment	37
Add at least one domain	34
<ul style="list-style-type: none"> <li>• Social-emotional development</li> <li>• Gross/fine motor skills</li> <li>• Science</li> <li>• Creative arts</li> <li>• Mental Health</li> <li>• Domain not specified</li> </ul>	
N = 35 Head Start programs	

At the same time, staff in a similar proportion of programs (34 percent) suggested adding at least one new domain to the NRS. Half of the programs that wanted to add domains specified an assessment of social-emotional development should be included, although many expressed concern that this domain would be difficult to measure through the NRS. Other topics that staff suggested included gross and fine motor skills, science, creative arts, and colors.

### Improving the Spanish-Language Version

Programs made a number of recommendations about improving the Spanish language version of the assessment. Often, the issue of dialectical differences came up and staff suggested more flexibility to substitute words in the directions to account for the variety of dialects and differences in word usage among Spanish-speaking Head Start families and children. These families come from a variety of regions across the United States and Latin America, and word usage varies widely. Local program staff members have made this recommendation in every round of site visits made for the Quality Assurance Study. Beginning in fall 2004, the Office of Head Start allowed for discretionary use of some terms, such as replacing “senala” (to point) with “apunta” or “indica” if the assessor thinks this word is more regionally appropriate. Some assessors, however, believe that even more flexibility is needed; others may not be aware of the change.

Finally, staff in a few programs said that the items and pictures in the Spanish version should be the same as the English version. These staff members were concerned that because the items were somewhat different, the difficulty levels of the two assessments were not equivalent.

### **Improving Assessment Procedures**

Assessors and other program staff members made a number of specific suggestions for improving the assessment battery and procedures; many of these suggestions have also been made during previous rounds of site visits. To improve efficiency and conserve staff and financial resources, staff in some of the sample programs suggested combining the NRS and the local assessment. Staff in many programs had difficulty understanding why both assessments were necessary, although none expressed dissatisfaction with their local assessment. Local program staff made a broad range of suggestions to improve specific sections of the NRS assessment, noted above and described fully in Chapter II.

### **SUMMARY**

When asked to list the contributions the NRS had made to their programs, staff members in sample programs were able to attribute a number of positive effects of the NRS. Primarily, the contributions that staff listed were that it (1) helped to improve classroom practice and inform teacher training, (2) proved Head Start is effective, (3) raised program accountability by providing national comparisons, and (4) validated local assessment results. In a few programs, a contribution of the NRS was to raise staff morale. In two programs, staff indicated that a contribution of the NRS was that it helps to get children accustomed to a testing environment.

Local program staff raised several concerns about the NRS and its implications for future directions Head Start may take; most of these concerns also had been raised in previous rounds of site visits. The concern expressed most often by local staff members was about how the NRS results would be used at the national level. Many staff said that they still did not have a clear understanding of the purpose of the NRS and its implications for local programs where children did not perform well on the assessment. Local staff also expressed concern about whether the NRS results accurately reflect program performance, the amount of staff time and financial resources dedicated to the NRS, and whether it was valid to compare fall and spring assessments.

To improve NRS implementation, programs suggested that the Office of Head Start share more information about how the results would be used, send training materials and outcome reports to programs sooner, provide more information about the assessment's development and validity, and consult more with the Head Start community about future changes to the assessment. Many programs requested written materials about the NRS for parents.

Regarding the assessment battery, programs were split on whether new domains should be added. Many staff thought that at least one new domain—particularly social-emotional

development—should be added, but at the same time they did not want the assessment to become longer. (The social-emotional component of the NRS was implemented for the first time in fall 2006, after our site visits.) Program staff continued to recommend that the NRS be combined with the local assessments. They also suggested changes to improve the Spanish-language version and to modify specific assessment procedures and items.

**This page has been intentionally left blank for double-sided copying.**

## CHAPTER VII

### IMPLICATIONS FOR SYSTEM IMPROVEMENT

---

The value of the NRS for local program purposes is dependent upon how well local Head Start programs implement it, whether they accept it as a valid and reliable assessment, and whether they can use the information effectively for program improvement. Accuracy of administration, validity and reliability are also important to the other OHS goals for the NRS: to develop targeted technical assistance and to enhance the ability to report for federal accountability efforts. On the basis of findings from site visits to nationally representative samples of programs over the first three years of NRS implementation and our understanding of concerns raised by the Office of Head Start, expert consultants including Head Start program directors, and others in the field, we make the following suggestions for system improvement of the NRS. These suggestions are of several types: (1) ideas for increasing communication with local Head Start programs to improve knowledge and understanding of the purposes of the system, (2) ways to improve access to and usefulness of NRS results for local programs, (3) options for supporting programs in administering the assessment (including guidance on assessing children in Spanish and children with disabilities), and (4) ideas for changes to the assessment battery. Over the course of five rounds of visits to representative samples of Head Start programs, we have seen evidence that feedback on problems has led to improvements in the system. However, we have also found evidence that some challenges persist and deserve further attention.

#### COMMUNICATION

Several concerns raised by local program staff during three years of the Quality Assurance Study appear to be fueled by uncertainty and lack of information. For example, not knowing the specifics of how the NRS results will be used at the national level has led to speculation about whether it will become a “high-stakes” test, helping to make decisions on funding or employment at the program, center, or classroom level. Increased

communication about the Office of Head Start's plans could address some of these concerns. Suggestions for improving communication follow:

**Provide more information about how the Office of Head Start will use the NRS assessment results.** Since the NRS began in fall 2003, programs have requested more clarity of information about the purpose of the test and how the results will be used. More than two-thirds of programs were still expressing concerns about the purpose in spring 2006. This lack of understanding is linked to fears that data may be used, both at the national and local levels, to reduce funding for the national program or for their own program. Forty percent of programs feared that results would be used to discredit Head Start, and a third were concerned that growth scores could affect future funding. Efforts made by the Office of Head Start to convey the purpose of the NRS have not been sufficient to allay program fears.

**Provide a technical report or background materials about how the NRS assessment battery was developed and the validity and reliability of each task, including predictive validity for later school performance.** Validity of the test is another ongoing concern that programs continue to bring up in successive rounds. Some of this concern stems from program staff members' incomplete understanding of the principles and practices of standardized testing. However, in other cases, concerns highlight potentially problematic elements of the test that could continue to be improved.

Concerns about validity fall into two categories: first, program staff members have expressed concerns that directly administered, on-demand tests may not be valid for this population developmentally or for low income children generally. On the question of overall validity, data are now available to link the NRS with other research data, such as from kindergarten follow-up in the Head Start FACES study, in order to demonstrate whether in fact the NRS is predictive of later performance in school. Other validity concerns reflect potential differential item performance for groups of children. For example, staff members continue to raise concerns about items on the vocabulary test being biased against subgroups of children. Analyses of NRS test data could be done to investigate that issue, leading either to changes in items or reassurance that children's skills are not being unfairly measured by test items. Communicating the results of such analyses to Head Start program staff could go a long way toward reassuring programs of the value of their investment of time and resources.

**Explore options for increasing consultation with Head Start practitioners on future changes to the NRS.** In addition to questions about the validity and reliability of the assessment, local program staff often ask about who has been involved in its development. In particular, they ask whether early childhood educators like themselves have been consulted. During the spring 2006 visits, staff from several programs suggested that the Office of Head Start consult more closely with the practitioner community as it further considers modifications and additions to the assessment. For example, an advisory group of Head Start directors could be convened periodically to provide input to proposed changes to the assessment battery, training materials, report formats, or other NRS products. In fact, this Quality Assurance Study is itself an attempt to gather feedback from this program staff



---

members who are administering the NRS. It also provides an important source of information related to reliability of administration.

**Provide written materials about the NRS for parents.** During site visits, some staff reported feeling unsure about how to report NRS results to parents and Policy Councils or respond to their questions about the NRS. Parents and members of Policy Councils have requested clarification about why they cannot view their child's individual results, expressed concerns about bias of specific items, and raised the long-standing questions about the purpose and use of NRS results and how they could impact their program and Head Start nationally. An Office of Head Start brochure for parents could help local program staff provide parents with a clear and consistent message about the purpose of the NRS and help staff to respond appropriately to frequently asked questions. Programs made similar requests for guidance in the 2004–2005 rounds.

#### **ACCESS TO AND USEFULNESS OF ASSESSMENT RESULTS FOR LOCAL PROGRAMS**

Each round of data collection reflects more programs that are attempting to make use of the NRS results to improve their classroom practice, but they continue to express a need for guidance in how to accomplish this goal. As in spring 2005, program staff suggested several ways to make the reports more useful to them and requested help in interpreting the results and implementing appropriate changes to their programs:

**Send reports sooner after each round of data collection, preferably in time to use them in planning for the next program year.** Programs received reports on the spring 2005 assessment results and on children's progress from fall to spring (the 2004–2005 Growth Report) in January 2006. Some had hoped to use the spring 2005 results to plan for the 2005–2006 program year and about 20 percent reported that they were disappointed that they did not receive the results in time. Instead, programs developed technical assistance and staff training plans without the benefit of NRS growth results from the previous year.<sup>1</sup>

**Consider providing results disaggregated to lower levels and allow for easier comparisons with related national, regional and local data.** Programs varied in the level of reporting they preferred, but the majority wanted results at least at the center or classroom level. More than a quarter suggested providing results at the individual child level. Programs also mentioned having comparisons available directly on their baseline or growth reports, rather than requiring that they look up comparisons online from the CBRS website. They requested the ability to make specific comparisons, such as with fellow delegates within a grantee, or with programs in neighboring counties. In addition, programs requested reporting formats that would allow them to incorporate their program's outcome data into presentations and reports to stakeholders. All programs had shared their results beyond the directors and lead trainers responsible for NRS oversight. Nearly half of programs reported

---

<sup>1</sup> The Office of Head Start has subsequently been able to move up the time of reporting, so that programs receive interim reports without national comparisons within a month of submitting their scoring forms, and final reports with national comparisons within a few months.

using the results to target their staff development efforts. Slightly more than 10 percent of programs collected the data at an individual level and used the results to identify skill levels needing attention, although the Office of Head Start does not recommend this practice. Clarifying appropriate uses of data would be helpful to program staff and further support their efforts in collecting the information.

**Provide programs with more guidance on how to use the NRS results for program improvement, including building more linkages with local assessment results.** The degree to which programs found the NRS growth reports useful and relevant for program improvement efforts is an important indicator of how likely programs are to use the reports as envisioned—as one of several tools to inform local decisions on strategies for improving program quality and effectiveness. If programs see the results as providing a new source of information that is accessible and clear, they are more likely to value the time and effort that go into the collection of the information. More than in previous years, programs saw the reports as user-friendly.

This round of site visits yielded more-specific information about local assessment tools: more than half the programs used Creative Curriculum’s assessment, and another 20 percent used the High/Scope COR—in both cases, because these aligned with their programs’ primary curriculum. More than two-thirds of programs expressed satisfaction with their local assessments, especially the ability to track child outcomes over time and to get results quickly. Another prized feature was the availability of results across subgroups of interest. However, nearly 30 percent of programs were not entirely satisfied with their local assessment.

Programs that made comparisons between NRS and local assessment results generally found the results aligned, although some programs refused to compare, citing different formats and emphases. Fourteen percent of programs reported gaining new information about ELLs from the NRS that they had not had previously. Some programs valued the ability to compare their results with those of similar programs or with national averages. Now that multiple years of data are available, a few programs expressed an interest in tracking trends over time.

**Provide access to resources to support local program improvement efforts undertaken in response to assessment results.** The overwhelming majority of programs reported making changes to classroom practices in response to NRS results. Nearly half increased their emphasis on letter naming, and more than a quarter purchased classroom supplies in order to bolster instruction in literacy or math. Some of these purchases seemed tied to specific items on the tests, such as globes, rulers, and graphing supplies.

In each site visit round, programs have requested more guidance and resources—such as training, technical assistance, recommended curricula, and access to consultants—for making improvements in their programs if the NRS results indicated a need to do so. As in past rounds, some staff said that if they are to be held accountable for improving children’s performance on the NRS, they need guidance and help in making changes in their programs that could produce the desired results. One of the stated goals of the NRS is to target

---

provision of technical assistance, and programs are suggesting clearer linkages of their NRS results and availability of related resources.

### **SUPPORT FOR ADMINISTERING THE ASSESSMENT**

Timely shipment of training and assessment materials and modest changes in the NRS training protocols have the potential to help staff feel more confident and conduct the assessments more smoothly and accurately. Specifically, recommended changes include the following:

**Ensure that programs receive materials and scheduling information on time.**

During every round of site visits, late receipt of training and assessment materials was mentioned as a significant problem for many programs. Problems with timely delivery continued to plague programs in spring 2006: nearly half experienced delays in receipt, and thus were delayed or blocked from having adequate training sessions. As the NRS becomes a routine part of Head Start program activities, local programs need to incorporate plans for training assessors, conducting the assessments, and reviewing the results into their local plans and schedules. Program staff would appreciate receiving schedules with key NRS dates and materials well in advance so that they can schedule training and assessments at convenient times and coordinate the NRS assessments with other staff training, assessments, and program activities that need to be completed each fall and spring. As in past rounds, staff reactions to spring 2006 training materials and information were mixed. At least one staff member at most programs had watched one of the past NRS broadcasts, or the newest Webcast, but opinions of their usefulness varied.

**Ensure that adequate refresher training is carried out to reduce errors in administration.** As in previous rounds, there was considerable variation in program approaches to conducting refresher training on the English-language assessment. Adherence to training protocols was higher in spring 2006 than in spring 2005. However, some programs did not comply with suggested methodologies, either because materials did not arrive in time, assessors were considered experienced and knowledgeable, and/or assessors lacked time for training. Six out of 35 programs were required to train and certify new assessors in spring 2006, although the training conducted was shorter than recommended by the Office of Head Start, and in rare cases, new assessors were not formally certified.

For the first time, we conducted analyses that demonstrated a link between offering assessor trainings and assessor performance on the assessment. Assessors in the two programs that failed to offer refresher training were much more likely to make administration errors, such as gesturing, non-neutral encouragement, and coaching. Training differences did not affect scoring. Assessors from those programs also received lower certification scores on the English assessment—another reason why recertification might be useful to correct administration problems before assessment begins (see recommendation below). Although we are unable to determine that the link with training is causal, it is certainly worth considering as an issue.

In summary, lack of thorough new assessor and refresher training in many programs, coupled with a tendency not to seek technical assistance, may have contributed to some of the errors in administration and scoring we noted on our visits.

**Consider requiring recertification of assessors during refresher training.** Programs might be more likely to conduct comprehensive training if recertification was required by the Office of Head Start for each assessment round. One way to ensure that assessors are prepared to implement changes in assessment procedures and can administer new items reliably is to recertify them at the end of refresher training. The current practice is to certify assessors only when they are initially trained. While this step would build in additional quality control procedures at the program level, it must be weighed carefully against programs' ongoing concerns about staff burden.

**Provide more guidance on how to interact appropriately with children during the assessment.** Assessments went more smoothly in spring than in the fall round, as children had gained confidence about skills, familiarity with the assessment process, and comfort working with adults. About half of the programs had some children with challenging behaviors, such as bored/distracted, overly chatty, nervous, or anxious about their performance. A small number of programs still requested changes in administration, primarily to be able to use a warmer tone with children; these programs appear not to understand the concept of neutral encouragement and were afraid to stray at all from the script to use encouraging prompts. Program staff who are not the children's teachers are concerned about how best to encourage and redirect children they do not know. Because some staff members are still unsure about what they can and cannot say to children during the assessment, additional guidance on how to provide appropriate, neutral encouragement and the types of statements that constitute coaching would be helpful.

## **GUIDANCE ON ASSESSING CHILDREN IN SPANISH**

The NRS is intended to assess the skills of children with English and/or Spanish as their home language(s) or of children with other home languages who can pass the English language screener. Twenty-four out of 35 programs needed to assess children in Spanish. Spanish assessment certification has improved since spring 2005. The mean certification score was 97, and 96 percent of assessments were completed by assessors who had a certification score over 85. Among Spanish assessors, the most frequent errors were in scoring or mispronunciation or insertion of articles. At the same time, assessors and trainers continued to have some concerns about administering the Spanish version of the NRS and expectations for growth in Spanish-language skills. Some possible approaches to addressing these concerns include the following:

**Allow for more regional differences in Spanish terms used in the assessment.** Staff continued to express concerns about the appropriateness of the language in the Spanish assessment for Spanish speakers from different regions. Although the NRS does allow for some discretionary use of regional terms, assessors need to be made aware of this flexibility and perhaps be provided even more options. Given the emphasis on sticking with the script, specific guidelines on appropriate substitutions may be needed. Incorporating

these acceptable terms into the easel materials would be most helpful and would help to standardize acceptable regional responses.

**Clarify expectations for growth in Spanish skills.** Some assessors questioned why children are asked to respond in Spanish when English is the language of instruction in their programs (except in Puerto Rico). This becomes particularly relevant at the spring assessment, when most Spanish-speaking children are likely to pass the English screener and complete the English-language assessment. Further explanation of how growth in skills will be measured for Spanish-speaking children would be helpful because some staff believed it was unfair for children to be assessed in Spanish when they are instructed in English. Some children assessed in Spanish in the fall could no longer pass the Spanish screener in the spring, suggesting that instruction in English might have eclipsed their Spanish language skills.

#### **GUIDANCE ON ASSESSING CHILDREN WITH DISABILITIES**

All but two programs in the sample assessed some children with identified disabilities during spring 2006. The majority of disabilities were speech and language delays. Thirty-one out of 294 English assessments and 6 out of 70 Spanish assessments were of children with an IEP. Although the majority of programs felt comfortable with the accommodations they made, some requested more guidance on how to assess children with disabilities. More guidance would be helpful in the following areas:

**Clarify when a child should not be assessed because of a disability.** The CBRS allows programs to indicate that a child was not assessed because of a “severe disability” or because an IEP prohibits it. However, 13 percent of programs were not certain about what constituted a severe disability. Although the Assessor’s Guide provides some guidance about appropriate accommodations, not all assessors were aware of these materials; some sought more specific information on whether to assess autistic children, for example, or children who were nonverbal or had speech impairments. In about a third of programs, staff opted to complete assessments with children who may have been eligible for an exemption under NRS guidelines, resulting in difficult assessments. While many programs felt that they had received adequate training and guidance, staff from 13 programs complained about the inadequacy of information about assessing children with disabilities. They requested more information in their local training materials, as opposed to needing to seek technical assistance on the helpline.

**Provide more examples of appropriate accommodations.** Many program staff mentioned accommodations they used to make sure children could be assessed fairly, such as splitting the assessment into two parts, speaking slowly, and ensuring that the testing setting had no distractions. However, as in past rounds, a small number of programs made accommodations that are not consistent with NRS protocols. For example, one program used a sign-language interpreter for a hearing-impaired child, and another covered letters on the alphabet plate to assist a visually-impaired child. In one program, assessors simplified the testing process for children with developmental delays, changing the script or pointing at letters when the child could not point. Programs need additional guidance of accommodations that conform to NRS rules of administration.

## CHANGES TO THE ASSESSMENT BATTERY

Some parts of the assessment continued to be areas of concern to staff or to be associated with relatively high rates of error. In light of these concerns, it might be appropriate to review the assessment in the following areas:

**Consider amending the Language Screener and Vocabulary tasks and carrying out analyses to allay concerns about bias.** For PreLAS Simon Says, there were a few suggestions about changing the instruction to “put one hand on top of the other” to saying “put one hand on top of the other hand” and also specifically prompting the child to put the paper down if he or she does not immediately do so. In Art Show, staff continue to assert that the colored pictures of children at the bottom of pages are distracting. On the Vocabulary task, 15 programs raised the issue of cultural appropriateness and regional bias of items on the test. Staff in nine programs had a sense that items were confusing or intended to trick children, or were too difficult for these ages. Analyses of NRS data to clarify whether in fact there is any bias against specific subgroups of children on this test would help to allay these concerns.

**Consider amending the pie, graphing, and counting questions in the Early Math section.** For the pie item, programs suggested that children were counting the number of pieces remaining in the pie rather than subtracting, and for the graphing question, that children were confused because assessors were no longer allowed to point to the teddy bear. The counting task instruction “Ready? Go” still caused children to race through the task, and staff suggested substituting “Start now.”

**Consider alternative approaches to the Letter Naming task.** As in past rounds, nearly half of the programs in spring 2006 found the current format and directions for the Letter Naming task to be time-consuming and frustrating for the children. Difficulties in administration ranged from layout (too many letters, mix of upper and lower-case letters) to the method of asking the question. Staff suggested asking the child to name letters that the assessor pointed to, so as to ensure that children could focus and not accidentally skip letters. Some children were likely to give the sound rather than name of letters, and to expect to find a sequence of letters (given A-E on the first panel). Difficulties in the score sheet layout were also mentioned, since it does not mirror the panel layout.

## SUMMARY

Based on our visits to a representative sample of Head Start programs, we suggest implications for system improvement in two major areas: (1) helping program staff better understand the purposes, interpretation, and potential uses of the NRS through improved communication and guidance (including a special focus on children with disabilities and children assessed in Spanish) and (2) making improvements in the training, administration guidelines, format, and content of the battery itself. Through the Quality Assurance Study, MPR will continue to work with the Office of Head Start, expert consultants including Head Start program staff, and the implementation contractors to foster improvements to the entire assessment system.

## REFERENCES

---

- Administration on Children, Youth and Families. “The Head Start Leader’s Guide to Positive Child Outcomes.” Washington, DC: U.S. Department of Health and Human Services, 2003.
- Chromy, J.R. “Sequential Sample Selection Methods.” Proceedings of the Survey Research Methods Section of the American Statistical Association, 1979, pp. 401–406.
- Dunn, L.M., and L.M. Dunn. *Peabody Picture Vocabulary Test—Third Edition*. Circles Pines, MN: American Guidance Services, 1997.
- Duncan, S., and E. DeAvila. *PreLAS 2000 Technical Report*. Monterey, CA: CTB/McGraw Hill, 1998.
- Horn, Wade F. “Improving Head Start: A Common Cause.” *Head Start Bulletin*, no. 76, 2003. Available at [[www.headstartinfo.org/publications/hsbulletin76/hsb76\\_02.htm](http://www.headstartinfo.org/publications/hsbulletin76/hsb76_02.htm)].
- Paulsell, Diane, Linda Rosenberg, Renée Nogales, Charles Nagatoshi, Susan Sprachman, Louisa Tarullo, and John Love. “Meeting the Challenge: How Head Start Programs Implemented the National Reporting System.” Princeton, NJ: Mathematica Policy Research, Inc., December 2004.
- Paulsell, Diane, Anne Gordon, Renée Nogales, Susan Sprachman, Cheri Vogel, Barbara Schiff, and Louisa Tarullo. “Implementation of the Head Start National Reporting System: Fall 2004 Update.” Princeton, NJ: Mathematica Policy Research, Inc., January 2005.
- Paulsell, Diane, Anne Gordon, Renée Nogales, Patricia del Grosso, Susan Sprachman, and Louisa Tarullo. “Implementation of the Head Start National Reporting System: Spring 2005 Update.” Princeton, NJ: Mathematica Policy Research, Inc., January 2006.

Scientific Software Development. *Atlas.ti: Visual Qualitative Data Analysis Management Model Building in Education Research and Business*. Berlin: Scientific Software Development, 1997.

Zill, Nicholas. *Early Math Skills Test*. Rockville, MD: Westat, Inc., 2003a.

Zill, Nicholas. *Letter Naming Task*. Rockville, MD: Westat, Inc. 2003b.