Contract No.: 233-02-0086; HHSP233200700001T
MPR Reference No.: 6337-500

MATHEMATICA
Policy Research, Inc.

Implementation of the Head Start National Reporting System: Spring 2007

Final Report

December 2008

Louisa B. Tarullo
Cheri A. Vogel
Nikki Aikens
Emily Sama Martin
Renée Nogales
Patricia Del Grosso

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.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Louisa B. Tarullo

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

ACKNOWLEDGMENTS

We 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 40 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. Barbara Schiff Kennen, Renée Nogales, and Susan Sprachman led the training of our talented group of observers/site visitors. The site visitors also included Brian Roff, David Eden, Betsy Santos, Charlotte Cabili, Emily Sama Martin, Nikki Aikens, Louisa Tarullo, Cheri Vogel, Tim Silman, and Liz Condon. Barbara Carlson, our sampling statistician, assisted by Yuhong Zheng, randomly selected programs, centers and children for site visits and assessment observation. Emily Sama Martin and Charlotte Cabili contacted all programs to ensure their participation. Xiaofan Sun did the programming to calculate scoring accuracy. Patricia Del Grosso and Emily Sama Martin coded the site visit summaries, and Tim Bruursema managed the project finances. Alfreda Holmes formatted the final document. Laura Bernstein was our editor. Diane Paulsell, who originally conceptualized this implementation study and led its previous reports, provided valuable insights during the review process. Kim Boller and Sally Atkins-Burnett served as our quality assurance reviewers. 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, Jatzin Alvarado, and Gloria Gonzalez, 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**

Chapter **Page**

Chapter **Page**

Chapter **Page**

VI (*continued*)

Chapter **Page**

TABLES

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

FIGURES

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.[1] 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).

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. Starting in fall 2006, a new Social-Emotional Development (SED) rating form was distributed to programs for completion by teachers in fall and spring of each program year. The Office of Head Start has provided each program with a summary report of average results for all children in the program who were assessed, available within about a month after each fall and spring administration and in final form, with national averages included, several months

[1] The Improving Head Start Act, signed in December, 2007, discontinued the NRS. This report focuses on NRS implementation in the spring of 2007, and analyses were completed prior to the discontinuation of the NRS.

after the final completion date of administration. Reference tables were 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), and second and third year reports were submitted in subsequent years (Paulsell et al. 2005, 2006; Vogel et al. 2008). ACF again contracted with MPR and J&A to continue the study with the same two project components.[2]

This report documents Head Start programs' experiences with the NRS during the fourth year of implementation. The report is based on information collected through visits to a nationally representative sample of 40 Head Start programs in spring 2007 with an oversample of American Indian/Alaska Native (AI/AN) programs. During these visits, staff observed the assessments of a random sample of children and interviewed program staff members about their experiences implementing the NRS. This round differed from the previous round in three ways. First, in an effort to gain more information about uses of assessment information in AI/AN programs, we added questions to interview and focus group protocols that focused specifically on perspectives on assessment from AI/AN programs. Second, since a new aspect of the NRS, the Social-Emotional Development (SED) rating form, had been added to the NRS in spring 2006, we added questions to the protocols to learn more about this rating form that was completed by children's teachers in the fall and spring of the Head Start year. Finally, because previous reports had shown a decline in time devoted to training in programs, we also directly observed training sessions in a selected set of these programs.

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

[2] 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.

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.[3] 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 was not new to Head Start, a national assessment system implemented consistently for all 4- and 5-year-olds was a significantly new and different approach. The NRS, initiated in April 2002 when the Bush administration announced the *Good Start, Grow Smart* early childhood initiative, was a key element of the "Strengthening Head Start" component of this initiative. It created 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 which 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 discuss 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.

[3] Information Memorandum No. 18, issued 8/10/2000, on Child Outcomes, Performance Measures, and Program Self-Assessment

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 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. 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 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.

Starting in fall 2006, a new Social-Emotional Development (SED) rating form was distributed to programs for completion by teachers in fall and spring of each program year. The inclusion of this new element had been recommended by program staff and the TWG. Chapter VII contains detailed information on the new rating form.

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 could now generate assessment completion reports at the program or center level, perform data searches, and view and operate the CBRS in Spanish. They could 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 could also enter or make changes to multiple children at one time. In addition, grantees could 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 first two years of the Quality Assurance Study demonstrated that, while there was room for improvement, Head Start staff were able to administer the NRS assessment with a fairly good degree of fidelity to the protocol (Paulsell et al. 2004; Paulsell et al. 2005, 2006). Moreover, analyses of site visit interview data and assessment observations yielded a number of helpful suggestions for improving the assessment battery and process. Results from the Year 3 study showed that error rates remained very low overall, but that they had increased over time. Moreover, there was an increase in programs that did not follow ACF guidelines for refresher trainings. This decrease in training fidelity is important, since programs that did not follow all of the recommendations for refresher training had higher error rates than programs that did follow them (Vogel et al. 2008). The Office of Head Start decided to continue the study into a fourth year to collect data on the level of fidelity of administration and ongoing information about programs' experiences and perspectives on NRS implementation. For the reasons cited above, we also observed the trainings held at five programs in spring 2007.

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 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. In spring 2007, the primary focus of this report, we visited a nationally representative sample of 40 programs with an oversample of American Indian/Alaska Native (AI/AN) programs that was designed to capture their unique perspectives and experiences regarding child assessment. During these visits, staff observed the assessments of a random sample of children and interviewed program staff members about their experiences implementing the NRS. This spring, we added questions that focused on perspectives on assessment from AI/AN programs. We also added questions to the protocols to learn more about a new aspect of the NRS, the Social-Emotional Development (SED) rating form that was completed by children's teachers in the fall and spring of the Head Start year to report on children's behavior and social skills.

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. Systems development support includes recommendations based on the QA study findings, as well as a set of activities aimed at improving the implementation and use of assessments in Head Start programs.

The purpose of this report is to provide updated findings from the Quality Assurance Study—focusing primarily on NRS implementation in spring 2007—and make recommendations for system improvement.

QUALITY OF ASSESSMENT ADMINISTRATION

As in earlier rounds of the NRS Quality Assurance (QA) Study, most of the 312 English language assessments observed in spring 2007 met or exceeded the standard of quality used in certifying assessors. The average certification score was 91, and 75 percent of observed assessments exceeded the certification standard of 85 points. The inter-rater reliability of assessment scale scores remained high.[4] Further, quality of the Spanish-language assessments observed (90 in 12 programs) was also relatively high, on average (89 points).

For the English-language assessments, scoring errors along with all types of administration errors due to coaching, inappropriate gestures, non-neutral encouragement, straying from the script, and inserting articles such as "a" and "the" were higher in spring 2007 than in spring 2006; no error categories were less frequent than the previous spring. Likewise, errors increased from spring 2006 to spring 2007 on the Spanish-language assessment, with the exception of "other" errors (inserting articles or mispronouncing words during the TVIP section).

Although quality was high overall, some areas of the child assessment were difficult for staff. Modifications to the assessment easel or additional guidance and training might be helpful in certain components of administering the assessment. The majority of these areas were also problematic in previous years of the NRS, including:

- Coaching, particularly in the Simon Says and Tío Simón sections
- Gesture errors on certain Art Show items, (especially B5 to B10), Early Math items (especially E14, E18, and E19), and Conocimientos Básicos de Matemáticas items (especially EE16 to EE19)
- Scoring the counting item correctly in the Early Math (E20) and Conocimientos Básicos de Matemáticas (EE20) sections

[4] We examined the reliability of total scores for each section of the NRS by calculating an intraclass correlation coefficient (ICC) for each scale. 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 well into the .90s for all subscales except counting, which was .88. Findings are similar to previous rounds of the NRS Quality Assurance Study.

Coaching and gesturing errors were somewhat more prominent as compared with spring 2006. Scoring errors on the counting item, however, remained essentially unchanged—it was scored incorrectly on 25 percent of observed English-language assessments (bilingual assessors–the term used to describe those staff administering the assessments in Spanish–have scored this item incorrectly on approximately 17 percent of assessments in spring 2006 and 2007). 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.

Considering children's reactions to the assessment, seventy percent of sample programs reported that most children had reacted positively. In spring 2006, this had been true for about 50 percent of programs. As in previous years, assessors reported that some children enjoyed the one-on-one time with staff and the opportunity to play a game, along with the chance to demonstrate what they had learned during the year. Assessors reported that children's behavior was much less of a concern in the spring than it had been in the fall, a finding similar to previous rounds of the NRS QA Study. The most common behavioral issue was that children became bored or restless during the PPVT and/or the Letter Naming task and needed a lot of redirection. To address these challenges, assessors attempted the assessment at a later time or split the assessment into multiple sessions, took a quick stretch or bathroom break, or offered the children extra neutral encouragement or small incentives.

Most programs (32 out of 40, or 80 percent) administered the child assessment to English language learners. Spanish was the most common other language spoken, followed by Chinese and Arabic. Staff reported that children usually could pass the English language screener in the spring, even if they had not passed in the fall. However, assessors often lacked understanding of how to determine the best language in which to assess a Spanish-speaking child, if their view was in conflict with information provided by parents. As has been the case in previous rounds of site visits, some bilingual assessors were critical of the Spanish version of the NRS because they felt it was unfair that children who use colloquial words from different regional dialects are not given credit on the NRS when these words are logical responses but are not responses accepted as correct on the assessment.

Nearly all programs (37 out of 40 in the sample) that assessed children with disabilities reported using a wide range of accommodations for the assessment. In most cases, programs successfully administered the NRS to children with special needs. These programs either did not need to make any accommodations, or else used modest ones, such as allowing the child extra time to respond, taking breaks, or repeating questions. Most assessors felt well prepared to administer the NRS to children with disabilities and making necessary accommodations. However, a few programs would like the training materials to explicitly address the issues of conducting the assessment to children with special needs and making appropriate accommodations for these situations.

LOCAL APPROACHES TO NRS TRAINING

Comprehensive training that NRS assessors perceive as valuable may increase staff buy-in to the assessment, which, in turn, could contribute to the quality of collected data. To

promote excellence in assessor training, OHS provides annually updated materials for each program to use when training its assessors.

At the local level, nearly all (93 percent) programs provided some form of refresher training to staff in spring 2007. Approximately 60 percent of programs provided comprehensive spring refresher training that incorporated the training video, the Assessor's Guide, role play, and a question and answer session. Another 28 percent of programs provided a training that incorporated three of these four elements, usually leaving out the role-play activity. Only one quarter of programs reported that their lead trainer did follow-up observations of some staff after the training. Nine out of ten programs that had new assessors in spring 2007 provided training to them. At 28 percent of the programs, staff felt that training twice per year was too much and suggested reducing the frequency to once annually. Staff in the three programs that did not train staff were significantly more likely to make administration errors than programs where staff had been trained. Similarly, at the 23 programs where staff had a comprehensive training, staff were significantly less likely to make administration and scoring errors than at the 16 programs where the training had been less than comprehensive.

In spring 2007, 70 percent of programs assessed children in Spanish. A third of these offered comprehensive Spanish-language assessment training to staff, and 42 percent either encouraged self-review of the Spanish assessment materials or offered no additional training on Spanish assessment materials. Although these training figures appear to indicate a lower level of preparation for Spanish-language assessment staff, many of these staff attended a refresher training on the English assessment, or were at sites where a comprehensive Spanish-language refresher training was impossible to conduct with only a single Spanish-language assessor.

MPR observed the training of 170 assessors at five programs in spring 2007. An experienced and certified lead trainer led all observed training sessions, and the trainings included a video and distribution of the Assessor's Guide. Training at four of five programs also incorporated the role-play activity. The length of refresher trainings ranged from 50 minutes to about 4.5 hours. We observed four areas in which the staff had difficulties conducting the local NRS trainings, including (1) ineffective role-play activities; (2) providing misinformation; (3) not referring to the Assessor's Guide as needed; and (4) not providing bilingual trainings. However, we also observed several particularly effective elements of training, such as a review of proper procedures for completing the answer sheet and techniques to set children at ease.

In interviews and focus groups with program staff, we asked for reactions to and feedback on the NRS training materials that OHS provides. Seventy percent of the programs visited in spring 2007 reported that someone had watched a webcast or satellite broadcast in the past. However, fewer than half of these (28 percent of all programs) reported that someone on his or her staff had either watched the spring 2007 webcast or read the transcript. Programs this year reported fewer problems with materials than in past rounds, and 12.5 percent of programs described an improvement in the timely and accurate delivery of assessment materials. Most program staff reported favorable impressions of the

2007 training video and Assessor's Guide, and of their experiences with requesting and receiving technical assistance.

LOCAL APPROACHES TO IMPLEMENTATION

Overall, the 40 Head Start programs we visited in spring 2007 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 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 all but one program, the lead trainer was also responsible for overseeing implementation of the SED rating form.

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 13 assessors. Consistent with findings from previous years, in spring 2007, 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. 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 assessing children.

Nearly all programs communicated with parents, Policy Councils, tribal leaders/elders, and other stakeholders 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. Similar to programs visited in spring 2006, three-quarters of programs sought to obtain parents' written consent for the NRS assessment.[5] 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. These included parent requests to see their child's individual results, concerns about bias and appropriateness of specific items and the language of the assessment, and questions about the purpose of the NRS and how the results would be used. Other stakeholders also had questions about how the results would be used to improve program performance.

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

[5] We are not aware of a requirement that programs obtain specific parental consent for administration of the NRS assessment, beyond general parental approval for program assessment and screening procedures.

USING THE SOCIAL-EMOTIONAL DEVELOPMENT RATING FORM

Like the NRS cognitive child assessment, if the SED rating form is to be useful at the local level, local Head Start programs must be able to implement it with fidelity to the model, accept it as a valid and reliable instrument, and understand how it can lead to local program improvements. Unlike the NRS cognitive child assessment, ACF does not require special training or certification for staff to fill out the SED rating form that is designed to capture information about individual child behavior and social skills. Among the programs visited in spring 2007, many lead trainers felt the rating form and accompanying materials were straightforward and required little training. As a result, most programs did not provide a formal training for staff, instead distributing rating forms and the instruction sheet to raters. Most programs reported that the rating form took little to no preparation and was easy for staff to complete. Only SED raters in one program and lead trainers in two programs expressed a desire for formalized training or more instruction from ACF. While most programs felt the forms were relatively easy to fill out and not burdensome, staff often stressed the time constraints at the start and end of the program year when the rating forms need to be completed.

As with the NRS cognitive assessments, program directors assigned a lead trainer the responsibility of overseeing implementation, including scheduling and tracking the completion of rating forms and submitting rating forms by the deadline set by the Office of Head Start. In all but one program, the lead NRS trainer was also responsible for overseeing implementation of the SED rating form. In selecting raters, a primary goal was to have the rating forms completed by someone who knows the child well. Among the 40 Head Start programs we visited in spring 2007, all programs used classroom teachers to complete the rating forms for children enrolled in the center-based option; programs used home visitors, home-based teachers, and classroom teachers to complete the forms for home-based children. Programs adopted these staffing approaches because they felt assigned staff were most familiar with the children, they wanted to be responsive to instructions provided by ACF, or selected staff were also responsible for completing other local assessments and screenings. More than three-quarters of programs used at least two staff members to complete forms for individual children, with teachers often consulting teaching assistants or working collaboratively with them (or other staff) to fill out forms.

More than half of programs communicated with parents, Policy Councils, tribal leaders/elders, and other stakeholders about the SED rating form in some respect. Remaining programs indicated that they did not discuss or inform these groups about the SED separately from the NRS cognitive assessment. While the means of communication varied, programs made an effort to inform parents at the start of the program year. None of the programs sought to obtain parents' written consent for the SED rating form alone; instead, the majority of programs considered the rating form as being covered by a blanket consent form for all assessments or a separate consent for the NRS cognitive assessment. No programs had any parent refusals, and parents (and stakeholders) rarely raised concerns.

Most programs did not offer concrete plans for using the SED data in the future. However, more than two-thirds of programs felt that the SED could make a contribution at the child, program, or national level. Some programs thought that the SED rating form

complemented local assessments and screenings, and some also expressed appreciation that a social-emotional development component has been incorporated into the NRS. Moreover, some programs expected the rating form to inform local training, technical assistance, and program improvement efforts. Many programs felt that the items included in the rating form were comprehensive and well targeted.

On the other hand, local program staff from 65 percent of programs raised some concerns about the SED rating form. Twenty percent of programs said that they should receive SED results in a more timely fashion and at the child level in order to use them for program planning and improvement. In addition, some programs felt that guidance from ACF on how to address problem areas would be helpful. Local staff also expressed concern about whether the rating form imposed a burden on staff at the beginning and end of the program year; duplicated local assessment efforts; and contained items that were perceived by some staff as being too negative, subjective, or difficult to quantify based on the rating categories. Staff in a few programs also expressed concerns about the rating form for children in tribal programs, ELLs, and children with disabilities.

USING THE NRS FOR LOCAL PROGRAM IMPROVEMENT EFFORTS

The majority of Head Start staff who participated in the site visits found the 2005-2006 Growth Report easy to understand. Program 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). A majority of the 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. Some programs reported purchasing a new curriculum. Others mentioned enhancing parent involvement to support classroom activities. Because the NRS reports did not provide center-, classroom- or child-level information, 37 percent of programs reported tracking the item responses of individual children (compared with 11 percent in 2006); this practice was more likely in programs which used teachers as assessors than those which did not. Seventeen percent of programs mentioned using NRS data to make comparisons with national data or with data on comparable programs (compared with 14 percent in 2006), and 57 percent of programs reported making comparisons of their data across multiple years (compared with just over half of programs in 2006).

With regard to using the NRS results in the future, more than half of the 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 intended to continue incorporating NRS outcomes as one source of information for program-planning efforts.

PERSPECTIVES OF LOCAL HEAD START STAFF ON THE NATIONAL REPORTING SYSTEM

Program staff cited a number of contributions the NRS has made to their programs, notably the national comparison it provides, an indication that programs have begun to appreciate the primary aim of the NRS and to use the data as intended. Staff also reported that the NRS contributed to their program by raising program accountability, improving classroom practice and teacher training, measuring child knowledge, proving Head Start is effective, and raising staff awareness about assessments and OHS expectations. In all, staff at 85 percent of programs reported that the NRS made some contribution.

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. In 35 percent of programs, 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. It is notable that concerns about the purpose of the NRS were about half as prevalent in spring 2007 than in spring 2006.

To improve NRS implementation, programs suggested that OHS share more information about how the results would be used, send training materials and outcome reports to programs more promptly, provide additional information about the assessment's development and validity, and consult more with the Head Start community about future changes to the assessment. Nearly 20 percent of programs requested materials about the NRS for parents and other program stakeholders.

Regarding the assessment battery, programs were split on whether new domains should be added. Most thought that the assessment needed no additional domains, but approximately one third suggested adding at least one more. As in previous rounds of NRS site visits, program staff recommended 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.

In our experience talking with staff at selected tribal programs, no concerns, contributions, or suggestions were identified by staff as being unique to tribal programs. At two programs staff appeared to take exception to the idea that tribal Head Start programs would be any different from non-tribal programs. Staff in one program mentioned a concern about language of administration and local culture, but maintained that these concerns could also apply to non-tribal programs.

IMPLICATIONS OF FINDINGS FROM THE QUALITY ASSURANCE STUDY

Based on our visits to a representative sample of Head Start programs in Spring 2007, we have suggested 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) helping programs understand how to use and interpret data that they collect from regularly administered local assessments. Given that, as of the time this report is being finalized, the NRS has ended, we continue to underscore the importance of planfully collecting, interpreting, and using locally collected data on children's performance to improve practice. Through the ongoing Learning from Assessment task in the Quality Assurance Study, MPR continues to work with the Office of Head Start and expert consultants including Head Start program staff, to develop materials that will help programs use their own locally collected data more effectively.

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.[6] 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.

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 provided each program with a summary report of average results for all children in the program who were assessed, available several months after each fall and spring administration. Reference tables were also created to allow programs to compare their scores to national averages, regional averages, and 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 in the Administration on Children and Families (ACYF/ACF) contracted with Mathematica Policy Research, Inc. (MPR), and its subcontractor Juárez & 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

[6] The Improving Head Start Act, signed in December, 2007, discontinued the NRS. This report focuses on NRS implementation in the spring of 2007, and analyses were completed prior to the discontinuation of the NRS.

of Head Start in December 2004 (Paulsell et al. 2004), and second and third year reports were submitted in subsequent years (Paulsell et al. 2005, 2006; Vogel et al. 2008). ACF again contracted with MPR and J&A to continue the study with the same two project components.[7]

This report documents Head Start programs' experiences with the NRS during the fourth year of implementation. The report is based on information collected in spring 2007 through visits to a nationally representative sample of 40 Head Start programs. 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 2007. We also observed the refresher trainings held in five programs. 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 2007 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.[8] 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 (ACYF 2003). The child assessments required of all programs to 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

[7] 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), ACF, U.S. Department of Health and Human Services.

[8] Information Memorandum No. 18, issued 8/10/2000, on Child Outcomes, Performance Measures, and Program Self-Assessment

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 was not new to Head Start, a national assessment system implemented consistently for all 4- and 5-year-olds was a significantly new and different approach. The NRS, initiated in April 2002 when the Bush administration announced the Good Start, Grow Smart early childhood initiative, was a key element of the Strengthening Head Start component of this initiative. It created 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 then 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 Office of Head Start contracted with Westat, Inc., and its subcontractor Xtria, LLC in August 2002, which 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 discuss 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 of 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 subsets 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 used for 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 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. 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. Children who could not pass the English screener and whose home language was other than Spanish were not assessed.

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 instructions for pointing to 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, facilitate the identification of eligible children, and 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 to each child 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 program staff could generate assessment completion reports at the program or center level, perform data searches, and view and operate the CBRS in Spanish. They could 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 staff could also enter or make changes to data about multiple children at one time. In addition, grantees could produce the 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 first two years of the 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; Paulsell et al. 2005, 2006). Moreover, analyses of site visit interview data and assessment observations yielded a number of helpful suggestions for improving the assessment battery and process. Results from the Year 3 study showed that error rates were quite low overall, but that they had increased over time. Moreover, there was an increase in programs that did not follow ACF guidelines for refresher trainings. This decreased fidelity to the training protocol is noteworthy, since programs that did not follow all of the recommendations for refresher training had higher error rates than programs that did follow them (Vogel et al. 2006). For these reasons, we also observed the refresher trainings held at five programs in spring 2007. The Office of Head Start decided to continue the study into a fourth year to collect data on the level of fidelity of administration and ongoing information about programs' experiences and perspectives on NRS implementation.

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 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. In spring 2007 we visited a nationally representative sample of 40 programs with an oversample of American Indian/Alaska Native (AI/AN) programs in order to explore their unique perspectives and approaches to child assessment. During these visits, staff observed the assessments of a random sample of children and interviewed program staff members about their experiences implementing the NRS. For the first time in spring 2007, we also studied the Social-Emotional Development rating form that had been newly implemented in spring 2006.
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 program staff 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 came 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 40 Head Start agencies conducted in spring 2007. 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 40 Head Start agencies, with 5 AI/AN programs and 35 non-AI/AN programs, and a random and representative sample of about 10 children per program 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 40 Head Start programs in spring 2007, we used the most current (2005–2006) Head Start Program Information Report (PIR) as the basis for a sampling frame.[9] 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.[10] For the primary sampling unit, we used 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 the data analysis computer package 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. We explicitly stratified (sorted) by whether a program was AI/AN (ACF regional office 11) or not, and agencies were selected within stratum with equal probability. We selected 35 programs within the non-AI/AN stratum and 5 within the AI/AN stratum. We implicitly stratified the frame of Head Start agencies by Census region. Within a region, we sorted by racial and ethnic distribution, metropolitan status, and enrollment size.[11]

In preparation for the site visits conducted in spring 2007, we initially selected twice the number of programs that we were targeting (80 rather than 40), then formed sequential pairs of programs within stratum. 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. Five programs in the sample were replaced by the agency paired with them, so a total of 45 programs were ultimately released into the sample. We ended up with only 39 participating programs for child assessment observations because one program could not be visited in person. However, staff from the program were interviewed by phone, allowing us to include their responses in the program level analyses. Reasons for nonparticipation of the originally

[9] 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.

[10] 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 directly serving no 4- and 5-year olds, according to the 2004–2005 PIR. At ACF's request, we also excluded programs in the Migrant/Seasonal farm worker program (ACF regional office 12) and programs under the transitional management of CDI. Earlier rounds of the Quality Assurance Study focused specifically on migrant program issues.

[11] 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 less 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 2005–2006, and had two categories: (1) 200 or fewer, and (2) more than 200.

sampled programs included ineligibility and refusal to participate. (See Table I.1 for the characteristics of sample Head Start agencies.)

Centers Within Programs. Many of the Head Start programs provided services at two or more centers. Given time and resource constraints for conducting these 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. From that, we 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 (on average). In the five largest agencies (based on the 2005–2006 PIR), we aimed for 15 assessments, and in the 5 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.

The Sample of Head Start Programs and Children Observed

The sample of 40 Head Start programs that participated in the spring 2007 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 368, ranging

Table I.1. Characteristics of the Sample of Head Start Agencies, Spring 2007

Characteristics	Number of Agencies
ACF Region	
Region I	0
Region II	7
Region III	5
Region IV	6
Region V	7
Region VI	4
Region VII	1
Region VIII	3
Region IX	2
Region X	0
Region XI (Tribal Programs)	5
Total	40
Head Start Grantee Status	
Grantee that Operates Programs Directly and Does Not Have Delegate Agencies	27
Grantee that Operates Programs Directly and Has Delegate Agencies	3
Delegate Agency	10
Total	40
Head Start Program Option	
Center-Based Only	29
Combination of Center-Based and Home-Based	11
Total	40
Type of Agency	
Community Action Agency	9
School System	7
Public or Private Nonprofit	17
Tribal Government or Consortium	2
Government Agency	5
Total	40
Metropolitan Status of Agency's Service Area	
Metropolitan Area	23
Non-Metropolitan Area	16
Puerto Rico	1
Total	40
Number of Head Start Centers the Agency Operates	
1 Center	9
2 to 5 Centers	12
6 to 10 Centers	9
11 to 20 Centers	5
21 or More Centers	5
Total	40
Number of Enrolled 4- and 5-Year-Olds	
1 to 50 Children	5
51 to 100 Children	9
101 to 200 Children	9
201 to 300 Children	6
301 to 400 Children	1
401 to 500 Children	1
501 to 600 Children	1
601 to 700 Children	2
More than 700 Children	6
Total	40

Source: Head Start Program Information Report (PIR), program year 2005-2006.

Note: 40 programs were sampled for the spring 2007 study, and child observations were completed at 39 of those (though staff interviews were completed at all 40)

from 16 to 3429 across the sample programs. All selected programs provided primarily center-based services; the number of centers operated by the programs ranged from one to 73. Out of the 9 programs that also offered a home-based option to some families, funded home-based slots ranged from 8 to 354 across programs.

In spring 2007, our primary sample of 405 children included 55 children with an identified disability and 82 children whose home language was Spanish. When children from the primary sample could not be observed, we released 403 additional children from the replacement sample, bringing the total number of children in the sample to 808. Of the total sample of 808, 107 were children with an identified disability, and 160 were from primarily Spanish-speaking homes. Of the 808, we completed 405 observations of child assessments; 303 of these were from the primary sample, and 102 were from the replacement sample.[12] Nineteen children from the total sample were ineligible to participate (not going to kindergarten the following year) and 128 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; see Table I.2). This number is close to the number in the last round in spring 2006 (135). 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 2007 included 52 with an identified disability and 86 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).

The Child Assessment Observation Form is based on the Child Assessment Certification Form developed by Westat for certifying NRS assessors.[13] For each item in the NRS child assessment (including the warm-up section and 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 Peabody Picture Vocabulary

[12] At 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.

[13] 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.

Table I.2. Reasons Why Assessment Observations Were Not Completed in Spring 2007

Reason	Primary Sample	Replacement Sample	Total
Ineligible for the NRS			
Family Dropped Out of the Program	18	1	19
Eligible Non-Completes			
Child absent	49	26	75
Child assessed prior to site visit	11	8	19
Scheduling problem or insufficient time to complete the assessment observation during the site visit	5	10	15
Child could not be assessed due to behavior problem	8	3	11
Spanish assessor not available during the site visit	5	2	7
Parent refused permission for the assessment	1	0	1
Total eligible non-completes	79	49	128
Grand total	97	50	147

(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
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 2006 results
12. Concerns and suggestions for improving the NRS

The focus group guide for staff who conducted 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 identified and organized 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.

Starting in spring 2007, we added questions to the protocols to learn more about a new aspect of the NRS, the Social-Emotional Development (SED) rating form that was completed by children's teachers in the fall and spring of the Head Start year. We also added questions that focused on approaches to local assessment and general perspectives on assessment from American Indian/Alaska Native programs.

Qualifications and Training of Site Visitors

A team of 16 site visitors conducted the spring 2007 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 assessors to conduct child assessments, interviews and focus groups with program staff, and provide quality assurance and assess program implementation. Five of the site visitors were fluent in Spanish. Many staff members who conducted spring 2007 site visits had also conducted visits in spring 2005 or 2006 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 2007, prior to the spring 2007 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 met the criteria and were judged to be reliable at the 85 percent level.

Site Visit Activities

Spring 2007 site visits were conducted between April 16 and June 6. 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, MPR conducted a quality control review of all observation forms and narrative reports as soon as possible after each visit. We then carried out data entry of 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 the 39 programs with child assessment observations 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.[14]

[14] To construct the weights for programs, we first calculated the probability of selection of each program. 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 pairs 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 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 African American or Hispanic population, and metropolitan status, and used the inverse of

(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 (see 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 40 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.[15] Throughout the report, we note instances in which different staff members within the sample programs disagreed about a particular topic, although this would be more likely to happen in independent interviews than within focus groups.

ROADMAP TO THE REPORT

We now turn to describing the results of our assessment of the NRS implementation in spring 2007. In Chapter II, we describe the quality of the cognitive 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 assessments, including coordinating and staffing the assessments and communicating with parents. Chapter IV also reports what program staff told us about their costs for NRS implementation. In Chapter V, we describe ways in which program staff have used the NRS cognitive child assessment results for program improvement, how they plan to use their results along with the results of local assessments, future plans for using the results, and their reactions to the reports on outcomes for their program. 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, concerns about the NRS, and suggestions for improvement. In Chapter VII, we describe the administration of the Social Emotional Development (SED) rating form, including training and projected use of the data. Based on findings from the spring 2007 site visits, in Chapter VIII we synthesize recommendations for improving program assessment systems.

(*continued*)
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.

[15] 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 Spring 2007 Site Visits, by Category

Protocol Source
Grantee Director (or, if delegate, Head Start Director)
Lead NRS Trainer
Assessor/Social Emotional Development (SED) Rater Focus Group
Site Visit Summary
Overall Impression of Assessments
Program Characteristics
Center-Based Only
Some Home-Based
Conducts Spanish Assessments
Tribal Program
Implementing the NRS Assessment
Scope (Assessments, Centers, Number of ELLs)
Assigning Staff to Administer the Assessment
Assigning Spanish Assessors
Assigning SED Raters
Conducting the Cognitive Assessment (Scheduling, Timing, Locations)
Completing the SED Rating Form
Assessors' Experiences Conducting the Assessments
SED Raters' Experiences Completing the Rating Form
Children's Responses to the Assessment Process
Difficulties Encountered and Strategies Used
Assessing Children with Disabilities
Assessing English-Language Learner Children
Experience of Spanish-Language Assessors
Experience of Teacher-Assessors
Communication with Parents and Other Stakeholders about Cognitive Assessment
Communication with Parents and Other Stakeholders about SED Rating Form
Costs of NRS Implementation
Implementation Errors
Scheduling Assessments
Conducting Assessor Follow-Up Observations
Tracking Progress in Completing Assessments
Update on Program Operations
Training
Approach to Training for Cognitive Assessment
Approach to Training for SED Rating Form
Adequacy of Refresher Training
National Training Materials and Guidance
Suggestions for Refresher Training and Other Training Issues
Suggestions for National Training Materials
Lessons and Implications
Reactions to the SED Rating Form
Changes to Classroom/Curriculum
Concerns About the NRS
Contributions of the NRS
Suggestions for Improving the Cognitive Assessment
Suggestions for Improving the SED Rating Form
Suggestions Unique to Tribal Programs
Suggestions for Other Aspects
Local Assessment Process
Reactions to the NRS Outcomes Reports
Reactions to SED Report
Using Cognitive Assessment Data
Using SED Data
Tribal Programs' Perspectives on Using Results
Future Plans to Use Data
Other Issues Raised

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

CHAPTER II

ADMINISTERING THE CHILD ASSESSMENT

Each year in the NRS Quality Assurance study, we evaluate the implementation of the NRS child assessment by assessing the reliability of administration, the prevalence of errors in administration, and staff perceptions of the NRS overall. Understanding the overall quality of assessment administration is an important first step for improving the assessment system. Then, identifying patterns of errors in administration and scoring is also important, since some errors may indicate a need for more training and guidance. Although assessor errors are generally few, learning more about their patterns is illuminating. Also, items that are difficult to administer may need to be modified or removed from the assessment battery. After previous data collection rounds, information collected during site visits led to changes in such areas as adding acceptable alternative answers, simplifying gestures and clarifying instructions, and modifying questions administered in the battery. The focus on errors in this chapter is not intended to be critical of assessor or program performance, but to offer constructive support for improvement of all aspects of the system of child assessment activities.

In this chapter, we will describe the overall quality of assessment administration in spring 2007 and make some comparisons to results from earlier rounds. As in previous reports for the NRS Quality Assurance Study, we based our analyses on the extent to which the administration of the assessments we observed met certification standards, measures of inter-rater reliability, and the types and frequency of administration and scoring errors. Similar to the spring 2006 round of assessments, we present weighted, rather than unweighted, percentage estimates of error data, unless otherwise noted (a discussion of the weights is presented in Chapter 1). We will examine these issues for our observations of both the English- and Spanish-language versions of the assessment. However, it is important to note that our observation sample for the Spanish assessment is just 98 observations (90 of which completed all five sections), and thus should not be used to generalize about Spanish assessments nationally. Finally, we will describe the experiences of

assessors, NRS lead trainers, and Head Start program directors based on our interviews and focus groups.

We conducted tests of statistical significance for the changes from spring 2006 to spring 2007 for the following key measures: total errors of each major type (script errors, non-neutral encouragement, coaching, incorrect gestures, pronunciation errors, and scoring errors) along with the mean certification score.

APPROACH TO EVALUATING THE QUALITY OF THE NRS ASSESSMENTS

During our site visits, we examined the quality of assessment administration by conducting structured observations of a sample of assessments—421 assessments administered by more than 121 different assessors across 40 sampled programs.[16] During these observations, site visitors[17] 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, one of the federal contractors that developed NRS materials, for the purpose of certifying local Head Start staff to administer the assessment. The form was designed to document three types of 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. These scores were compared to each other to identify scoring errors; we treated our site visitors as though they were trainers and accepted their scores as correct.

2. **Administration Errors:** The Assessment Observation Form was designed to record the frequency of five types of administration errors, by item: (1) straying from the assessment script, (2) coaching, (3) non-neutral encouragement, (4) gesturing errors, and (5) other errors, such as pronouncing words incorrectly and/or inserting articles ("a" and "the") in the vocabulary section (see Box this page, in which the last two types of errors are grouped as "Other Errors"). 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 errors in pre-assessment activities such as not entering the child's 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

[16] The full sample for the spring 2007 NRS Quality Assurance Study was 40 programs. Data from one monolingual Spanish program in Puerto Rico, however, is only included in the discussion of Spanish certification scores later in this chapter.

[17] Site visitor were trained by MPR researchers who had been certified and deemed reliable assessors during the initial regional Training-of-Trainers conference in summer 2003.

preparing materials before the child was present, and not having materials ready (i.e., the sheet of paper required for two Simon Says items). It also included other procedural errors such as not providing appropriate prompts if the child stopped identifying letters in the Letter Naming task. Frequencies of these procedural errors are presented in text boxes later in this chapter.

Types of Administration Errors

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

Coaching: Encouraging a 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, saying "That's right!," "You are so smart!," or "That's wrong."

Gesturing Errors: Making incorrect hand gestures, such as pointing instead of circling, circling incorrectly, forgetting to point, or unscripted pointing.

Other Errors: Pronouncing words incorrectly or inserting articles ("a" and "the") on the vocabulary section; and failing to slow the child down when necessary to record answers.

We also conducted focus groups with assessors—including those we observed and some we did not—to learn about their experiences administering the NRS child assessments. In addition, we interviewed Head Start directors and staff who trained assessors (see 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 2007 refresher training and summarizes the content of six local NRS trainings observed by site visitors, provides further insight into possible sources of errors.

The next three sections describe the quality of the sample of 323 English assessments observed.[18,19] 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 procedures, administration, and scoring errors observed for each segment of the assessment. After describing assessors' performance on the English assessments, in the fourth section of this chapter we report similar information for the Spanish-language version for the 98

[18] The primary focus of this report is the quality of the observed NRS assessments in spring 2007. Nevertheless, we make some comparisons between assessment quality in fall 2003 and 2004 and spring 2004, 2005, 2006, and 2007 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] and Vogel et al. 2008).

[19] As discussed in Chapter I, we present weighted estimates of assessment errors in this chapter. Appendix C provides weighted estimate tables of question-by-question assessment errors for both English and Spanish assessments.

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 (see Box) on the Child Assessment Certification Form. Errors in scoring are determined by comparing trainer and trainee scores; trainers' scores are accepted as the correct scores. After the observation, total administration and scoring errors are summed by type. The number of errors for the five error categories (straying from the script, coaching, non-neutral encouragement, other administration errors, and scoring errors) is then calculated by assigning a rating of 1 to 5 to each of the five error categories depending on the total number of errors in that category (see Box on this page). Ratings in each section are multiplied by four and summed to produce a certification score between 20 and 100. To become certified, trainees must obtain a score 85 or higher.[20] That is, an assessor can make as many as two of each type of administration error in a given assessment and still receive a score of 100. The total number of potential errors on the certification forms is quite high – 90 potential scoring errors and 224 administration or procedural errors; however, certification requires no more than 19 errors in a given assessment.

Certification Ratings

5: 0–2 errors

4: 3–5 errors

3: 6–8 errors

2: 9–11 errors

1: 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 2007, programs did a reasonably good job of administering the assessments correctly and consistently. Across the 312 English assessments with full data, the mean certification score was 91, indicating that, on average, administration of the assessments exceeded the certification standard by six points (91 – 85 = 6).[21] Furthermore, 75 percent of observed assessments achieved a certification score of 85 or above (see Table II.1 on next page). Of the 25 percent below a

[20] Functionally, because scores are each divisible by 4, this means assessors must obtain a combination of either all 5s, three 4s and two 5s, or one 3 and four 5s.

[21] By "full data," we mean that we have data on the entire NRS assessment. All or parts of the scoring data were missing for 11 assessments, so certification scores and average scoring errors could not be calculated. In these cases, the child did not pass the language screener, and thus only scores for Sections A (Simon Says) and B (Art Show) were available. In results presented below, we use the slightly smaller sample to examine the certification score and the level of scoring errors but use the larger sample of 323 to examine other types of errors.

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

Certification Score	Number of Assessments	Percentage of Assessments	Cumulative Percentage of Assessments
56	2	0.1	0.1
60	1	0.0	0.1
64	1	0.5	0.6
68	6	1.7	2.3
72	8	3.1	5.4
76	21	5.9	11.3
80	27	5.7	17.0
84	**34**	**8.2**	**25.2**
88	**36**	**12.2**	**37.4**
92	**50**	**15.9**	**53.4**
96	**59**	**19.3**	**72.7**
100	**67**	**27.3**	**100.0**
Total	**312**	**100**	

Source: Spring 2007 observations of English NRS child assessments.

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

N = 312 (11 children did not pass the language screener). Percentages are weighted estimates.

passing score, 8 percent achieved a score of 84, only one point below the passing score.[22] Thus, about 83 percent of all observed English assessments either met the standard or achieved a score one point below the standard. Approximately half (46 percent) fell in the 96-to-100 range. Overall mean certification scores for the spring 2007 assessments were about the same as certification scores on assessments observed in spring 2006, but the percentage meeting the certification standard in 2006 was 83 (versus 75 in Spring 2007) with an additional 5 percent within one point.[23]

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 cost issues.

[22] 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 4 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.

[23] The difference in the mean certification score between spring 2006 and spring 2007 is not statistically significant.

The full NRS assessment (sections A to E) was designed to be administered in 15 minutes. During spring 2007 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 15.3 minutes. Eighty-six percent of assessments were between 10 and 20 minutes, with some outliers due to special circumstances. These durations were similar to previous rounds of site visits; 15.3 minutes in spring 2005 and 15 minutes in spring 2006. Full assessments conducted by classroom teachers lasted 2 minutes longer than those conducted by other staff (the difference in minutes, 16.8 versus 14.8, is statistically significant, $p < .01$).[24]

Intraclass Correlation Coefficients for NRS Scales

Language Screener	0.96
Vocabulary	0.94
Letter Naming	0.95
Early Math Skills	0.97
Counting (Item E20)	0.88

N = 323 English assessment observations

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 (see Box on this page).[25] 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 well into the .90s for all subscales except counting, which was .88. Findings are similar to previous rounds of the NRS Quality Assurance Study.

ERRORS IN PROCEDURES, ADMINISTRATION, AND SCORING

During the assessment observations, site visitors coded errors made in administration for each item in the battery, including assessment items, introductory sections preceding each segment, and practice items. Approximately eight percent of assessors whom we observed had administered the NRS assessments for the first time in spring 2007 (compared to seven percent in 2006). As with previous rounds of the NRS, the overall error rate on English assessments was low in spring 2007. On average, assessors made just 2.5 errors per assessment. Only straying from script errors increased significantly since spring 2006, although we noted a pattern of small nonsignificant increases across all other types of errors.

[24] 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 their likelihood of committing a coaching error.

[25] 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.

As compared to the initial round of assessments in fall 2003, only scoring errors are lower on average; instances of straying from the script, coaching, giving non-neutral encouragement, and using incorrect gestures have increased.

For this round of the NRS Quality Assurance Study, we examined the differences in mean error rates between spring 2007 and 2006. The mean number of scoring errors per assessment was 2.5, which is about the same as spring 2006 (2.2). 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 Table II.2).

Of all types of errors, script errors increased the most between 2006 and 2007. Straying-from-script errors rose from 2.1 errors per assessment in spring 2006 to 4.8 errors in spring 2007, a difference that is statistically significant ($p < .01$). Similarly, coaching errors, instances of non-neutral encouragement, and gesturing errors also increased over time, although in small and not practically meaningful ways (all increases but non-neutral encouragement were statistically significant $p < .10$ to .01). Gesturing errors increased the second most, despite changes made in the gestures and the instructions to simplify this aspect of assessment administration.[26] 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.

Table II.2. Mean Administration and Scoring Errors for Observed English Assessments

Type of Error	Fall 2003	Spring 2004	Fall 2004	Spring 2005	Spring 2006	Spring 2007
Straying from the script	3.1	1.5	1.3	1.1	2.1	4.8
Coaching	0.6	0.6	1.5	0.8	1.1	1.5
Non-neutral encouragement	0.4	0.6	0.5	0.3	0.9	1.0
Incorrect hand gestures	1.9	2.6	2.7	1.9	3.1	3.7
Inserting articles	0.1	0.1	0.1	<0.01	0.1	0.3
Scoring errors	5.6	2.3	2.9	2.1	2.2	2.5
Total Errors (all types)	13.1	9.8	12.1	8.0	12.7	13.8
N =	346	300	286	316	294	312

Note: Sample sizes for the scoring errors are slightly lower in some years than those listed in the table because these they 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), 293 (spring 2006), and 312 (spring 2007). All except fall 2003 are weighted estimates.

[26] 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").

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 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 low. However, in order to provide information that can be useful for improving training and materials in the future, we focus our discussion in the rest of this section on those few errors observed.

Administering the Set-Up and Warm-Up Sections

The most common set-up errors observed were failure to: (1) fill out the score sheet cover page before the assessment started, (2) set up the assessment in a quiet area, (3) fill out the date correctly, and (4) fill out child's ID correctly (see Box on this page). (Examples of assessment environments that may have interfered with children's abilities to respond correctly that program staff described during interviews and focus groups is discussed further in Chapter IV.) Other set-up errors, such as not having materials prepared in advance or making the testing environment as comfortable as possible for the child, occurred far less often. Although we do not have data on the reason why it happened, it is worth noting that the incidence of incorrectly recording a child's ID increased nine percentage points (from 3 to 12).

Percentage of Assessments with Observed Set-Up Errors

Did not fill out score sheet cover before getting started	14
Did not set up assessment in a quiet area	12
Did not fill out the date correctly	12
Did not fill out child's ID correctly	12
Did not place easel correctly or placed in an awkward spot	5
Area was not set up with materials before child arrived	3
Arrangement or size of table and chair was awkward or uncomfortable for child	2
Did not have paper ready for Simon Says	2
Did not acknowledge nervousness of child	1

N = 323 English assessment observations; weighted estimates

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, assessors strayed from the warm-up script in 33 percent of assessments by misreading the script, omitting certain parts or choosing their own words to introduce the assessment to the child (Appendix Table C.1). This represents a sizeable increase since spring 2006, when assessors committed script errors on eight percent of assessments.

Administering and Scoring the Pre-LAS Simon Says

The Pre-LAS Simon Says task, based on the well-known children's game, serves along with the Pre-LAS Art Show as a language screener to determine whether the child has

adequate English proficiency to be assessed in English.[27] The task consists of an introduction, two practice items, and 10 assessment items; it is the same instrument used since the Year One assessments.

Errors in coaching, scoring, and straying from the script were most common in this section. Items A3 and A10 require that assessors look under the table to see if the child understands the commands "lift one foot" and "put your feet together." These tasks require careful timing by the assessor to avoid hinting that the child should do something with his or her feet, since assessors need to look under the table to see if the child performed the item correctly. Program staff members know that they should pause before looking under the table, since their movement could give a hint to the child on the appropriate task, but in about one-fifth of observed assessments, they looked before the child had responded. For the command, "pick up the paper" (A5), 14 percent of assessors coached by either looking at the piece of paper or pushing or touching the paper (Appendix Table C.1). Most script errors occurred in the two introductory statements (24 and 18 percent, respectively compared to 8 and 7 percent in 2006), perhaps as assessors tried to make the child feel comfortable with the upcoming activities. Scoring errors on item A9, "Simon says point to the middle of the paper," reached 7 percent, but no other item had errors more common than 4 percent (not shown). 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 they 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. Despite the higher numbers of scoring errors for A9 than for other items, these errors were actually less common in spring 2007 than in spring 2006 (11 percent).

Percentage of Assessments with Procedural Errors on Simon Says

Looked below table early or did not look below table for item A3, "Simon Says lift one foot" and item A10, "Simon Says put your feet together"	21
Hinted by looking at or moving paper	14
Did not look up for first practice item	6
Child confused by paper and assessor did not give appropriate guidance	4
Incorrectly looked down for second practice item	4
Did not look for hands if hidden from view	2

N = 323 English assessment observations; weighted estimates

Site visitors also noted several procedural errors not captured by the certification error codes (see Box on this page). 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 six

[27] In the Spanish version, a comparable task (Tió Simón Dice) serves as part of a language screener to determine whether the child should be assessed in Spanish.

percent of assessments and incorrectly modeled the second item in four percent. Since the error rate for the second item was 20 percent in spring 2005, this four percent rate is far lower than in the past, although it is similar to the rate in spring 2006. 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 over time.

Administering and Scoring the Pre-LAS Art Show

The Pre-LAS Art Show task, the second part of the language screener, consists of two practice and 10 assessment items. On the English version, seven items require the child to name objects, such as a bee, a frog, and a pig. Three items require the child to tell the function of objects—a book, a cup, and a knife.[28] Like previous assessment rounds, gesturing errors emerged as the most common mistake made by program staff (Appendix Table C.1), the majority of which clustered around the object-function items (B5, B7, B9). For these items, assessors incorrectly gestured by pointing to rather than circling the book, circling rather than pointing to the cup (19 and 15 percent, respectively), or failing to run their finger along the knife (13 percent).[29] Moreover, between 29 and 34 percent of the assessors failed to point to the pictures a second time while asking, "What can you do with it?" The prevalence of these errors closely mirrored the rates observed in spring 2006.

Percentage of Assessments with Procedural Errors on the Language Screener

Did not complete language screener section before continuing with the assessment	20
Did not fill in correct response to indicate assessment path	5
Did not choose correct assessment path	2

N = 323 English assessment observations; weighted estimates

On the other hand, assessors strayed from the script more often than in spring 2006 for a handful of items. Site visitors observed about the same percentage of script errors in the introduction in 2007 as in 2006 (8 and 9 percent, respectively), a portion of the script when assessors may take liberty to deviate to help prepare the child for the next activity. However, items that required the assessor to ask, "What can you do with it?" (B6, B8, and B10) yielded script errors in 17 percent, 12 percent, and 7 percent of assessments, respectively. This was a larger number of errors than in 2006. It is unclear what could account for the increase in

[28] On the Spanish version, the Art Show (Exposicíon de Arte) contains no items about the function of objects; all items require children to identify the objects only.

[29] Precise gestures for items B5 and B9 help ensure that children focus on the correct part of the picture in answering the question. For example, if an assessor points to the book instead of circling it, the child may think that the assessor wants him or her to identify the picture of the cat in the book. However, a child's answer is not likely to be influenced by whether the assessor points to or circles the cup (B7). Still, if an assessor circled the cup, we counted that as a gesturing error according to the directions in the easel and Assessor's Guide.

script errors, especially since these questions are short and straightforward. Scoring errors for the Pre-LAS Art Show were relatively low and comparable to the previous spring.

At the conclusion of the Art Show section, assessors sum 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.[30] Assessors did not complete the part of the form showing the results of the language screener before moving on to the vocabulary task for 20 percent of assessments (see Box on previous page). This is a 200 percent increase in this type of procedural error from spring 2006 when assessors did not complete this part of the form on 4 percent of assessments. However, despite this high error rate, only 2 percent of assessors in spring 2007 did not choose the correct path on the screener by either failing to continue or stopping the assessment appropriately.

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

The vocabulary task includes an introduction, 4 practice and 24 assessment items, each of which includes a plate with four pictures. The child is instructed to point to a quadrant on each plate, which illustrates both objects 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 2007. As was the case in all previous rounds, assessors generally made few administration errors in the PPVT section, with one notable exception—straying from the script errors increased during the introduction and practice items (Appendix Table C.1). Assessors strayed from the introductory script on 19 percent of observed assessments (compared to 9 percent in spring 2006) and 0-12 percent of assessments across the practice items (compared to 0-3 percent in spring 2006). On two practice items, two percent of assessors incorrectly inserted an article (for example, "point to *the* ball" instead of "point to ball"), but, like previous spring rounds, these kinds of errors rarely occurred during the scored (non-practice) portion of the PPVT. In spring 2007, the mean number of script errors for the PPVT section was 0.8 versus 0.4 in spring 2006. Vocabulary scoring errors, however, increased over time, ranging from one to seven percent, compared with zero to four percent in spring 2006 (the mean number of scoring errors was 0.7, compared with 0.1 in spring 2006). Similarly, site visitors cited mispronunciation errors on several items; for example, one assessor repeatedly instructed children to "point to *pick*ing" instead of "point to *peek*ing," which often confused them. Another assessor said "decorate" and "surprise" instead of "decorate*d*" and "surprise*d*." No pronunciation errors were recorded in spring 2006.

Assessors also made a small number of procedural errors during the vocabulary section. On 2.3 percent of assessments, program staff neglected to use encouragement in cases when

[30] 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.

a child seemed reluctant to give an answer ("It's okay to guess"), or redirecting the child to the task at hand ("We can talk about that later"). Four percent of assessors did not suggest that the child guess when they should have, and three percent did not help the child master pointing, (for example, clearly pointing to one quadrant so the assessor can see the child's answer).

Administering and Scoring the Letter Naming Task

In the Letter Naming task, children were presented with three plates of upper- and lower-case letters (eight to nine per plate) and asked to point to and name all the letters they recognized.[31] In spring 2007, as in spring 2006, assessors made relatively few administration errors, such as coaching and giving non-neutral encouragement (Appendix Table C.1). Still, incidence of these two types of errors increased slightly over time. For example, program staff coached children on in three to six percent of assessments (depending on the letter plate) in 2007, compared to only two percent in spring 2006. At one program, observed assessors tended to point to letters that a child had not yet identified while asking, "Do you know this one?" Scoring errors were 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 2006 (12 to 18 percent versus 7 to 14 percent, respectively).

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

	Spring 2007	Spring 2006
First plate	7	9
Second plate	9	9
Third plate	11	9

N = 320 English assessment observations (2007); 293 English assessments (2006); weighted estimates

During the Letter Naming task, assessors are instructed to provide prompts in response to various incorrect answers, such as a child saying the sound of the letter, giving a non-English name, or saying "zero" for the letter "O." Assessors failed to provide these probes appropriately in 7 to 11 percent of observed assessments, which reflects a comparable percentage of administration errors from spring 2006 (see Box on this page). Additional errors included a few instances when assessors did not ask the child to continue with this task,[32] could not help the child pace his or her pointing and identifying letters so that the assessor could score correctly, or did not score a letter correctly if named at least once. Observers also reported that about 11 percent of the children (up three percentage points from spring 2006) noticed and commented on or reacted to the scoring.

[31] The Spanish version contains four panels because four additional letters—CH, LL, Ñ, and RR—are included.

[32] Site visitors have observed children who name one or two letters and then stop. Assessors are instructed to encourage the child to continuing naming letters by asking, "Do you know any others?"

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. As in previous rounds, the graph and counting questions were the major source of errors, as well as the measurement item, to a lesser degree. In addition, script errors increased throughout the Early Math section, for reasons that are unclear (Appendix Table C.1). The mean number of math script errors on the spring 2007 assessment jumped to 1.8 from 0.7 in spring 2006; the range of assessments with straying from the script errors increased to 3 to 19 percent from 0 to 7 percent (2006). Gesturing errors also continued to increase, despite changes in Year Two to make gesturing instructions more consistent. Instances of non-neutral encouragement and coaching remained about the same.

Overall, scoring errors were relatively low and similar to error rates in spring 2006. Nonetheless, assessors continued to struggle with accurately scoring the counting task (E20), with this item scored incorrectly in one-quarter of assessments. (Likewise, 27 percent did so 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. A small number of assessors made procedural errors on this item. During three percent of observed assessments, program staff did not control a child's counting by saying, "Please don't jump around. Count the marbles in order, okay? Start here."[33] Similarly, children were not encouraged to keep counting when he or she stopped at the end of a row of marbles on two percent of assessments.

Like past site visit rounds, incorrect gesturing was another substantial source of errors in Early Math Skills, although instances are concentrated among a handful of items. For item E5, assessors are instructed to sweep across the three nests while asking, "How many eggs are there altogether?" Gesturing errors remained about the same—9 percent in spring 2007 compared to 10 percent in 2006. On the measurement item (E14), gesturing errors occurred on 13 percent of assessments (the same as spring 2006). Assessors should circle the response quadrants while saying, "Point to the number that shows how many inches tall the teddy bear is" *without* pointing to the bear. (This gesture was removed in Year Two.) Finally, assessors continued to have difficulty administering graph items (E18 and E19). Gesturing errors were observed in 51 and 34 percent of assessments, respectively, compared with 27 and 31 percent in spring 2006. These difficulties usually resulted from assessors' pointing to the pictures of pets (dogs, cats, rabbits), rather than their names, as instructed, or by circling the entire page rather than just the graph.

[33] Importantly, children do not have to count the marbles in order to get credit, though it does make it easier for assessors to follow a child's counting and then accurately score their final correct response. Children get credit for this activity as long as they count the objects in numerical order (1, 2, 3…) and maintain a one-to-one correspondence while pointing to the objects; it is irrelevant if the child's finger 'jumps around' the plate, as long as these two conditions are met. Assessors must understand this rule, but they must also be able to control the child's counting as necessary.

IMPLEMENTATION OF THE SPANISH-LANGUAGE ASSESSMENT

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 Two, the NRS protocols now call for children whose home language is Spanish to be assessed on the Spanish version first, then English—a change from Year One. 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.

Eighty percent (32 of 40) of programs in our sample reported that at least one NRS-eligible child spoke Spanish at home. Site visitors observed 98 administrations of the Spanish-language version (90 full assessments) across 12 programs and 35 assessors.[34] 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 2006 sample for the NRS Quality Assurance Study (up 11 percent, or 69 percent of programs that assessed Spanish-speaking children). Therefore, differences across years may be a function of the samples from different years not being equivalent rather than true changes over time. Four out of 35 bilingual assessors observed were conducting NRS assessments for the first time in spring 2007.

While procedures for the language screener rarely caused confusion for program staff in spring 2006, site visitors observed more types of these errors during the 2007 observations. For example, assessors failed to choose the correct path after completing the Pre-LAS activities on 1.7 percent of assessments (this never occurred during the spring 2006 visits). Two lead trainers described that their staff frequently raised questions or are confused by the rules for continuing with the entire NRS assessment, depending on the child's home language and on the number of mistakes made on the first two sections. For example, one lead trainer noted that some assessors wrote on the score sheet, "Child does not speak English," and seemed to be confused as to when it was appropriate to continue or stop the assessment. Assessors also made some quality control errors, including not filling in the correct language routing item (5 percent), despite choosing the correct path, and not completing the language screener boxes before continuing with the vocabulary section (20 percent).

As was the case with the sample of programs in spring 2006, site visitors did not raise any concerns about the fluency levels of observed bilingual assessors (i.e., they had no difficulty reading or pronouncing Spanish words in the script). Furthermore, site visitors rarely reported mispronunciation errors, which are observed during the PPVT and TVIP sections.

[34] 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.

Meeting the Certification Standard

Across the 90 complete Spanish assessments observed (8 had missing data or the child did not pass the screener), the mean certification score was 89--4 points above the required minimum score of 85. This mean represents an eight-point decrease from the mean certification score on the Spanish assessments in spring 2006, and is two points lower than the mean certification score for English assessments (91) in spring 2007.[35] Approximately 68 percent of all observed Spanish assessments were completed by assessors with a certification score higher than 85 compared with 75 percent of the English assessments (see Table II.3).

Table II.3. Distribution of Certification Scores Across Observed Spanish Assessments

Certification Score	Number of Assessments	Percentage of Assessments	Cumulative Percentage of Assessments
24	1	0.1	0.1
36	1	0.1	0.2
60	2	1.5	1.7
68	1	2.9	4.6
72	4	6.8	12.2
76	6	9.0	20.3
80	5	5.0	25.3
84	**4**	**6.8**	**32.1**
88	**8**	**6.7**	**41.0**
92	**15**	**20.9**	**59.7**
96	**25**	**25.1**	**84.8**
100	**18**	**15.2**	**100.0**
Total	**90**	**100**	

Source: Spring 2007 observations of the Spanish-language version of the NRS child assessment.

N = 90 (eight children had missing data or did not pass the language screener). Percentages are weighted estimates. Bolded rows indicate scores at or near the certification level.

The average duration of the Spanish assessment across all observations in which the child passed the language screener (and thus completed the entire assessment) was 18.4 minutes, ranging from 9 to 32 minutes in length. In contrast, the English version lasted only 15.3 minutes, on average. In previous rounds, the Spanish-language assessment also took longer than the English version.[36]

[35] The difference in mean certification scores for English and Spanish assessments is not statistically significant.

[36] During interview and focus groups, staff have remarked that sometimes Spanish-speaking children are more shy and difficult to "draw out," which may account in part for the longer assessment times. The difference in the average length in minutes of English and Spanish assessments is statistically significant ($p < .001$).

Errors in Procedures, Administration, and Scoring

Overall, the Spanish assessments were conducted with relatively few errors (Appendix Table C.2). As was the case for the English sample, Spanish assessors made few errors by coaching, giving non-neutral encouragement, or inserting articles. Bilingual assessors made more scoring errors but fewer gesturing errors than the English assessors.[37] The most frequent types of errors were made in scoring, straying from the script, and gesturing (see Box on this page). The mean number of errors on Spanish assessments for nearly all types of errors was higher than in spring 2006 (5.3 compared to 2.1, respectively). Straying from the script errors jumped fourfold (from 1.1 errors in 2006), while scoring errors more than doubled (from 2.1 errors). Only other errors, including inserting articles or mispronouncing words during the TVIP section, decreased.

ASSESSORS' EXPERIENCES ADMINISTERING THE CHILD ASSESSMENTS

During site visits, assessors discussed their experiences administering the NRS assessment, and described aspects of the process that worked well and those that were difficult for them or the children. Within sites visited in spring 2007, like each round of the NRS 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 2007,[38] (1) assessors' reports on children's reactions to the assessment and how they responded to children's behavior, (2) the length of assessments, (3) staff reactions to the assessment process, (4) experiences of assessors who were also the children's teachers, (5) assessors' experiences administering the NRS to children who were English language learners, and (6) assessors' experiences administering the NRS to children with disabilities. Data are largely derived from the focus groups with assessors held at each site, with additional material provided from the NRS lead trainer and Heat Start director interviews.

Mean Number of Errors per Spanish Assessment

Straying from the script	4.8
Coaching	1.1
Non-neutral encouragement	0.8
Incorrect hand gestures	2.3
Mispronunciation/inserting articles	0.1
Scoring errors	5.3
Total errors (all types)	14.1

N = 98 Spanish assessment observations (90

[37] The difference in mean total errors between English and Spanish assessors for gesturing ($p < .001$) and scoring ($p < .001$) was statistically significant. The mean number of gesture errors was higher on the English assessments (3.8 versus 2.3), while Spanish assessments were more likely to contain scoring errors (5.3 versus 2.5).

[38] While program staff were instructed by site visitors to concentrate on their experiences in spring 2007—unless otherwise asked to comment on specific changes over time—it is possible that some information collected on assessors' experiences pertains to previous rounds. When staff described a specific incident and indicated the time frame, we omitted the information from our analysis here (for example, if an assessor recounted an experience with an ELL from a previous program year.)

As in previous years, assessors reported feeling that compared to the fall, administering the assessments was much easier in the spring, largely because the children seemed more comfortable, confident, mature, and knowledgeable about the information and skills being tested. There was less need for probing and redirecting, and 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.

Children's Responses to the Assessment Process

Seventy percent of programs reported that most NRS-eligible children responded positively to the child assessment and were cooperative with assessors, which represents an increase from 51 percent during the previous spring NRS assessment round (see Box on next page).[39] Because it is hard to see why children new to the assessment each year should act more comfortable with it, these changes are likely a function of assessor comfort with the assessment. Approximately one-third of programs (35 percent) reported that children enjoyed the one-on-one attention with an adult and welcomed the opportunity to interrupt their classroom routine (compared to 23 percent in spring 2006). Moreover, three programs (8 percent) said that many children are eager to go with the assessor to demonstrate their skills, one of which reported that children frequently get very excited when they know the answer and yell, "I'm so smart!" Some assessment teams (18 percent) remarked that children often looked forward to playing games; a few programs even referred to the child assessment as "The Game." In fact, one assessor described how she sits with children who are not yet eligible to do the NRS assessment so that they can turn the easel pages, look at the pictures, and 'play' just like their classmates. Staff from another program noted that children really enjoy looking at the colored pictures in the Art Show and Early Math sections, which seem to hold their attention throughout the tasks.

Assessors encountered fewer challenges with children who were shy, unresponsive, or uncooperative in the spring as compared to the fall assessments, a change that staff have noted since the NRS implementation began. Staff from the programs (91 percent) said the assessment went much more smoothly in the spring, for several reasons (see Box on next page). They speculated that children had mastered skills over the course of the program year, and thus were less anxious or hesitant in giving responses. Staff reported that children seemed more confident overall, not only because of familiarity with the process and assessments in general, but also because of familiarity with the Head Start environment as

[39] In calculating percentages, all programs are weighted equally, regardless of their size or the number of people interviewed. The common denominator is always 40 programs—unless noted—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 assessment. In general, specific concerns or activities were counted if at least one respondent at a program mentioned them. Disagreement among respondents within programs is noted when it occurs; data are largely from the focus groups with NRS assessors.

well as with teachers and other staff members who serve as NRS assessors. Children exhibited better attention spans and less timidity during the spring assessment. Twenty percent of programs noted that the spring assessments went faster than the fall version since the assessors did not have to spend time redirecting children or encouraging them to guess (not shown). Four programs noted that some children remembered the content of the NRS from fall 2006, even specific questions. One assessor recalled a child who frequently anticipated what she was going to ask on the PPVT (for example, "You want me to point to the tornado?").

Children's Responses to the Assessment Process

	Percentage of Programs
Easier in the spring	91
Problem behaviors in some children (see below)	75
Most children responded positively	70
Assessment was too long	48
Enjoyed the one-on-one attention	35
Excited about participating, playing games	18
Unable to complete assessment with some children	13
Eager to demonstrate what they knew	8

Reported problem behaviors:

- Difficulty concentrating; short attention span
- Bored or fatigued
- Shyness or non-response
- Wanted to talk throughout the assessment
- Nervous, anxious, or afraid
- Discouraged when they did not know the answer
- Noted that assessor was scoring responses

N = 40 Head Start programs; unweighted estimates

During the spring 2007 site visits, 75 percent of the sample programs reported that a small number of children exhibited challenging behaviors, such as those who had difficulty staying focused, grew bored, distracted, or "antsy," wanted to chat during the assessment (especially during the Pre-LAS Art Show and PPVT), were nervous, afraid, unresponsive, or were generally uncooperative. The most common behavior issue cited by assessors was that some children could not concentrate and had short attention spans. In certain cases, children also appeared to grow bored or tired. Staff from three programs noted that sometimes children started pointing to the same quadrants in the vocabulary section in an apparent effort to hurry the assessment along; some of these assessors described situations in which the child slid down in the chair. Importantly, however, all programs noted that these challenging behaviors were atypical, especially in the spring. Cases of children refusing to be assessed were rare as well. At the time of the site visits, staff from 13 percent of programs had not been able to—or anticipated that they would not be able to—assess one or two children due to something other than a disability (for example, one child was very stressed about the assessment and his parents requested that he be exempt from it).

To address these challenges, assessors reported using a variety of strategies, including those suggested in the NRS training materials[40] (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 on this page). Some assessors thought it was effective to talk to children before the assessment to make them feel comfortable. For example, during one assessment that an MPR site visitor observed, the assessor showed a lot of energy and enthusiasm for the activity, assured the child that he would do well, gave him a 'high five,' and promised a sticker upon completion. About 18 percent of programs reported using an incentive, such as promising the reward of a sticker, cracker, or lollipop at the end, or allowing the child to bring a toy into the testing room, or offering to let him or her turn the easel pages. During the spring 2007 site visits, assessors also noted children's difficulties with the assessment, both during certain sections and with specific items. As in spring 2006, assessors reported that overall children had the most difficulty with specific items from the PPVT, followed by the Letter Naming and Early Math sections (see Box on next page). At the same time, far fewer programs raised concerns about specific sections of the NRS assessment (in 2006, for example, 74 percent of programs reported challenges with the vocabulary section, and 34 percent with Early Math). Few staff raised concerns with the Simon Says (Pre-LAS) and Art Show (Pre-LAS) Language-Screener sections. The rest of this section describes particular items assessors reported that children had trouble with and portions of the assessment that concerned assessors.

Assessors' Strategies for Addressing Children's Behavior

- Try assessment later that day or on a different day
- Break the assessment into multiple sessions
- Redirect attention using techniques such as saying the child's name, repeating question, giving extra neutral encouragement, or assuring child that they will be done soon
- Take a quick stretch, bathroom, or water break
- Spend time getting to know child beforehand; engage in "small talk"
- Give out stickers or lollipops at end of assessment
- Let child hold a toy or promise a toy when assessment is finished
- Assess more confident children first to reassure shy ones
- Allow child's teacher or parent to sit in room next to child
- Schedule next attempt with a different assessor
- Promise cracker at the end to keep child focused
- Allow child to turn easel pages
- Talk about how kindergarten students need to know what is on the NRS; ask child to show what he/she knows
- Schedule next attempt during home visit
- Use a pretend voice to convince child to cooperate
- Push through to finish assessment in one sitting

N = 40 Head Start programs

[40] One program appreciated that the Office of Head Start now allows local programs to use more frequent neutral encouragement than the specific scripted statements included in the easel script; this was not the case during the first two years of the NRS implementation. There is no guidance in training suggesting that assessors cannot use a warm tone—just that they cannot give non-neutral encouragement.

Simon Says (Pre-LAS) and Art Show (Pre-LAS). Once again, nearly all sample programs reported that overall children enjoyed the Simon Says and Art Show tasks. They view these sections as fun activities for children that serve as a good warm up as well as an effective language screener. Similarly, almost no programs had any concerns about the Pre-LAS Art Show.

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

PPVT	38
Letter Naming	28
Early Math	8
Simon Says	3
Art Show	3

N = 40 Head Start programs; unweighted estimates

PPVT-III (Adapted). As in the spring 2006 site visits, the PPVT task generated the greatest overall difficulty for children; assessors in just over one-third of the programs mentioned some kind of issue with the vocabulary section. The most common issue cited by far (11 programs) was the length of this activity. Assessors frequently described how children become bored, restless, and fatigued. Staff from a handful of programs noted that some children lose interest and start to point randomly to the quadrants without thinking through their answers. To help them focus, one assessor reported using additional neutral encouragement.

Fewer programs than in past rounds of the NRS Quality Assurance Study cited concerns with the content of the PPVT. While it was the most prevalent issue raised about the vocabulary section in spring 2006 (38 percent), this year only three programs criticized the cultural appropriateness of the NRS assessment and whether it is regionally biased. Although these issues are not representative of programs as a whole, they may indicate problems that can affect scores for at least a small number of children. For example, staff from a rural community said that children have never seen a *fountain,* and their concept of a *fence* is the metal type used for ranching, not the wooden picket fence depicted in the vocabulary section. Another group of assessors in a program that serves a large number of non-Spanish ELL children explained that the local ethnic word for *window* sounds very much like the English word for *fence*, and in almost every case children point to the picture of the window on that plate. In addition, staff from three programs believed that some items (*selecting, surprised, group, globe, knight)* were confusing, especially for preschoolers. They asserted that two people (shown in a distracter plate) could be considered a *group* (the correct quadrant shows four people), and that items shown in two different quadrants resemble a *vase.* Two programs observed that children had difficulty with the black-and-white pictures and suggested having color pictures or photographs. Finally, one program noted that it can be difficult for assessors to move through this section efficiently since some children like to tell stories about the pictures, no matter how many times staff try to redirect them to listening and pointing.

Letter Naming. More than one-quarter of programs (28 percent) reported difficulty in administering the alphabet task. This represents a substantial decline from the previous two spring site visits when nearly half of the sample programs reported trouble with this section. Four programs critiqued the design and layout of the plates, saying that there are too many

letters on each plate. As in previous rounds, some programs suggested that assessors be allowed to point to each letter and ask, "What letter is this?" or "Do you know this one?" or recommended reducing the number of letters per plate. They felt that these strategies would help to ensure that children do not skip any letters they know or forget which ones they had already identified, and could help to focus those who may become overwhelmed by many letters on one plate. Staff from one program asserted that the assessment seems to be testing not only alphabet knowledge but also memory and the spatial relationships of the letters. Three programs noted that the Letter Naming section could be quite frustrating and uncomfortable when children only know a few letters but must listen to the assessor repeatedly prompt, "Do you know any of these?" Likewise, staff from two programs said that the alphabet task can be long and burdensome for children who struggle to name letters.

Early Math Skills. During recent years, fewer programs have criticized the Early Math section during the spring assessment round (74 percent in 2005, 34 percent in 2006, and only 8 percent in 2007), which may reflect an effort to improve the math questions over time. Only two programs raised concerns with specific items. One group noted that children sometimes became confused during the counting exercise (E20) and wanted to start over with "one" at each different-colored row. Another program described how several children gave the wrong answer to the question, "How many eggs are there altogether?" (E5). Staff said children appeared not to hear the word "altogether" and instead counted the nests or the eggs in a single nest. As such, some assessors repeated the question, emphasizing the word "altogether." (This technique is considered to be coaching and is against NRS assessment procedures.)

Length of Assessments

Forty-eight percent of programs reported that, in general, the NRS assessment was too long and length may have contributed to challenging behaviors from certain children. Program staff reported that some children lost focus and became bored, tired, or frustrated, citing the PPVT and Letter Naming as causing the most difficulties for children. (Concerns about the length of these sections have emerged repeatedly in previous assessment rounds as well.) One assessor described how she moves briskly through the vocabulary plates to minimize children's fatigue or lack of focus, while another added supplemental neutral encouragement.

Staff Reactions to the Assessment Process

The majority of assessors felt quite comfortable managing the assessments, recording answers, and addressing challenging behaviors; they had ample experience dealing with these situations in the classroom or in other professional experiences. Nonetheless, some programs described difficulties they encountered, either with managing children's behaviors or facilitating the assessment. Six groups of assessors noted that while they feel confident administering the NRS assessments, they still struggle with remaining neutral and not using extra encouragement or coaching. In addition, four programs wanted more clarity on how to determine when it is appropriate to complete an alphabet plate in the Letter Naming section. For example, one group noted that when they asked the scripted prompt, "Do you

know any others?" children sometimes repeat letters that they had already named, and the assessors did not know whether to nod without saying anything or just sit silently. They felt because of the guidance in the easel, this task lasted longer than necessary and created an awkward atmosphere for the assessor and child. Similarly, staff from one program wondered how long they needed to allow for a non-responsive child to provide an answer before moving onto the next item. These program concerns may reflect a need for improvements in training.

Other administration challenges emerged as well. Although these were only mentioned by staff in a handful of programs, they may indicate a need for improvements in training or support for assessors. One group of assessors noted that some children have a tendency to react to scoring answers. In most cases, children are simply curious and it was easy to refocus their attention back to the task at hand. However, one assessor described a child who appeared to "bait" the assessor by declaring, "I'm getting those wrong." Since she could not comment on his performance, she ignored these comments and continued with the questions. Staff from one program said that they were generally uncomfortable with the NRS procedures. If at any time they went 'off script' to address a child's behavior, they became flustered and had difficulty getting back on track with the script. Lastly, one group of assessors noted that children enrolled in the home-based option always come to the center to be assessed. Parents usually watch the assessments and interject comments, such as "You know that answer," which places undue pressure on the child and makes the assessor feel uncomfortable.

Experiences of Assessors Who Are Also Children's Teachers

Teacher-assessors may have different impressions about the assessment process from those who do not work with the children on a daily basis. In spring 2007, a little more than half of sample programs (55 percent) used at least some teachers as the NRS assessors. Of those, most permitted these individuals to assess children enrolled in their own classrooms. One program in our sample left the staffing decision up to each center supervisor. During the focus group, a teacher from this program wished that she could administer the assessments to children in her classroom because she believed they would be less nervous and thus would perform better. Staff from two programs who assessed children from their classroom and other classrooms agreed that the former group was noticeably more comfortable during the test, which could have affected their performance. The difference in mean total errors between teacher-assessors and other staff for coaching (1.4 errors versus 1.3 errors) and non-neutral encouragement (1.1 and 1.0) were not statistically significant; teachers were no more likely to make these types of errors than other program staff.

The extent to which children's performance on the NRS matched teachers' expectations of their abilities varied, as in the past. Of the 18 programs where teacher-assessors commented on the children's performance during the focus groups, assessors from 7 programs said that their expectations of the children's skills usually lined up with performance on the assessments. In contrast, two programs observed that the assessment's one-on-one format enables some children to perform better than they do in class. Teachers explained that they often assumed that some children who did not raise their hand in class or

talk a lot just did not know the answers, but that in fact they were smart children who tended to become quiet in large groups.

Nine programs reported that children tend to demonstrate less knowledge on the NRS assessment than they had demonstrated in the classroom. Three assessment teams felt that this inconsistency emerged specifically in the Letter Naming section as a result of the section's design. Children who had previously demonstrated solid alphabet knowledge in the classroom often became confused as to which letters they had already pointed to and identified and inadvertently omitted letters that they knew. Staff from other programs remarked that the children may not do as well as expected because they (1) become shy or anxious, (2) grow bored and point to pictures randomly or give answers quickly, or (3) have a more difficult time responding to questions in such a formal, "depersonalized" testing environment. A few programs noted that this pattern of demonstrating less knowledge than the teacher believes they possess could be particularly frustrating for staff, as programs in previous assessment rounds have observed. One teacher-assessor declared, "You just know they know the answer! They know it in class!"

Programs rarely reported gleaning additional insight while assessing children from their own classrooms in the spring (note that this question is different from the contributions of the NRS overall that are included in Chapter VI). Many teacher-assessors explained that by this point in the program year, they are very familiar with the children's capabilities through classroom experiences and results from local assessments and screenings. Still, one teacher-assessor noted that she uses the Letter Naming exercise to gauge how well her children have mastered the alphabet over the course of the Head Start program year. In addition, one lead trainer recalled assessors reporting that the NRS helped them to identify potential behavior problems and "red flags" in cases when they are not familiar with the child. Two programs noted 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 a child could follow basic instructions or count above 10.

Teacher-assessors from two programs noted that it was challenging not to praise the child during the NRS assessment. They reported that some children became concerned when they did not hear positive feedback since they were accustomed to it in the classroom. For one program, maintaining a neutral tone became particularly difficult while assessing children with disabilities.

Experiences Assessing Children with Limited English Language Skills

The majority of Head Start programs (32 out of 40 in our sample) had assessed or planned to assess English language learners; one program was located in Puerto Rico where Spanish is the language of instruction, although staff also assessed a few Spanish language learners this year. Among these, 27 programs had planned to assess at least one child in Spanish or had already done so, and 13 programs had assessed or expected to assess English language learners who did not speak Spanish. As in spring 2006, the most common home languages identified by staff—other than Spanish—were Chinese and Arabic dialects;

Vietnamese and Native American languages (Kikapu, Lakota, and Navajo) were cited more frequently by program staff than in spring 2006. In spring 2007, our sample included 5 AI/AN programs, more than in other years. Other languages included Afrikaans, Albanian, Creole, German-Dutch, Italian, Korean, Portuguese, Somali, and Yiddish.

By the spring 2007 assessment round, programs reported that most English language learners were or would be able to complete the English version of the NRS assessment, having been exposed to enough English language instruction in the classroom. At the same time, staff from four programs noted that there were some children who completed the language screener, only to struggle on the remaining three sections because they did not know enough English. As one director described, "They make it through Art Show and Simon Says but then the rest of it is a huge hurdle." Assessors from another program remarked that some children pass the screener, but they take longer to process each question and usually perform poorly on the PPVT due to limited vocabulary fluency.[41]

While most programs in our sample did not raise concerns about the effectiveness of the Pre-LAS activities to gauge language ability, five programs described how difficult it was to use the home language recorded by the program to determine if a child from a Spanish-speaking household should be tested in Spanish first. They mentioned that often there is a discrepancy between the language reported by parents and the children's actual language fluency. Four of these five programs used strategies to minimize the likelihood of testing children in a language they did not comprehend. One assessment team typically consults with the child's teacher beforehand to find out the child's primary language used in the classroom and use that language version of the child assessment. Another assessor stopped the assessment after the language screener when one child—who was classified in program records as a Spanish speaker—said, "I don't speak Spanish, my mother speaks Spanish."[42] One teacher, who was quite concerned that children did not understand the questions on the Spanish or English versions, frequently repeated the question or command in the other language to ensure that children at least comprehended what she was asking them to do. At one program, if a child whose primary language is Spanish did not pass the screener on the English version in the fall but substantially increased his or her English fluency over the year, the child was not assessed at all in Spanish in the spring. These problems, although only occurring in a small number of programs, may indicate a need for clarification of the recommended procedures for deciding which children to assess in Spanish.

Aside from challenges associated with assessing Spanish-speaking children, or those identified as such by their parents, two programs had concerns with the child assessment as

[41] It was unclear when these children enrolled in Head Start. Presumably an English Language Learner who enrolls later in the program year may be less likely to pass the screener in the spring or fully understand the assessor throughout the remaining three sections. Indeed, assessors from five programs noted that whenever English language learners begin Head Start in the fall, they could always complete the child assessment in the spring.

[42] According to the NRS protocols, the Spanish-language version of the assessment should have been completed for this child.

it pertains to other English language learners. These programs enroll a majority of English language learners who speak their primary languages at home and also receive some classroom instruction in these languages, which they feel makes it more difficult for them to do well on the English-language assessment. One program director described how children often "withdraw, get ashamed, and shut down" when they cannot respond in their home language. Likewise, staff report that they get very frustrated when they cannot respond to questions from the children in their language—not questions on the NRS but if children become confused about the procedures or what they should be doing. They also do not understand why non-verbal responses can be accepted on certain Early Math items, but responses in languages other than English (or Spanish) cannot be accepted.

Although the spring 2007 assessment round marked the eighth time that local Head Start programs were conducting the NRS, staff from eight programs in our sample appeared confused as to the appropriate procedures to follow when assessing English language learners. Examples described by program staff included:

- Assessors were uncertain as to when they should proceed past the language screener, or whether they should proceed to the English assessment at all (four programs);

- Assessors were confused about which language to assess Spanish-speaking children first (two programs);

- Assessors did not prompt English language learners (non-Spanish) for the names of the letters in English when children responded in their home language;

- A program with 22 percent Hispanic children (according to its lead trainer) did not administer the Spanish-language assessment; no details were provided as to why this program only administered the English version of the child assessment.

Experiences Assessing Children with Disabilities

Training materials state that children with severe disabilities may be exempted from the assessment, but those with milder disabilities should still be assessed.[43] Most programs in the sample (93 percent) assessed at least one child with identified disabilities. Therefore, most programs needed to consider what kinds of accommodations they could make for these children. Like past assessment rounds, speech and language delays emerged as the most common type of disability encountered by assessors, followed by cognitive and

[43] One assessor assigned to assess a child with severe behavior challenges allowed the child to stand, invited his teacher to sit in the room, and also liberally paraphrased the script to try and redirect the child back to the exercise. A second program permitted a speech therapist to be present during the assessments to "translate" for the children since the NRS assessor could not decipher their responses. Two programs reported assessing children who presumably should have been excluded from testing. One child was non-verbal due to cerebral palsy, and another child with Down Syndrome was assessed, answering the questions "the best way he could," according to his NRS assessor.

developmental delays, unspecified behavioral or social/emotional problems, and autism. Other less common disabilities cited by program staff included ADD/ADHD, unspecified physical or occupational impairments (for example, a condition that required the use of a walker), visual impairment or blindness, orthopedic disabilities, Down Syndrome, oppositional defiant disorder, cerebral palsy, and spina bifida.[44]

Reported Accommodations Made by Head Start Programs for Children with Disabilities

- Repeat questions or ask child to repeat answers
- Split assessment into multiple sessions, take breaks
- Give extra encouragement and redirection to children with behavior issues
- Speak slowly, go slowly, and allow more time
- Invite the speech therapist to be present because assessor could not understand child's answers[a]
- Child can use assistive devices for pointing
- Target staff member most familiar with a child's speech patterns to administer the NRS assessment
- Consult with disability specialist to determine if child should be assessed
- Permit child to stand and have teacher present
- Give credit to child with speech impediment even if assessor cannot discern exactly what was said
- Alter script for child with a behavior disorder[a]

N = 37 Head Start programs (three programs did not assess children with disabilities in spring 2007)

[a]Not consistent with NRS protocols.

The extent to which assessors were able to complete the NRS assessments on children with disabilities varied considerably across programs. Among the 37 programs that assessed children with special needs in spring 2007, about two-thirds reported that they successfully completed them. These programs either did not encounter any challenges, or assessors dealt with periodic challenges by providing minor accommodations (for example, allowing extra time to respond or using redirection techniques). On the other hand, 14 programs did not complete assessments in certain circumstances. For example, in most cases, staff would not even attempt to assess children with more severe disabilities (for example, an acute cognitive delay). Two programs allowed assessors to begin the NRS but stopped after the language screening if the child struggled to respond or understand the assessor's questions or instructions. For the most part, only a small number of children (usually one to four percent) could not complete the assessment.

Programs identified several types of accommodations that they used in assessing children with special needs (see Box on this page). 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.[45] For the English-language assessments, the difference in

[44] In our sample of 323 English child assessments, 46 observations (10 percent) were of children with an Individual Education Plan (IEP), according to staff report. Nine out of 98 observed Spanish assessments (nine percent) were administered to children with a diagnosed disability. Percentages are weighted estimates. None of the assessors observed during our site visits made accommodations for children with disabilities.

[45] For the English-language assessments, the difference in mean total errors between assessing children with disabilities and those without for straying from the script ($p < .001$) and coaching ($p < .001$) was statistically significant, which could be explained by assessors attempting to reword the question or repeat the

(continued)

mean total errors between assessing children with disabilities and those without for straying from the script ($p < .001$) and coaching ($p < .001$) was statistically significant, which could be explained by assessors attempting to reword the question or repeat the question for a child with special needs. Assessments administered to children with disabilities had an average of 6.9 script errors (compared to 4.1 errors for children without disabilities) and an average of 3.0 coaching errors (as opposed to 1.3). For English- and Spanish-language assessments combined, assessors were more likely to coach children with disabilities (statistically significant, $p < .01$). Two programs arranged for children with speech and language disabilities to be assessed by either their own teacher or the disability specialist because they were familiar with the children's speech patterns. In addition, assessors often used strategies such as taking breaks, giving frequent encouragement and neutral praise, or completing portions of the assessment on different days, especially for children with behavioral issues who had difficulty cooperating or staying focused. At two programs, assessors routinely consult with the disability specialist to determine if children should be exempted from even attempting the NRS assessment due to their IEP.[46] In contrast, staff from nine programs reported that assessors did not need to make accommodations for children with special needs. Assessors from two programs noted that while they do not make special accommodations for children with delays or disabilities, they recognized that these children do not perform as well as their peers. (None of the assessors observed during site visits made any accommodations.) A few programs reported having made an accommodation in violation of the NRS assessment guidelines (such as changing the script for a child with a behavioral disorder).

Of the 37 programs that assessed children with disabilities during spring 2007, 11 programs offered feedback on whether they had received sufficient training and support for these assessments. Assessors from five programs reported that they had received guidance and that it has been adequate. One program found the Assessor's Guide to be helpful in explaining when it is appropriate to exempt children with special needs from the NRS assessment. In contrast, two programs would like to receive additional technical assistance on assessing children with disabilities. While local programs have the discretion when to exclude a child from being assessed, they only exempt a child if his or her IEP states that no assessment should take place.

Staff from three programs could not recall having received any guidance on assessing children with special needs, nor did they indicate a need or preference for additional training or support. Assessors from a fourth program observed that while they do not need any more technical assistance on how to assess children with special needs, they did not feel that

(*continued*)
question for a child with special needs. Assessments administered to children with disabilities had an average of 6.9 script errors (compared to 4.1 errors for children without disabilities) and an average of 3.0 coaching errors (as opposed to 1.3). For English- and Spanish-language assessments combined, assessors were more likely to coach children with disabilities (statistically significant, $p < .01$).

[46] We are not aware of particular instructions given to programs in training regarding whether or not to attempt assessment with children with disabilities.

the training video effectively addresses this topic. They stated that while the video effectively addressed how to handle uncooperative children, it did not discuss the issue of administering the NRS assessment to children with disabilities nor offer tips for managing these testing situations.

SUMMARY

As in earlier rounds of the NRS Quality Assurance Study, most of the 312 English language assessments observed in spring 2007 met or exceeded the standard of quality used in certifying assessors. The average certification score was 91, and 75 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 (90 in 12 programs) was also relatively high, on average (89 points).

For the English-language assessments, scoring errors along with all types of administration errors due to coaching, inappropriate gestures, non-neutral encouragement, straying from the script, and inserting articles such as "a" and "the" were higher in spring 2007 than in spring 2006; no error categories were less frequent than the previous spring. Likewise, errors increased from spring 2006 to spring 2007 on the Spanish-language assessment, with the exception of "other" errors (inserting articles or mispronouncing words during the TVIP section). This increase across nearly all errors may be a function of the amount of time that has passed since the NRS implementation began (and some staff may feel less compelled to adhere strictly to NRS protocols). For example, the mean number of script errors reached the highest since the NRS Quality Assurance Study began (4.8 errors per assessment). It is possible that many assessors have become so comfortable with the assessment procedures that they have taken certain liberties in stating the script. However, it could also reflect that different staff are completing assessments now than in 2003.

Although quality was high overall, some areas of the child assessment were difficult for staff. Modifications to the assessment easel or additional guidance and training might be helpful in certain components of administering the assessment. The majority of these areas were also problematic in previous years of the NRS, including:

- Coaching, particularly in the Simon Says and Tío Simón sections
- Gesture errors on certain Art Show items, (especially B5 to B10), Early Math items (especially E14, E18, and E19), and Conocimientos Básicos de Matemáticas items (especially EE16 to EE19)
- Scoring the counting item correctly in the Early Math (E20) and Conocimientos Básicos de Matemáticas (EE20) sections

Coaching and gesturing errors were somewhat more prominent as compared with spring 2006. Scoring errors on the counting item, however, remained essentially unchanged—it was scored incorrectly on 25 percent of observed English-language assessments (bilingual assessors have scored this item incorrectly on approximately 17 percent of assessments in spring 2006 and 2007). 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 remained a section with which many programs continued to struggle. Assessors continued to challenge the validity of this section, since children may very well know some or all of the alphabet letters, but the structure of this test might not enable children to demonstrate what they know. Assessors believed that allowing them to point to each letter and say, "What's this?" or "What's the name of this one?" could help them ensure that children do not inadvertently omit a letter that they lost track of while scanning the plate.

Seventy percent of sample programs reported that most children had reacted positively to the child assessment. As in previous years, assessors reported that some children enjoyed the one-on-one time with staff and the opportunity to play a game, along with the chance to demonstrate what they had learned during the year. Children's behavior was much less of a concern in the spring than it had been in the fall, a finding similar to previous rounds of the NRS Quality Assurance Study. The most common 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 is too long. To address these challenges, assessors frequently attempted the assessment at a later time or split the assessment into multiple sessions, took a quick stretch or bathroom break, or offered the children extra neutral encouragement or small incentives.

Most programs (32 out of 40, or 80 percent) administered the child assessment to English language learners. Spanish was the most common other language spoken, followed by Chinese and Arabic. Staff reported that children usually could pass the English language screener in the spring, even if they had not passed in the fall. However, assessors often lacked understanding of how to determine the best language in which to assess a Spanish-speaking child, if their view was in conflict with information provided by parents. As has been the case in previous rounds of site visits, some bilingual assessors were critical of the Spanish version of the NRS because they felt it was unfair that children who use colloquial words from different regional dialects are not given credit on the NRS when these words are logical responses but are not responses accepted as correct on the assessment.

Nearly all programs (37 out of 40 in the sample) that assessed children with disabilities reported using a wide range of accommodations for the assessment. In most cases, programs successfully administered the NRS to children with special needs. These programs either did not need to make any accommodations, or else used modest ones, such as allowing the child extra time to respond, taking breaks, or repeating questions. A small number of programs inappropriately assessed special-needs children whom they should have exempted or neglected to use appropriate accommodations. In contrast, about one-third of the 37 programs did not complete the child assessments. Staff usually exempted children with severe disabilities, or, in a few cases, stopped after the first few sections if it became clear that the child was struggling to understand the tasks and questions. Most assessors felt prepared to administer the NRS to children with disabilities and making necessary accommodations. However, a few programs would like the training materials to explicitly address the issues of conducting the assessment to children with special needs and making appropriate accommodations for these situations.

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

CHAPTER III
LOCAL APPROACHES TO TRAINING ASSESSORS

The validity of the NRS data depends heavily on how thoroughly assessors understand the importance of standardized assessment administration, and how carefully each assessor conducts the child assessment and records child responses. Local assessor training is a key component that supports the reliability and usefulness of the NRS assessment data. OHS has emphasized the importance of training since the NRS began and has provided national training opportunities for lead trainers, training materials for local assessor training, and resource help lines for technical assistance.

Due to the need to conduct a large number of assessments across the country, OHS required an extensive cadre of certified assessors to implement the NRS. Recognizing the short timeframe in which assessors were to be trained, OHS 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 administer the assessments at the local level. OHS 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 completing the initial round of training 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 new items. In addition, some programs have provided separate trainings and certification for new assessors not previously trained. In summer 2006, prior to the fourth year of NRS implementation, OHS conducted three regional ToT conferences (in Los Angeles, Atlanta, and Kansas City, MO) for programs that had new lead

NRS trainers.[47] For the fall 2006 assessment, OHS and its contractors prepared materials for local programs to use in comprehensive fall trainings for all local assessors. In preparation for administering the spring 2007 assessment, programs then held local refresher trainings for those staff members who had attended fall training, and a limited number of more comprehensive trainings for new assessor trainings.

This chapter describes the approach programs used in spring 2007 to train NRS assessors in spring 2007. Based on individual interviews with Head Start directors and NRS lead trainers, focus group discussions with assessors in the 40 sampled programs, and observations of local NRS trainings at a subset of sampled programs, we describe local programs' approaches to conducting refresher and new assessor training for the English- and Spanish-language assessments. We also describe the perspectives of local program staff on the training, materials, and support they have received from OHS and its contractors.[48]

LOCAL RESPONSES TO SPRING 2007 TRAINING MATERIALS AND GUIDANCE

In this section, we describe the range of approaches that programs used to prepare both their English- and Spanish-speaking assessors for the NRS, as well as the specific training approaches that programs used for new NRS assessors. While we have discussed staff experiences with the regional ToT conferences in previous reports for the Quality Assurance Study, we omitted those questions in this round to allow time for new questions about the SED rating form. Among the 40 programs we visited, lead trainers at five raised this topic in the absence of a direct question about it. All five reported that scheduling or budget problems at their program had prevented them from attending a ToT conference in summer 2006. The trainers suggested either increasing the frequency of these conferences or holding them at locations where the travel costs for local programs would not be prohibitive.

PROGRAM APPROACHES TO LOCAL NRS REFRESHER TRAININGS

OHS provided programs with guidelines and refresher trainings for experienced assessors. According to the guidelines, a complete training session consisted of four parts: (1) viewing the video, (2) question-and-answer session, (3) role playing, and (4) distribution of the Assessor's Guide. OHS sent the Assessor's Guides, videos, role-playing scripts, assessor certification forms, new easels, and score sheets to programs. Programs also received a recommended refresher training agenda that included about 1.5 hours of

[47] Training sessions for replacement NRS leads have been conducted every year.

[48] In calculating percentages in this chapter, all programs were weighted equally regardless of their size or the number of staff interviewed. The common denominator when discussing the training for the English assessment is always 39 programs, regardless of whether a given program had an opinion on a certain topic. (One program located in Puerto Rico is excluded in discussions about training for the English-language version of the assessment.) If several programs did not provide information on a given topic, we presented the number of applicable programs (for example, "Six programs decided…") as opposed to a percentage of programs. In general, specific concerns or activities were counted if at least one respondent at a program mentioned them. Disagreement among respondents within programs was noted when it occurred.

activities, with an option to add a second role play and extend the training to two hours and 20 minutes. This is a reduction in suggested training time from spring 2006, when OHS suggested that programs spend 4.5 hours on the English refresher training. The revision was made in response to feedback from prior rounds of this study, in which programs noted that the recommended training protocol was too long. A review of the Assessor's Guide during training was eliminated from the OHS-suggested agenda, as well as the second round of role plays and discussions of changes to the assessment and addressing behavioral problems. The spring 2007 guidelines for training bilingual assessors recommended an additional 30 minutes to be spent reviewing the Spanish training video and conducting a question-and-answer session.

Among the 40 programs we visited in spring 2007, 37 (92.5 percent) provided staff with a refresher training of some kind, and in 23 (58 percent) of programs the training included all four elements in the OHS guidelines. Of the three that did not provide a refresher training, two programs encouraged assessors to do a self-review, distributed the Assessor's Guide and new easels, and made the video available for staff to watch on their own time. Some assessors at these two programs reported during the focus groups that they did not review any training materials for the spring 2007 assessment, so we count these programs as not having provided training. The third program, a delegate agency, reported that its program agency instructed staff that a refresher training was not required because there were no new assessors on staff. (One of the 37 programs providing refresher training is located in Puerto Rico, where all training and assessments are in Spanish, so we eliminated that program from the discussion of English refresher training, but included it in discussion of Spanish training.)

English Refresher Trainings

Programs reported a range of activities conducted during their refresher trainings. The majority of programs (59 percent) that scheduled a refresher training in spring 2007 provided their assessors with the full range of activities suggested by OHS, including watching the video, answering questions, role playing, and distribution of the Assessor's Guide (see Box on this page). This is a considerable increase over spring 2006, when only 27 percent of sampled programs provided comprehensive refresher training to their staff. The suggested duration of the spring 2007 refresher training declined over the past year, and it is possible that programs found the abbreviated agenda to be more feasible.

Approaches to the Spring 2007 Refresher Training for the English-language Assessment

	Percentage of Programs
Included all required elements Video, Q & A, role play, and distributed guide	**59**
Included 3 of 4 required elements	**28**
Video, Q & A, guide	23
Different set of 3 elements	5
Included 2 of 4 required elements	**5**
Self study or no refresher training	**8**

N = 39 Head Start programs

Programs also reported considerable variation in the amount of time that they dedicated to NRS training sessions. Some programs devoted half a workday or more, while others

spent less than one hour providing refresher training. The average length of training among the 36 programs that provided refresher training in English was slightly longer than two hours. This slightly exceeds the 1.5 hours training length suggested by OHS. Not surprisingly, programs that chose to provide a refresher training that included all four required elements reported longer training sessions, on average, than programs that provided only some elements. Programs that offered full refresher training reported an average training length slightly longer than two-and-a-half hours, while those that offered partial sessions lasted an average of one-and-a-half hours. For lead trainers, the relative difficulty of scheduling a group training for staff and the number of staff to be trained were key factors in determining the amount of time devoted to refresher training this spring.

Among the 41 percent of programs that did not provide full training with all required elements, most provided refresher training of some kind to their staff. Only three programs (eight percent of the total) provided no training at all, and two of those three made materials, such as the video and Assessor's Guide, available for staff to review on their own. Most programs that did not schedule full refresher training did choose to provide three of the four required elements. The training element most commonly excluded was the role-playing activity. For programs where a role playing activity was included, staff suggested improving the scripts by adding more problematic child behaviors, and by separating the adult and child portions of the script into separate documents (rather than listing them side by side) for ease of use.

At several programs, MPR collected conflicting information from administrators and assessors about the length and content of training activities. For example, at 3 of the 23 programs that we counted as administering a full training, the focus group participants did not recall any role-play activities, while the lead trainer at those programs maintained that the role-play had been a part of training. At the four programs where staff disagreed about the range of activities included in training, we credited the programs with providing training that any member of the staff recalled. In cases where the staff disagreed on the length of training (for example, the director reported that the training was shorter than the focus group recalled), we used the midpoint of recalled training lengths for calculating cross-program averages.

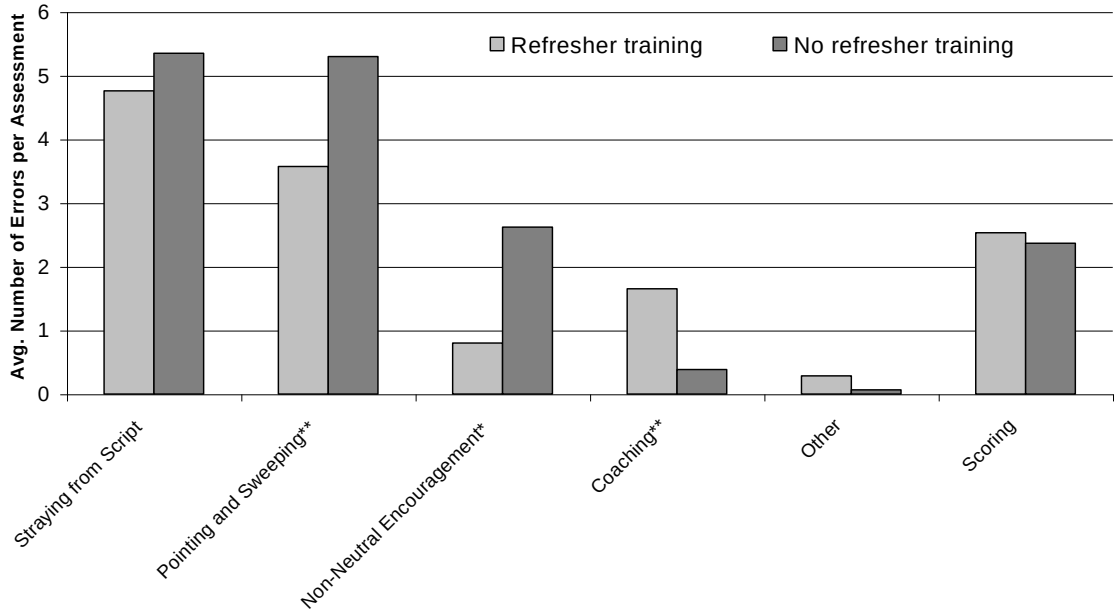
In addition to questions about training length and content, we also asked lead trainers if they conducted follow-up observations to monitor the quality of administering assessments. In 25 percent of programs (including the Spanish-only program in Puerto Rico), the lead trainers conducted some form of monitoring this program year. Of these, sixty percent reported that they only observed a subgroup of assessors, such as those who struggled during training, new assessors, or assessors in the center nearest to the lead trainer's office. Only two programs reported formally tracking errors with a certification form. The others said that their monitoring process was more informal, typically consisting of lead trainer observation followed by a discussion about ways the assessor could improve.

In spring 2007, a large number of programs expressed a desire for shorter or less frequent training for experienced assessors. At 28 percent of programs, staff indicated a preference for having only annual trainings for experienced assessors. They reported that

the current frequency of training complicates their schedules and contributes to staff boredom and frustration with the NRS assessment process. Staff at three of these programs suggested the additional step of shortening the current training schedule by eliminating the role-play exercise. Staff at a single program did not always agree; at one program where the lead trainer had suggested eliminating the role-play activity, assessors in the focus group said it was the most useful aspect of training and would like more opportunity for this activity.

The lack of refresher training appeared to be related to the likelihood that assessors would make errors during NRS assessments. For example, we observed a significantly greater incidence of errors in pointing and sweeping and non-neutral encouragement at the three programs that offered either no refresher training or self-review of the materials only (Figure III.1). In contrast, staff members at these programs were about as likely to commit scoring errors and stray from the script as staff that had attended refresher training, and were also significantly less likely to make coaching errors. We also compared the average certification scores (the extent to which assessments we observed met certification standards) of programs that offered refresher training (91 percent) and those that did not (89 percent). While programs that offered refresher training scored higher on certification, on average, the difference was not statistically significant.

Figure III.1. Observed Errors for Programs that Offered/Did Not Offer Any Refresher Training (English)



Source: MPR observations of 39 Head Start programs administering the English assessment (n=316 assessment observations).

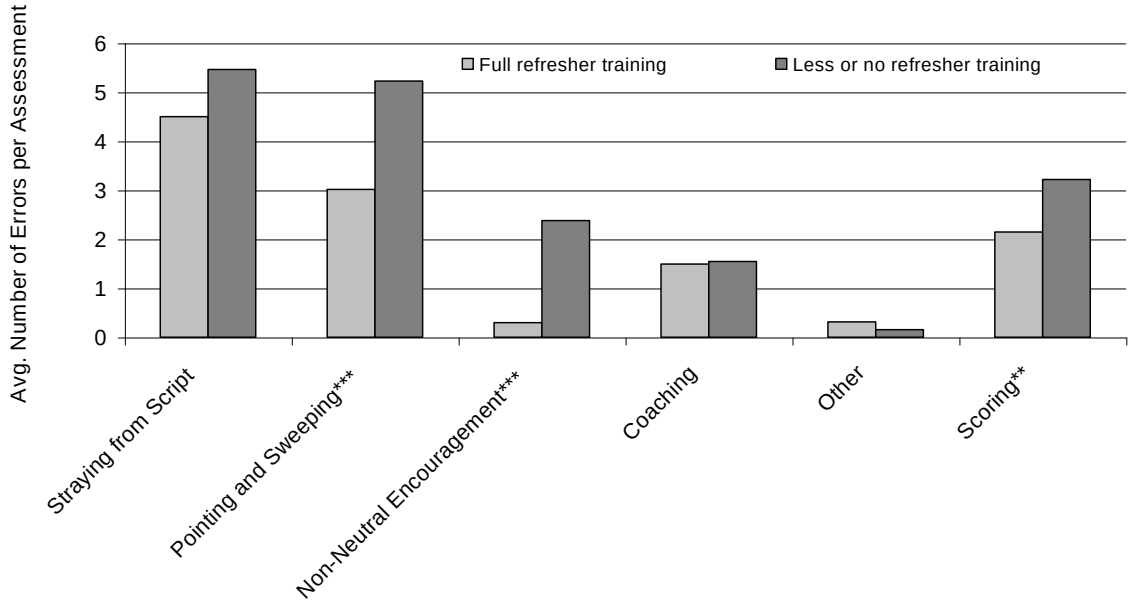
Notes: "Other" errors include omitting gestures, along with inserting articles (for example, "a" and "the") and mispronouncing words during the Peabody Picture Vocabulary Test-III (PPVT).

*Statistically significant at p<0.01

**Statistically significant at p<0.05

In addition, we compared the 23 programs that offered a full refresher training that incorporated the video, role-playing, question and answer session, and distribution of the Assessor's Guide to those 16 programs that offered less (or no) refresher training. Some differences in the frequency of errors emerged, with those programs that provided fewer training elements making significantly more errors on average (Figure III.2). For example, assessors at programs with less training were significantly more likely to make gesturing errors and to provide non-neutral encouragement. Moreover, programs that opted for less comprehensive refresher training were significantly more likely to make scoring errors, which could have important implications for the accuracy of the data that OHS receives. As described earlier, most programs that provided less comprehensive training chose to omit role-playing. In this activity, assessors typically practice recording scores in the score sheet and evaluating the accuracy of responses. It may be that, in the absence of a role-playing activity to practice scoring, the programs made more scoring errors during the assessment.

Figure III.2. Observed Errors for Programs that Offered/Did Not Offer Comprehensive Refresher Training (English)



Source: MPR observations of 39 Head Start programs administering the English assessment. (n=316 assessment observations).

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

** Statistically significant at p<0.05

*** Statistically significant at p<0.01

Spanish-Language Refresher Trainings

Of the 40 programs MPR visited in spring 2007, 70 percent assessed children in Spanish. This section summarizes the range of activities that programs conducted while providing refresher training to their Spanish-language assessors. Notably, two additional programs described having enrolled children who speak Spanish at home. These programs did not assess children in Spanish, however, and therefore did not conduct refresher training for Spanish-language assessors. The lead trainer at one program said that all children spoke English fluently enough to be assessed in English, but she did not explain why the program had declined to use the Spanish-language assessment or how staff members determined fluency. At the other program, the director thought the lead trainer was assessing the Spanish-speaking children, but the lead trainer did not speak Spanish and had not received training on administering the Spanish-language version of the NRS.

Evaluating the amount of time that program staff spent in the Spanish-language training was a challenge. Programs with a large share of Spanish speakers, where all or nearly all assessors would conduct the NRS in both languages, typically conducted a single training that encompassed both the English- and Spanish-language preparation. Even if staff recalled the length of the Spanish portion of the training, certain activities (for example, reviewing the Assessor's Guide, a question and answer session, and role-playing) from the English section would also be applicable to the Spanish-language administration. Therefore, while several programs reported an additional half hour or hour to review the Spanish video and/or role-play in Spanish, this time is not comparable to spending an hour training Spanish-language only assessors who had not attended the English refresher training. In fact, OHS training guidelines instruct programs that Spanish-language training will take an additional hour beyond their English training time.

Approaches to the Spring 2007 Refresher Training for the Spanish-Language Assessment

	Percentage of Programs
Included all required elements Video, Q & A, role-play, and distributed guide	**32**
Included 3 of 4 required elements	**11**
Included 1 or 2 of 4 required elements	**7**
Self study or no refresher training	**42**
Self study	21
No refresher	21

N = 28 Head Start programs (note that percentages do not sum to 100 because two programs (seven percent) did not describe the activities included in the Spanish-language refresher training)

While 32 percent of programs that conducted Spanish-language assessments reported offering a comprehensive training to staff, a greater number (42 percent) of programs reported offering either no training or materials for self-study only (see Box on this page). However, this proportion is misleading. At one quarter of programs assessing children in Spanish this spring (including two of the six that did not offer Spanish assessment training), a sole assessor or the lead trainer was responsible for conducting all Spanish

assessments. A comprehensive training session—including role-play—would have been impossible at a program with only one bilingual assessor. The assessor attended the English training and then received Spanish assessment materials. At the other four programs, Spanish-language assessors attended the English-language training, but did not receive additional training on the Spanish-language assessment. Among the six programs that had Spanish-language assessors do a self-review before the NRS administration, one was a program at which English assessors also did self-reviews, two only had one Spanish-language assessor, and the remaining three trained all of their assessors together on the English assessment, and then encouraged the bilingual assessors to review the additional Spanish material on their own. In short, while a lower intensity of training for Spanish-speaking assessors will not necessarily result in bilingual assessors who are not familiar with the assessment practices nor reliable on the instrument, this lower intensity does run contrary to the OHS suggested training guidelines.

Program Approaches to Training New Assessors

For the spring 2007 round of NRS assessments, OHS told lead trainers that staff who were certified in fall 2006 should receive a refresher training, while new assessors should receive the complete, full-day fall 2006 certification training, followed by the spring refresher training. One quarter of programs had staff members who conducted assessments for the first time that spring. Nearly all of these programs (9 out of 10) provided some form of training and certification for new staff. At the tenth program, the new assessor was initially certified in 2004 but had never administered the NRS assessment; this person attended only the refresher training because the lead trainer did not think a previously certified assessor needed the new assessor training. Additionally, an assessor at an eleventh program had been conducting the assessment for several years but had never participated in a new assessor training and certification.

Among the nine programs that held training for new NRS assessors, staff described a range of approaches, from full-day training and certification with children to providing a refresher training only to certifying staff during role-play. At seven of the nine programs, all new assessors were certified after the local training; the remaining two programs had not yet held a training session for new assessors at the time of the site visit and are not included in the following analyses and discussion. These seven programs are classified into high, medium, and low intensities of training for new assessors:

- **High intensity.** Two programs offered a full-day training for new assessors and certified them using children who were not eligible for the NRS. At one program, the two new assessors were trained concurrently with thirteen experienced ones at a somewhat more in-depth refresher training (including extra time for additional rounds of role-playing) intended to accommodate assessors with and without previous NRS assessment experience. The other program trained two new assessors exclusively for a day and they then attended the two-hour refresher training with other staff.

- **Medium intensity.** The two programs in this category offered less than a full-day training requested by OHS, but somewhat more than the training they offered to experienced assessors during refresher training. One program trained a group of four new assessors for five hours, including the video, discussion, distribution of the Assessor's Guide, and several rounds of scripted and unscripted role-play. At the end of this training, assessors took the certification quiz. The second program trained three new assessors for a total of six hours: the first three hours ran concurrently with the refresher training for four experienced assessors, and the final three focused on additional role-play, feedback, and certification.

- **Low intensity.** Three programs provided minimal training for new assessors. One program included two new assessors in its hour-long refresher training, and then certified them afterwards based on role-play with more experienced assessors (rather than using the certification quiz, as recommended by OHS guidelines). At a second program, an education manager, who had conducted the fall 2006 assessments, took on the lead trainer role and convened a group meeting to watch the video, distribute the Assessor's Guide and discuss any questions. She then instructed staff to practice role-playing on their own time before doing the assessment with children, but staff were not ever certified.[49] Finally, at the third program, the lead trainer met with the new assessor to describe the NRS, review procedures, and then instructed her to attend the two-hour refresher training. This program provided conflicting information on the certification process. The assessor said she was certified with a child who was not eligible for the assessment, while the lead trainer recalled certifying the new assessor while she role-played with another staff member. Neither individual mentioned using the certification quiz.

Across all sampled programs that trained new assessors in spring 2007, lead trainers described that gesturing errors and straying from the script were common mistakes for new trainees. One program, which had provided high-intensity training, proudly reported that there were no mistakes among its new assessors, and that all were certified on their first try.

OBSERVATIONS OF SPRING 2007 TRAININGS AT FIVE SITES

In spring 2007, MPR staff attended six local training sessions at five Head Start programs to observe how staff implemented the NRS training model and to identify aspects of the training materials and procedures that might be improved. We observed one training session of new assessors and five refresher trainings for experienced staff; of these, four were for English-language assessors and one was for bilingual assessors. In total, we observed 4 new assessors being trained and 166 experienced assessors participating in

[49] The education manager, who eventually became the lead trainer, joined the staff in the fall after the last national ToT conference. Her program, therefore, had no certified trainer on staff, and was not successful in its efforts to hire a certified trainer from another grantee to train its staff.

refresher training.[50] Although this group of programs represents a small cross-section, we offer detailed observations of these sessions to illustrate good training practices, as well as examples of inadequate training.

Structure of the Trainings

The OHS refresher training guidelines for the spring 2007 assessments instructed lead NRS trainers to incorporate four core elements: (1) show the video, (2) lead a question-and-answer discussion about the video, (3) conduct at least one round of role-play, and (4) distribute the Assessor's Guide. The suggested time frame for completing these activities was approximately 1.5 hours. Following the refresher training on the English version, bilingual assessors were expected to watch and discuss the Spanish video and to role-play the assessment. The bilingual training was recommended to take an additional 30 minutes.

A seasoned lead NRS trainer led the six local trainings we observed, and most of these trainers had been certified at one of the initial ToT regional conferences in summer 2003; one individual was certified in summer 2005. Education coordinators or specialists in the Head Start program primarily served as trainers. Certified staff members assisted a few lead trainers (for example, to help facilitate the role-play), and at one program the program director co-led the training session. Half of the trainings were part of an in-service day, while others were "pull out" sessions. Three of the five programs conducted trainings at off-site locations, including a conference room at a not-for-profit organization, a banquet hall, and a nearby coffee shop.

Components Included in Observed Spring 2007 Local NRS Trainings

	Number of Training Sessions
Watched video	6
Discussed video	3
Received Assessor's Guide	6
Discussed Assessor's Guide	1
Role played in pairs or triads	3
Role played as one large group	2
Used role-play scripts	3
Used score sheets during role-play	2

N = 6 training sessions at five Head Start programs; we observed separate English and Spanish trainings at one program. Five sessions were refresher trainings, and one session was for new assessors.

The five programs where we observed spring 2007 refresher trainings required that all experienced assessors attend. In fact, one lead trainer followed a strict attendance policy, refusing to allow anyone who did not attend to conduct child assessments during that assessment round. The length of refresher trainings ranged from 50 minutes to about 4.5

[50] MPR did not randomly select programs at which we would observe local NRS trainings. Instead, during conversations with program directors to discuss the Quality Assurance Study site visits, we asked if staff would be willing to be observed at the local training as part of the visit. While it was not a random process, the identified sites included urban and rural programs, programs with bilingual assessors, programs that would only be training a few assessors, and programs with large groups of trainees. Since the NRS was in its fourth year of implementation, it was more difficult to find a program that needed to train and certify new assessors, which is why most observed trainings were refresher sessions.

hours, with most programs exceeding the OHS suggested minimum length of 1.5 hours. Lead trainers generally followed the suggested training agenda, except for one program that opted out of role-play and did not distribute or review the easels (see Box, previous page). Instead, staff concentrated on showing the video and distributing copies of the Assessor's Guide. The other four programs—one of which had separate English and Spanish sessions—screened the video, set aside time to discuss the video and easel, distributed blank answer sheets and an Assessor's Guide to each participant (including the Spanish version if applicable). Three programs enabled assessors to role-play among themselves with the spring 2007 easel for at least two rounds using the scripts and practicing with the score sheets. However, one program with separate English and Spanish sessions only reviewed the spring assessment once as a group with the lead trainer reading the script through once; no practice took place among assessors. None of the five observed refresher trainings included a discussion of the Assessor's Guide.

The program that trained new NRS assessors adhered to some of the guidelines outlined by OHS, but also omitted several important elements. For example, trainings for new assessors should last a full day, but this training lasted approximately five hours. Participants watched the English-language video, performed three rounds of role-play and completed the answer sheets, took the NRS quiz, and received and discussed the Assessor's Guide.[51] At the same time, group discussion of the Guide lasted only about 10 minutes, and they could not discuss changes in the easels from fall to spring since the new easels had not yet arrived (the lead trainer used the fall 2006 easels for training). Moreover, the lead trainer—who read the training script verbatim—did not carefully monitor the role-play exercise nor did she give much feedback during this process or conduct parallel scoring during practice. She also did not plan to conduct any follow-up observations of the new assessors.

Observed Problem Areas in Local Trainings

Site visitors observed four areas in which the staff had difficulties conducting the six local NRS trainings, including (1) ineffective role-play activities; (2) providing misinformation; (3) not referring to the Assessor's Guide as needed; and (4) not providing information for bilingual assessors.

Ineffective role-play activities. Role playing the child assessments with scripts enables both new and experienced assessors to become familiar with the latest version of the easel, raise specific points about the assessment procedures (from responses on the scripts), and have a group discussion about issues that arose during the exercise. Moreover, while the role-play occurs, the lead trainer should rotate to different pairs or triads to monitor the

[51] The Spanish-language video was not shown, nor did any activities incorporate the Spanish version of the assessment. The training session included one bilingual assessor (out of four total) who was selected to be certified because the program enrolled one Spanish-speaking child since last fall. However, the child had a severe disability, and the program was in the process of deciding whether the NRS assessment would be appropriate for this child. The Spanish-speaking assessor was not scheduled to conduct any English assessments.

activity, giving feedback and emphasizing any issues observed during the subsequent group discussion. However, only two of the five programs demonstrated effective role-play activities. In these cases, the trainers walked around the room, listened to each group, and gave feedback as needed. One lead trainer offered suggestions to pairs of assessors, and then reviewed some key points at the end of the exercise. For example, she reminded assessors that they should no longer point to the teddy bear in the Early Math section, even though this was the accepted protocol during the first year of the NRS.

In contrast, the other four training sessions either did not do role-playing at all, or did not do it effectively. One lead trainer explained that the program omitted the role-play exercise because the assessment team was experienced, so the program decided that staff did not need to spend much time on the refresher training. Instead of following a traditional role-play model in which assessors practice in small groups of two or three, one lead trainer read through the easel script with one assessor who played a 'perfect' child, while the other assessors listened without keeping score.[52] At another program, the lead trainer did not closely observe the practice groups. Assessors were more likely to correct each other than be corrected by the trainer.

Providing misinformation. According to a site visitor, one lead trainer seemed to be unfamiliar with the easel and the NRS training materials, despite clear instructions from OHS that all trainers should familiarize themselves with the materials before conducting the trainings. At a new-assessor training, she occasionally answered questions incorrectly. A management-level staff member who was also an experienced NRS assessor sometimes corrected the trainer, but sometimes assessors still received inaccurate guidance. For example, when asked if it was acceptable to "get a child going" on the Letter Naming exercise by pointing to the first letter plate and asking for the name of the letter, the lead trainer responded that assessors could use this strategy if the child seemed reluctant to start the task. Additionally, while she corrected assessors for misreading the script, she then inserted articles ("a" or "the") on the PPVT section.

Not referring to the Assessor's Guide as needed. One resource available to Head Start programs to provide clarification for specific child assessment administration procedures is the Assessor's Guide, which is updated each fall and spring. While OHS guidelines do not require that staff review the Guide as a group, they expect lead trainers to refer to it to verify the correct answer to a question raised during training. During one observed training session, an assessor asked if *trinche* would be an acceptable answer instead of *tenedor* to identify a fork (item BB8 in Exposición del Arte). Instead of referring to the spring 2007 Guide—which lists *trinche* as one of two acceptable alternate answers—the lead trainer gave two different answers, both of which were incorrect. First, she told assessors to score any regional Spanish term as correct, and then she instructed them to mark any answer

[52] The lead trainer in this program had to train 90 assessors for the English training and 40 assessors for the Spanish. She may have decided to take this approach because the groups were so large. For the Spanish session, she modeled the appropriate gestures instead of just reading through the script because, as she explained, the smaller group could more easily see her gestures.

aside from *tenedor* as wrong. The site visitor noted that this group received other misinformation during the refresher training.

Not providing bilingual training. The "Instructions for Conducting Spring 2007 Assessment Refresher Training" direct Head Start programs to hold a separate session for Spanish-speaking assessors and show the Spanish-language version of the video, discuss the video, and to role-play using the Spanish easel and role-play scripts. The English training should not serve as a substitute for the Spanish one. Nevertheless, one program of the five did not hold a separate training for its single bilingual assessor. The lead trainer reported that she was not aware until watching the spring 2007 video that there was a separate set of NRS training materials for bilingual assessors, who would need to be certified in Spanish. Therefore, the bilingual assessor was only trained and certified on the English-language assessment. The lead trainer reported that she would have liked information from OHS on how to train and certify this assessor. (This year was the first time since the NRS began that this program enrolled a Spanish-speaking child; as noted earlier, it was possible that this child would not be assessed due to a severe disability.)

Effective Training Components

Despite MPR's observation of some problem areas at local trainings, lead trainers carried out several components appropriately. Most lead trainers reviewed the score sheet with assessors, reminding them to record carefully the background information on each sheet (for example, child's identification number and date of assessment), and to fill out all bubbled responses. Trainers from programs with bilingual assessors adequately explained that children who have Spanish listed as their primary language on the Program Information Report (PIR) should first be assessed in Spanish, and should take the entire assessment in English if they make fewer than 15 errors on the language screener. One lead trainer reminded staff that they should sum and record the total correct answer, not the errors, a mistake she had noticed during the fall 2006 quality review process. (The language screener proved to be a source of confusion at several local trainings that MPR observed in fall 2004; see Paulsell et al. January 2005.)

In addition, a few lead trainers offered helpful reminders and tips to facilitate the child assessment process such as (1) make eye contact with the child so he/she will feel comfortable and de-emphasize the formality of reading a script; (2) review completed answer sheets with site supervisors to make sure that all background information is accurate; and (3) try placing the score sheet on a clipboard on the assessor's lap so that children do not get distracted by what the assessor is marking.

LOCAL PROGRAM RESPONSES TO TRAINING AND RESOURCE MATERIALS

The degree of a local program's buy-in to the value and careful administration of the NRS appears linked to staff impressions of the communications and materials that they receive from OHS about the assessment. Directors, lead trainers, and assessors from all 40 programs in this year's sample shared frank opinions of the resources they had seen, and often gave thoughtful suggestions for further improvement of these materials. Below we

summarize reactions to the webcast and satellite broadcast methods of disseminating information. A discussion of program opinions about the delivery of materials follows, along with feedback about the training videos OHS provided for front-line assessor training. Finally, we discuss reactions to and suggestions for the Assessor's Guide and staff impressions of any technical assistance that they requested or received during the 2006-2007 program year.

Satellite Broadcasts and Webcasts

Prior to the 2005-2006 program year, OHS hosted live satellite broadcasts on August 4 and October 27, 2004, and February 17, 2005, to provide updates on the NRS. These sessions 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 findings from the Quality Assurance Study. Satellite broadcasts allowed OHS the opportunity to reach all Head Start programs and, because the programs could call in or email 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, OHS replaced the live broadcast with a prerecorded webcast video that local programs could view at any time. A transcript of the video was also made available to programs. In the spring 2006 webcast, topics 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 OHS online over the course of the program year. This decision came about partly in response to feedback from programs during previous study site visits, which indicated that the broadcasts were too lengthy and that the live questions from local staff were sometimes repetitive and not very useful.

OHS used a webcast format again in spring 2007, this time broadcasting the presentation live rather than prerecording it. Topics included discussion of the most recent NRS results and an update on the Quality Assurance Study, information on how OHS uses national data to guide program improvement efforts, and an introduction to learning from child assessments. The spring 2007 webcast was approximately 90 minutes long, and also included program questions submitted by phone, fax, and email during the live broadcast. After the live broadcast, the script, slides, and some representative questions were posted online for future reference by programs.

Seventy percent of the programs MPR visited in spring 2007 reported that someone had watched a webcast or satellite broadcast in the past. However, only 28 percent of programs reported that someone on their staff had either watched the spring 2007 webcast or read the transcript. This is a substantial decrease from what programs reported during site visits in spring 2006, when 40 percent of programs we visited watched the spring webcast or read the transcript. At 10 percent of programs, we received conflicting information from staff when directors reported that the lead trainer had watched the most recent webcast, but the lead trainer reported not having watched it.

Staff from one quarter of programs described specific reasons for not watching the spring 2007 webcast—more than half (15 percent) attributed it to lack of awareness of the webcast itself or of its date of availability (although the transcript continued to be available online). This included one program whose lead trainer said that staff did not watch because of the "several hundred dollar" cost she associated with viewing it. Although these webcasts are free to programs, this program had watched previous satellite broadcasts in facilities rented from a local college, and may have confused the cost of those facilities with the cost of the broadcasts. Two programs said that technology problems, such as slow Internet connections, prevented staff from viewing the webcast. Two other programs decided not to watch it since they did not find the previous broadcasts to be useful.

Staff who watched the webcasts provided information on aspects they found to be most useful—an overview of the NRS, its results, and national data trends helped them refresh their memories about the assessment and increased the level of buy in and perceived legitimacy of the assessment. Programs also appreciated reminders of changes to the easel and the assessment process, and thought that the question-and-answer session was a helpful tool for learning more about the NRS. Staff at one program explained that the webcast introduced using the CBRS to produce comparative reports (for example, data comparing the program's NRS results with state averages), which would be a helpful program assessment tool. Finally, the two programs that read the transcript rather than watched the webcast appreciated this additional option. The transcript allowed them to skim the content for the most relevant information without having to wait for a large electronic file to run.

Program staff also shared some specific recommendations for how OHS could improve the content of the webcasts. One program suggested that the webcasts include information on who developed the cognitive assessments and SED rating form, how the tools were developed, and how OHS analyzes and evaluates the NRS assessment results. Another program would like the webcasts to walk through the changes made since the last round of assessments (the lead trainer making this suggestion acknowledged that she had not watched recent broadcasts or webcasts and did not know if this was now included), and staff at a third program wanted to see more concrete suggestions on how to administer the assessment to children with special needs. In addition, three programs (7.5 percent) would like some specific advice from webcasts on how to use assessment results, effectively compare their program with others, and how to use any observations they make when comparing performance across programs (in fact, this information was provided in the last webcast; see Chapter VI for more information on program requests for help with using data). Finally, a lead trainer at one program suggested that the materials in the webcast be made available in VHS or DVD format because she would use this portable format to share the contents of the webcasts with principals, teachers, and small groups of other stakeholders. Another lead trainer echoed this request, explaining that the limited technical capabilities of their program forced them to drive to and rent facilities from a local hotel to view these webcasts. A DVD format would be easier for them to use.

Training Materials

In spring 2007, OHS made a concerted effort to send out training and assessment materials to programs even sooner than they had previously done. As in prior years, the OHS publications office shipped materials to sites on a rolling basis (based on program location and closing date—between mid-March and mid-April 2007). Packages of completed assessment materials were due to OHS from all programs by the end of June. Unlike previous years, site visitors did not hear a large number of complaints from programs about the timing of the arrival of materials. In fact, staff from 13 percent of programs mentioned that material distribution had improved noticeably in spring 2007, and several assessors at one program remarked that they hoped to see this continue because it facilitated their training and assessment process.

However, the spring 2007 material delivery was not without problems. Five programs (13 percent) told us that their materials arrived later than they would have liked. Three of these (two with early closing dates, and another that primarily offered home-based services) remarked that the structure of their programs meant that they would have to receive materials much earlier than they did this year to be confident that they could finish assessments on time. A third program, whose assessment materials were late, called the regional office for assistance and discovered they had been sent to the incorrect address. Furthermore, three other programs explained that materials had been delivered on time, but that the wrong number of supplies had arrived. For example, one program with eleven assessors at three widely spaced centers received only one easel. The lead trainer at another program described receiving 200 copies of the training materials each year since the NRS began, despite repeated requests that OHS send only 20. It is not clear from these comments whether programs had responded to OHS requests to update their CBRS information on addresses, program closing times and quantity of assessment materials needed. In addition to impressions about the timing of training materials, some programs shared questions and concerns about the change in training and easel materials from one year to the next. Staff at four programs expressed frustration at the financial and physical waste that they perceived to be associated with repeated printing of very similar materials. One program suggested that OHS just send the materials that had changed and tell them what pages to replace. In addition, one lead trainer wondered about how best to dispose of the assessment materials from the previous year when the new training materials arrived (shredding materials versus recycling). Finally, one lead trainer requested that OHS occasionally send binders in addition to easel pages, because the original binders have started to fall apart.

Spring 2007 Training Videos

The spring 2007 refresher training video had similar content to those provided in previous years. It demonstrated a full assessment and described useful techniques for organizing, distributing, and collecting assessment materials. It also suggested methods for handling children's challenging behaviors, and reminded assessors of the basic NRS administration techniques. The video began by thanking programs for their help with

collecting data and noted that OHS had responded to staff feedback by shortening the training and focusing the video on changes to easel items and instructions.

Almost all of the sampled programs (98 percent) either included the video in their group training or made it available for assessors to review on their own (three chose to offer it for self-review). Staff at 54 percent of the programs showing the video had favorable opinions about the content and format. At seven programs (17.5 percent), assessors identified the video as the most helpful part of their refresher training experience. Assessors at other programs also shared favorable, though less enthusiastic, impressions. Assessors in nearly one third of programs (33 percent) remarked that the opportunity to watch a perfect assessment, observe how the model assessor handled different scenarios, and be informed of changes to the assessment all helped them better administer the assessment. Staff at 15 percent of programs remarked that this year's video represented an improvement in both content and quality of production over others they have seen in the past. Some themes about what made the video helpful emerged across groups at different programs (see Box this page). An assessor in one group shared that the best part of the video was seeing OHS staff thank local assessors for their hard work in collecting these data. The only program that did not show the spring 2007 video was a delegate agency; its grantee agency said that this spring training was not required because there were no new NRS assessors.

Impressions of and Suggestions for the Spring 2007 Training Video

- **Positive impressions**[1]
  - Has some examples of handling child behavior
  - Illustrates neutral encouragement
  - Details changes to assessment
  - Answers to frequently asked questions are helpful
- **Suggested additions**
  - Assessment with a non-cooperative or non-responsive child
  - Assessment of a child with developmental delays
  - Advice on assessing children who speak neither English nor Spanish

[1]This is a selection of the numerous positive responses from the 40 programs, ordered from most frequent to least frequent. Negative responses (n = 2) are listed in the main text.

While the overwhelming majority of feedback about the video from program staff was positive, two programs (five percent) criticized the training tool. At one program, staff thought it was too repetitive. Staff at another program said that the video was problematic because the sample child appeared stiff and rehearsed.

Assessors at 18 percent of programs offered suggestions for improving the video (see Box). Staff at these programs had generally positive impressions of the video, but also observed areas where it could be enhanced. These suggestions focused on additional testing scenarios that would show the sample assessor reacting to challenging situations, such as children with learning disabilities, behavior problems, or limited English skills. One lead trainer suggested that the video should highlight changes that have been made to the assessment since the previous round.

Assessor's Guide

Each year, OHS distributes a detailed Assessor's Guide to programs as part of the NRS training materials. The Guide describes general NRS administration procedures, gives an overview of the answer sheet, and walks through each of the assessment sections in detail to provide answers to common questions. It also contains some tips for managing child behavioral issues and administering the NRS to children with disabilities. The Guide is intended as a first source of information for assessors who are unsure about administration or scoring, aiming to answer common questions and thus improve the reliability of the test.

In spring 2007, nearly all programs (95 percent) distributed the Assessor's Guide to NRS assessors either for review during training or for self-review on the assessors' own time. Of the remaining two programs, one provided no spring training at all, and the other program did not provide information on whether staff reviewed the Guide or not. More programs distributed the Guide this year than in the past. In spring 2006, only 70 percent of the programs gave copies to assessors.

Staff at approximately one quarter (23 percent) of programs identified favorable components of the Assessor's Guide, and no programs criticized the content. One lead trainer reported that the number of questions substantially declined after the staff began receiving the Assessor's Guide. A second lead trainer echoed this experience, saying that there are never questions at training that she cannot answer by referring to the Guide. At a third program, where the lead trainer had distributed the Guide for self-review, assessors found it to be a helpful tool and suggested that future local trainings include additional time to review it as a group.

Technical Assistance

Aside from the information provided in the training packets, local staff can contact one of two telephone helplines with questions about NRS implementation—one for questions about the child assessment, and one for questions about the CBRS. Programs are informed about these telephone numbers and corresponding email addresses in the trainer's binder, in the Assessor's Guide, in the CBRS user manual, and on the OHS website. Most programs (63 percent) said they had requested technical assistance (from the helpline, regional office, or another source) at some point since the NRS began, but the programs gave varied reasons for requesting this assistance. Reasons cited by programs for requesting technical assistance included requests for additional materials, questions about the CBRS, and how to score unusual responses.

Slightly more than half (55 percent) of programs reported using the assessment helpline during the 2006-2007 program year or at some point in the past. Among programs that had not used the helpline, staff expressed awareness of that resource as an option if questions arise. Few programs discussed their use of the CBRS helpline, in part because the questions MPR asked of staff this spring did not focus on the CBRS. However, one-quarter of programs mentioned calling the CBRS helpline at some point. Two programs said they had called their regional ACF office to ask questions about the CBRS, but both said that program officers from the regional office referred them to the CBRS helpline.

With one exception, programs recounted favorable impressions of the helpline, describing their experiences as prompt, helpful, and successful in resolving their questions. The exception was a program that called the helpline hoping to discover why there were no spring 2007 materials available for new assessor training, requiring lead trainers to certify staff on fall materials and then refresh them on spring materials. While the helpline told this lead trainer that OHS does not publish spring certification materials, the lead trainer said she did not feel she ever received a truly satisfactory answer to her question. The only materials she then had available for training new assessors were those from her fall 2004 ToT session, because there were no certification forms to accompany spring 2007 materials, and she had already discarded her fall 2006 materials. She did not say whether she had requested additional fall 2006 materials for use in certifying new assessors.

SUMMARY

Comprehensive training that NRS assessors perceive as valuable may increase staff buy-in to the assessment, which, in turn, could contribute to the quality of collected data. To promote excellence in assessor training, OHS provides annually updated materials for each program to use when training its assessors.

At the local level, nearly all (93 percent) programs provided some form of refresher training to staff in spring 2007. Approximately 60 percent of programs provided comprehensive spring refresher training that incorporated the training video, the Assessor's Guide, role play, and a question and answer session. Another 28 percent of programs provided a training that incorporated three of these four elements, usually leaving out the role-play activity. Only one quarter of programs reported that their lead trainer did follow-up observations of some staff after the training. Nine out of ten programs that had new assessors in spring 2007 provided training to them. At 28 percent of the programs, staff felt that training twice per year was too much and suggested reducing the frequency to once annually. MPR site visitors observed that programs that did not train staff were significantly more likely to make administration errors than programs where staff had been trained. Similarly, at programs where staff had a comprehensive training, staff members were significantly less likely to make administration and scoring errors than at programs where the training had been less than comprehensive.

In spring 2007, 70 percent of programs assessed children in Spanish. A third of these offered comprehensive Spanish-language assessment training to staff, and 42 percent either encouraged self-review of the Spanish assessment materials or offered no additional training on Spanish assessment materials. Although these training figures appear to indicate a lower level of preparation for Spanish-language assessment staff, many of these staff attended a refresher training on the English assessment, or were at sites where a comprehensive Spanish-language refresher training was impossible to conduct with only a single Spanish-language assessor. While the level of refresher training provided to Spanish-language assessors is lower than recommended by OHS, these assessors had more preparation than the count of Spanish-language assessment training activities suggests because many of them also attended training activities for administering the assessment in English.

MPR observed the training of 170 assessors at five programs in spring 2007. An experienced and certified lead trainer led all observed training sessions, and the trainings included a video and distribution of the Assessor's Guide. Training at four of five programs also incorporated the role-play activity. The length of refresher trainings ranged from 50 minutes to about 4.5 hours. We observed four areas in which the staff had difficulties conducting the local NRS trainings, including (1) ineffective role-play activities; (2) providing misinformation; (3) not referring to the Assessor's Guide as needed; and (4) not providing training for bilingual assessors. In contrast, some observed trainings featured particularly effective elements, such as a review of proper procedures for completing the answer sheet and techniques to set children at ease.

In interviews and focus groups with program staff, we asked for reactions to and feedback on the NRS training materials that OHS provides. Seventy percent of the programs visited in spring 2007 reported that someone had watched a webcast or satellite broadcast in the past. However, fewer than half of these (28 percent of all programs) reported that someone on his or her staff had either watched the spring 2007 webcast or read the transcript. Staff who had watched the webcast made suggestions for improving it, and those who had not shared some recommendations for making it easier for them to watch in the future. Programs this year reported fewer problems with materials than in past rounds, and 12.5 percent of programs described an improvement in the timely and accurate delivery of assessment materials. Most program staff reported favorable impressions of the 2007 training video and Assessor's Guide, and of their experiences with requesting and receiving technical assistance.

CHAPTER IV

LOCAL APPROACHES TO IMPLEMENTING THE NRS

Several dimensions of NRS implementation of the cognitive child assessment were left to the discretion of each Head Start program, resulting in different local approaches to critical components of the system that could have implications for the administration of the NRS. For example, different standards for assessor training, coupled with differences in timing and location of assessments, could influence the quality of the completed assessments. A better understanding of these implementation differences might be useful to the Office of Head Start for developing further guidance to programs on how to administer the assessment.

In this chapter, we describe the approaches to the NRS implementation in the 40 programs we visited during spring 2007. Specifically, we discuss the programs' approaches to coordinating the cognitive child assessments; assigning staff to conduct the assessments; and communicating with parents, Policy Councils, tribal leaders/elders, and other stakeholders about the NRS. Where appropriate, we discuss the specific experiences of tribal programs in implementing the NRS. We also discuss the perceived and real costs incurred by local programs. Our discussion in this chapter focuses on the NRS cognitive assessment. In Chapter VII, we cover the approaches programs used to implement the Social Emotional Development (SED) rating form.

COORDINATING THE COGNITIVE 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 three coordination issues: (1) who should have overall responsibility for NRS implementation, (2) where assessments would be conducted, and (3) when and how to schedule assessments.

Staffing for NRS Training, Coordination, and Oversight

As with the programs visited in spring 2006, program directors tended to delegate overall coordination and supervision of the assessment activities to the lead NRS trainer. Only eight percent of program directors had lead responsibility for the coordination of NRS activities. In one program, the director played some role in coordinating and overseeing NRS activities, but other staff within this program maintained primary responsibility.

Lead NRS trainers had responsibility for conducting refresher and new assessor training, certifying new assessors to administer the NRS cognitive child assessment, and monitoring and tracking progress in completing assessments. In the sampled programs, all but one of the lead trainers were also responsible for overseeing implementation of the SED rating form. (See Chapter VII for additional details on SED implementation.)

Most lead trainers were either education coordinators or program managers (see Box this page). Similar to spring 2006, no programs hired an outside consultant to manage NRS activities, relying instead on program staff to coordinate these activities. However, in some instances programs relied on staff from their grantee office to supervise and manage NRS activities. All of these lead trainers were program specialists or "other" staff (in other words, not classroom teachers or assistant teachers).

Three of the sampled programs reported that they had experienced turnover in the lead NRS trainer position from spring 2006 to spring 2007. In all cases, turnover occurred because the staff member appointed as lead trainer left the program. When we visited these programs, there were no vacancies in the NRS trainer positions.

Programs' Lead Trainers

	Percentage of Programs
Education coordinator [a]	55
Program manager [b]	28
Program director	8
Teaching staff	5
Other	5

N = 40 Head Start programs

[a] "Education coordinator" includes education coordinators, managers, directors, and specialists.

[b] "Program manager" includes such positions as family support coordinator, disability specialist, and program coordinator.

Assessment Locations

All programs conducted the spring 2007 assessments in locations within 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, teachers' lounges, cafeterias, conference rooms, multipurpose rooms, parent involvement and resource rooms, and kitchens. Some programs reported difficulty identifying appropriate, quiet spaces for the assessments—fewer than one-quarter of the sampled programs reported conducting some or all assessments in more open spaces, such as hallways, large rooms with other children present, 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. 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 (five percent), other adults (four percent), other children (one percent), and other distractions (seven percent). Among Spanish-language assessments, site visitors reported distractions, including general noise (less than one percent), other children (two percent), and other distractions (five percent).

Scheduling the Assessments

All sampled programs reported pulling children out of their classrooms to conduct the spring 2007 assessments, and programs used three different approaches for scheduling them (see Box this page). The majority of programs (55 percent) left the scheduling of assessments to the discretion of the assessors or the staff at each center. In 23 percent of the programs, the lead NRS trainer, program director, lead NRS data manager, or assessors scheduled the assessments in advance. In some instances this meant that assessors who were not classroom teachers communicated with teachers to create an ideal schedule for assessments. In other instances, this meant that assessors created a schedule in advance for approval from the lead NRS trainer, or that program staff took it upon themselves to schedule assessments for assessors. In the remaining 23 percent of programs, assessors were instructed to complete all assessments within a specific window of time, usually during a two- to three-week period, although some programs completed all assessments in one or two prior-scheduled days.

Programs' Approaches to Scheduling NRS Assessments

	Percentage of Programs
Assessor or center staff discretion	55
Scheduled in advance by lead trainer, director, data manager, or assessors	23
Scheduled within a set window of time	23

N = 40 Head Start programs

One-third of the sampled 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. The remainder of the programs reported conducting assessments at any time throughout the day. Approximately two-thirds of the sampled programs reported assessing at least one child in both English and Spanish in spring 2007. Of those that described how they scheduled assessments for children who needed to be assessed in both languages, 25 percent reported administering each version of the assessment consecutively with each child—first Spanish, then English. In contrast, the remaining three-quarters reported that staff routinely scheduled the Spanish and English versions on different days. Staff in most of these 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. 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 indicated that they had

limited numbers of bilingual staff members and had to wait for Spanish-certified assessors to schedule and complete the assessments. Although this situation was rare, it may indicate the need for further training, or increased availability of bilingual assessors in some programs.

Three programs reported that they did not administer 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. They took this approach even though the Office of Head Start instructed all programs to administer both the Spanish and English versions to all children whose home language is Spanish. Staff in one of these programs reported not administering the Spanish version of the assessment to children from Spanish-speaking households, even though many of these children exhibited great difficulty in responding to and performing on the English assessment. Staff in another program reported assessing some children from Spanish-speaking households in Spanish and English in the fall, but only in English in the spring because of perceived increases in their English skills during the program year.

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[53] only, and (3) using a combination of teachers and other program staff, such as program managers and area specialists. Programs have shown a decrease in assigning teaching staff to administer the cognitive assessments, with 15 percent of programs visited in spring 2007 using teaching staff exclusively to administer the assessments (compared to 29 percent in spring 2006, 33 percent in fall 2004, and 43 percent in spring 2004) (see Box this page). In spring 2005, programs had relied less heavily on only teaching staff to complete assessments (11 percent). In the current round, slightly more than half of the programs used a combination of teaching and non-teaching staff for the NRS assessments, while almost one third of sampled programs used only non-teaching staff to administer the assessments.

Programs' Main Approaches to Staffing the Assessments

	Percentage of Programs
Teachers and other staff	55
No teachers/assistant teachers	30
Teachers/assistant teachers only	15

N = 40 Head Start programs

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 one 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. Three programs employed policies that prevented teachers from assessing the children in their own classrooms.

[53] Non-teaching staff members assigned to administer the assessments include education coordinators, program managers, program directors, and other management and grantee staff.

Two-thirds of the programs sampled reported conducting the Spanish-language assessment in spring 2007. Programs used similar types of staff to administer the Spanish- and English-language versions (see Spanish Language Assessors Box), although more programs used either teaching staff only or non-teaching staff only to conduct the Spanish assessment.

Spanish-Language Assessors

	Percentage of Programs
No teachers/assistant teachers	46
Teachers/assistant teachers only	29
Teachers and other staff	25

N = 28 Head Start programs conducting assessments in Spanish

The number of trained assessors across the 40 programs varied considerably by program size (see Average Number of Assessors Box). Small programs, defined as having 200 or fewer total enrolled children (60 percent of the sample programs), had between two and fourteen trained assessors, with an average of six assessors; large programs with more than 200 enrolled children (40 percent of programs) had between 10 and 128 trained assessors, with an average of 19 assessors. The average number across all 40 programs was 13 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 staff.

Average Number of Assessors

Large programs	19
Small programs	6
All programs	13

N = 40 Head Start programs

Note: Large programs had more than 200 enrolled children; small programs had 200 or fewer 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 106 assessors, with an average of 12 assessors. In addition, some programs reported that they often had few choices in selecting Spanish-speaking assessors because only a limited number of staff spoke Spanish. In fact, one program reported utilizing staff from another nearby program to assess Spanish-speaking children because they did not have in-house access to staff or other individuals who spoke Spanish. Although rare, such situations underscore the need for increased bilingual assessors/staff in some programs.

More than half of the programs experienced some turnover in assessors who were certified in Years 1, 2, and 3. Almost all changes were the result of staff turnover in the programs rather than changes in programs' approaches to staffing the NRS. Only two programs reported changes in their overall approach to staffing the NRS cognitive child assessment that affected the number of trained assessors. In these programs, due to competing demands, staff decided to reduce the number of assessors to only those who expressed an interest in continuing involvement with the NRS. As noted above, some programs also reported training additional assessors in Year 4 to increase the overall number of certified assessors.

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. Eighty-five percent of the sample programs assigned non-teaching staff to conduct some or all of the assessments. Many of these programs recognized some drawbacks to using non-teaching staff alone, such as that the children felt more comfortable and familiar with teachers; children responded better to teachers; and teachers had better knowledge of children's behavior. One rural program reported that because it serves a large geographic area and many of its centers are far apart, assigning teaching staff was the most feasible approach. Using these staff both reduced the need for travel and any burden on other types of staff. However, many programs determined that the advantages of using non-teaching staff outweighed the disadvantages. Programs that took this approach identified four primary reasons for not relying exclusively on teaching staff alone. First, the most common reason reported by programs was that they thought teachers might coach or inappropriately encourage children.[54] Second, programs reported that teachers were often busy with other responsibilities and did not have time to participate in training and to conduct the assessments. Third, programs reported conducting the NRS assessments would constitute a significant loss of instructional time and would disrupt children's learning. Finally, programs chose not to use teachers because of the added cost of hiring substitutes while teachers conducted the assessments.

The majority of programs (55 percent) used a combination of teaching and non-teaching staff to conduct NRS cognitive assessments. Programs generally reported a need to utilize the range of staff available to complete assessments in a timely manner. Therefore, most programs did not make a purposeful effort to utilize one type of staff over another; instead, they attempted to staff the assessments in a manner that would allow them to be completed as efficiently as possible.

COMMUNICATING WITH PARENTS, POLICY COUNCILS, TRIBAL LEADERS/ELDERS, AND OTHER STAKEHOLDERS ABOUT THE NRS

All of the sampled programs shared information with parents about the NRS—through parent newsletters, handbooks, letters home, and presentations at parent meetings, enrollment and orientation sessions, and Policy Council meetings. During focus groups, many assessors also reported informally discussing the NRS with parents. Sixty percent of programs reported sharing the NRS outcome report(s) for Year 4 with their Policy Council. Among the sampled tribal programs visited, staff also reported sharing information about the NRS with tribal leaders. About two-thirds of programs also reported sharing information about the NRS with other stakeholders, including school boards, members of the local community, and in some instances funding sources. Eight percent of the programs

[54] 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.

that shared information with other stakeholders suggested that their children's outcomes on the NRS provided a source of pride for the program and a tool for recognizing the accomplishments of children in Head Start.

Consistent with the sample of programs visited in spring 2006, the majority of programs we visited in spring 2007 required parents to provide written consent for their child's participation in the NRS. In spring 2005, two-thirds of the sampled programs required written consent, compared to three-quarters of the programs visited in spring 2006 and 2007. Of these, slightly less than half distributed the consent forms to parents at registration or enrollment; the rest were distributed at the beginning of the program year. In all of these programs, consent obtained at enrollment or at the beginning of the year covered both the fall and spring assessments; none of the programs indicated that they requested consent for the NRS again in the spring. More than 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 and Questions Raised by Parents, Policy Councils, and Other Stakeholders

- Parents cannot see their child's results; questions about child's performance
- Appropriateness of language of assessment for ELLs
- Confidentiality of results
- Questions about items and domains included in the assessment
- Questions about the purpose of the NRS
- Questions about how the NRS results will be used by the Head Start for program improvement
- Questions about how the NRS results will be used by OHS or others outside the program
- Specific NRS items perceived as culturally or developmentally inappropriate
- Over-testing in Head Start

N = 40 Head Start programs

Note: Items in this box are ordered randomly.

Parent Refusals

Less than one-quarter of the 26 programs that required written consent from parents reported at least one parent refusal in the 2006–2007 school year or preceding years. Most of these programs reported a small number of refusals (one to five per program). Some programs indicated that parents who refused NRS consent also refused other assessments and immunizations offered by the program, while other programs indicated that parent refusals were a result of concerns specific to the NRS. Programs visited in spring 2007 did not typically describe how they handled or responded to parent refusals.

Concerns Expressed by Parents, Policy Councils, Tribal Leaders/Elders, and Other Stakeholders

Although most programs (65 percent) indicated that parents, Policy Council members, and tribal leaders/elders had some concerns and questions about the NRS, most reported that staff members 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 and whether they could see the child's results. Parents of some children who were English language learners were concerned about the language the child would be assessed in and their child's performance in English. This was particularly a concern for parents of English language learners who spoke a language other than Spanish. Some parents expressed concern about the confidentiality of results and whether they would follow their child into kindergarten or affect their child's ability to enroll in kindergarten. These concerns were most acute among parents of children with disabilities or developmental delays. Other issues raised by parents and Policy Council members included concerns and questions about the items and domains included in the NRS, the purpose of the assessment, how the scores would be used, perceived bias and developmental inappropriateness of some items, and the quantity of assessments done in Head Start. Parents and Policy Council members at tribal programs did not typically have concerns specific to tribal populations. Instead, their questions and concerns were similar to those of other programs, as outlined above.

Staff typically responded by telling parents that the NRS was a federal requirement that provided the government with information about how this program and Head Start were doing as a whole. To alleviate parent concerns, programs also reported sharing overall program performance with parents and sharing the assessment easel and copies of score sheets and SED rating forms with parents and members of the Policy Council. Some assessors reported informally describing individual children's performance with parents. Staff at several programs indicated that questions and concerns about the NRS have diminished over time as parents have become accustomed to the idea of the NRS and the other assessments taking place in Head Start.

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, the Policy Council, and other stakeholders. In fact, in 15 percent of programs, staff indicated that they might experience difficulty addressing parent and other stakeholder concerns and questions because they themselves were unclear about the purpose, background, and use of the NRS. In some of these programs, staff desired more guidance and written materials from the Office of Head Start to answer parent questions. Some staff also indicated that they experienced difficulty addressing concerns and questions because they could not provide specific information on an individual child's performance to the parent.

COSTS OF IMPLEMENTING THE NRS

As in previous rounds of site visits, few Head Start directors in the sampled programs were able to estimate a monetary cost of implementing the NRS. In fact, in half of the sampled programs, directors reported that the program incurred no significant monetary costs in implementing the NRS because it was viewed as a component of staff members'

responsibilities or was covered by OHS funding.[55] Although half of the programs indicated that there was no significant monetary cost associated with NRS implementation, nearly 60 percent reported significant in-kind costs to the program (see Box).

In-Kind Costs of NRS Implementation Cited by Program Directors

	Percentage of Programs
Any cost	58
Staff time	35
Travel costs	25
Substitute teachers	10
Overtime pay	8
Hiring staff	5
Supplies	3

N = 40 Head Start programs

Note: Percentages do not add to 100 because respondents could cite more than one in-kind cost.

A major in-kind cost mentioned by more than one third of the program directors was staff time. Some felt that redirecting staff members—whether teachers or non-teaching staff—from their regular responsibilities placed a heavy time burden on them. Ten 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, hiring new staff to assist with the CBRS and data entry, and other supplies (e.g., pencils). One-quarter of the sample programs cited travel costs associated with NRS activities, which staff accrued when traveling to a central location for assessor training, and when traveling from center to center to conduct the assessments.

Only 13 percent of the sampled programs estimated the monetary cost of NRS implementation. Of these, estimated costs for Year 4 ranged from $150 to $42,000. 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.

SUMMARY

Overall, the 40 Head Start programs we visited in spring 2007 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 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 all but one program, the lead trainer was also responsible for overseeing implementation of the SED rating form (see Chapter VII).

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

[55] ACF reports that OHS provided a one percent increase in progams' base grants to offset some of the costs of the NRS.

NRS. On average, the programs trained 13 assessors. Consistent with findings from previous years, in spring 2007, 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. 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 assessing children.

Nearly all programs communicated with parents, Policy Councils, tribal leaders/elders, and other stakeholders 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. Similar to programs visited in spring 2006, three-quarters of programs sought to obtain parents' written consent for the NRS assessment. 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. These included parent requests to see their child's individual results, concerns about bias and appropriateness of specific items and the language of the assessment, and questions about the purpose of the NRS and how the results would be used. Other stakeholders also had questions about how the results would be used to improve program performance.

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

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 provided 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 could 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 could be used in concert with other data sources to help guide programs in their quality improvement efforts.

In October 2006, the Administration for Children and Families distributed by mail the third NRS growth reports, covering the 2005-2006 program year, along with a report on spring 2006 assessment outcomes. The first set of interim reports had been made available online in July 2006, with a second set (including national comparison information) online in August. 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 2005 and spring 2006. Accompanying materials offered guidelines and suggestions for reviewing the reports and using them for local program improvement efforts. Our spring 2007 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 2007, as we had in spring 2006, we also obtained more detail about programs' use of local child assessments and how programs felt they compared with the NRS. (Programs' reports on proposed use of the new Social-Emotional Development Rating forms are discussed in Chapter VII.)

This chapter begins with a discussion of the reactions of program staff to the 2005-2006 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 2005-2006 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 2006 assessments, as well as the 2005-2006 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 2005-2006 Growth Report (dated October 2006) was divided into three main sections (see a sample program report in Appendix D). The first section provided programs with the HSNRS Spring 2006 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 2005-2006 program year, regardless of whether or not they had been assessed in fall 2005. These tables were presented separately for children assessed in English and in Spanish. For each of the four skill areas (English 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 third section of the 2005-2006 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 previous section, "T" bars superimposed on each bar in the graph showed national averages.

Finally, the reports provided tables that offered two additional sets of comparisons. In the first, "Spring 2006 Assessments," data provided a description of the group of children on whom the Spring 2006 HSNRS Child Assessment Scores were based, including number kindergarten-eligible, number of valid English and Spanish assessments, and number of children with disabilities. The final table, "Program Profile," provided basic information about program characteristics, as well as about the children on whom the Spring 2006 HSNRS Child Assessment Scores are based. Program characteristics have been used to assign programs to a set of categories and make it easier for programs to compare the performance of their children with that of children in similar Head Start programs. For example, the percentage of assessed racial/ethnic minority children was provided, and rated as low, moderate, or high relative to the national average of programs. Similarly, the percentage of children who have identified disabilities and of children who are English language learners was provided and rated as low, moderate or high. With this information, programs could log on to the CBRS website and compare their performance with that of programs serving similar populations.

Below, we describe the reactions of local program staff to the 2005-2006 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. (Some staff also included in their comments reactions to the fall 2006 baseline report.)

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 2005-2006 Growth Report were higher, lower, or about what they expected prior to receiving the report. Staff from 70 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). Often the results corresponded with results form local assessments or with NRS reports from previous years.

Programs' Perspectives on Growth Report Results

	Percentage of Programs
In line with expectations	70
Lower than expected	20
Higher than expected	5
Report not reviewed	5

N = 40 Head Start programs

In another 20 percent of programs, staff agreed that results were lower than expected either on one domain, for a specific subgroup of children, or overall. For example, staff in two programs commented that scores were lower than the previous fall's baseline or showed less growth than in the previous year. Another program cited lower scores for Spanish speakers than in the past. In 5 percent of the sample programs, scores were higher than staff had expected. Two programs reported that they had not reviewed the growth reports.

Usefulness of the NRS Reports for Local Programs

The degree to which programs found the reports useful and relevant for program improvement efforts was an important factor for the Office of Head Start to consider as it planned to move forward with NRS system improvement efforts. The views of local program staff were considered an indicator of how likely programs would be 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.

This year's interviews revealed that 82 percent of the programs viewed the reports as clear and easy to understand. This judgment compares favorably to 2006, when 75 percent of programs found the reports user-friendly, and to 2005, when less than half of programs did. Staff members commented on how much they appreciated the bar charts and graphics. Two programs specifically reported that they were able to understand the reports better than in past years. It was not clear whether this improved understanding was due to enhancements in the reports, or to increased staff familiarity with the formats. No programs requested technical assistance to interpret the results.

Despite efforts to improve timeliness, 30 percent of programs thought that the delayed timing of receiving the growth report—interim reports in summer and the final report in the fall of the 2006-2007 program year—diminished its usefulness. However, one program specifically praised the improved timeliness of report delivery.

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. Fifty-three percent of programs specified center or classroom level, while 23 percent requested data at the individual child level. Staff from 67 percent of the programs offered a range of other suggestions to enhance the way in which outcomes are presented in the growth report. Program recommendations on the format of the report included: (1) providing reports in electronic format so data can be reproduced (2 programs) and (2) providing reports in black and white text to make photocopying easier (1 program). Program recommendations on the content of the reports included: (1) providing results at the child, classroom, or center level; (2) providing results for specific skills within domains such as counting, addition, and subtraction (1 program); and (3) comparing results to assessment results of other kindergarten eligible, non-Head Start children (1 program).

HOW PROGRAMS HAVE USED THE GROWTH REPORT

In the 2005-2006 Growth Report, the Office of Head Start 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 or found the reports to be presented at a level that was not useful, 95 percent of programs identified at least one way in which 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

Nearly all of the programs shared the NRS results with staff beyond the directors and lead trainers who were directly responsible for overseeing NRS implementation (see Box). Three-quarters distributed results across various levels of program staff, including education specialists, center managers, NRS assessors, teachers or home visitors. Most programs shared the 2005-2006 Growth Report with stakeholders beyond local staff. Seventy-seven percent of the programs presented the results to their Policy Councils, and 10 percent shared the results with parents (compared with 43 percent in spring 2006). In addition, 55 percent presented NRS results to a board of directors or a tribal council, up from 45 percent in spring 2006, and two programs reported using the report results when applying for grants and other sources of outside funding. An additional two programs reported that while they may share the report in the future, they had not yet.

How Programs Shared NRS Results

	Percentage of Programs
Shared NRS results beyond director and/or lead trainer	95
Shared NRS results with	
Management team	90
Policy Council	77
Program staff	75
Board of directors or equivalent	55
Parents	10
Grant funding	5

N = 40 Head Start programs

Considering Changes to Classroom Practice

Programs reported using NRS growth report results to change program practices in such areas as curriculum, program self-assessment, and staff development. In making these changes, they usually referred to other sources of information as well, such as local assessment results. Programs reported that they also used NRS report results to gain a national perspective on the performance of children in their programs.

Sixty-five percent of programs reported using reports to inform changes to classroom practices and to make curriculum choices. Programs reported using results to enhance their

curriculum or to reconsider their chosen curriculum, in light of weaknesses in child performance identified by the NRS reports. All programs that identified using NRS reports to inform such changes also used information from local assessments when making these decisions. Some changes to classroom practice were based on the content areas emphasized by the child cognitive assessments or on the observations of assessors, rather than on the content of the baseline and growth reports themselves (see below for more information on specific changes to classroom practice).

In addition, forty-two percent of programs reported using reports as a component of their broader program self-assessment process. Again, the NRS reports were one of several sources that programs considered during this process.

Considering Changes in Staff Development

Forty-two percent of programs reported using NRS reports to identify training needs for staff. Programs reported offering training to staff on methods for providing instruction in literacy, print awareness, and early math. Once again, all programs that used NRS reports also cited using local assessment results when determining staff development needs.

Reasons for Not Using NRS Results

Twenty-two percent of programs (9 programs) reported that they have not used NRS growth or baseline reports. Their main reasons for not using the reports were: (1) data reported at a level that is too broad for program planning (2 programs); (2) programs prefer local assessment results (1 program); (3) timing of reports makes them difficult to use (1 program); and (4) programs question the validity of the child cognitive assessment (2 programs). Three programs did not provide reasons for not using the results.

Perspectives of Tribal Head Start Programs

Of the five tribal programs visited, four reported that they have not used the data, and one program reported only limited use of the data. Programs reported that the reports provided no new information and that they considered their local assessments more valid and useful. However, when asked whether tribal programs had any concerns about using child assessment data, four of the five programs reported no concerns that they felt would be specific only to tribal programs. One program specified that they did not use the results because of perceived language and cultural biases. Specifically, they reported that many children at their program speak a language other than English or Spanish; they are concerned that the NRS child cognitive assessment administration guidelines do not allow responses in the children's home language. Staff in this tribal program also mentioned having concerns about the cultural appropriateness of some vocabulary items.

AVAILABILITY AND USE OF LOCAL ASSESSMENT RESULTS

Since 2000, Head Start programs have been charged with carrying 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 second consecutive year, 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, almost a third of the programs used Creative Curriculum's assessment tool, which is completed three times a year, and another quarter used the Ages and Stages Questionnaire (usually considered a screening tool), with locally designed assessments and the High/Scope COR being mentioned by 20 and 17 percent of programs, respectively. Local assessments were carried out by teachers and teaching assistants.

Staff reported that they were generally very satisfied with their local assessment tools—67 percent expressed satisfaction, while another quarter of programs were generally satisfied with some concerns.

Comparing NRS Reports with Other Assessment Results

As part of their reviews of the 2005-2006 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. Fifty–seven percent of 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. They reported using NRS reports as further evidence that a domain was a specific weakness of the program. Others used NRS reports to flag areas of weakness and then carefully examined their local assessment results in that area. Some programs reported using the NRS results to help them assess the accuracy and reliability of their local assessment tool.

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. However, some programs noted some differences due to the fact that their local assessment results included all 3- and 4-year-olds in the program, not just the children who are kindergarten-eligible for the following year. Programs that did not compare results often reported that the NRS child cognitive assessment and their own local assessments were too different to make comparisons. One lead trainer commented: "The local assessment captures everything. Anything that the NRS provides is just bonus information. Most often it reiterates what the local assessments tell us."

Making Comparisons across Multiple Years

Fifty-seven percent of programs reported comparing NRS reports across multiple years. Many programs reported that their scores either remained consistent over time (15 percent) or improved (25 percent). One program described the incremental improvements the program had made over the course of three years. Other programs reported that although

they compared results across years, they did not think the comparisons were worthwhile because each year the program results report on a new cohort of children.

Table V.1. Developmental Screeners and Local Assessments

	Number of Programs[a]	Percent of Programs[a]
Developmental Screener		
Brigance	9	22
DECA	7	17
DIAL 3	7	17
LAP-D	4	10
ESI	3	7
Denver II	2	5
Other	13	32
Local Assessment		
Creative Curriculum	12	30
Ages & Stages Questionnaire	10	25
Locally Designed	8	20
HighScope COR	7	17
Galileo	3	7
LAP	2	5
Other	10	25
Conducted by:		
Teaching Staff	37	92
Reviews Results		
Teaching Staff	37	92
Management Team	37	92
Policy Council	21	52
Parents	20	50
Board	12	30
School District Staff	4	10
Tribal Leaders	3	7
Other Stakeholders	4	10
Uses of Local Assessment Results		
Target Staff Training and TA	31	77
Inform Changes to Classroom Practices	31	77
Program Planning, Assessment & Improvement	24	60
Individualization for Children	21	52
Satisfaction with Local Assessment:		
Satisfied	27	67
Satisfied but with some concerns	10	25
Not Satisfied	1	2

Source: Head Start National Reporting System Quality Assurance and System Development Project, Year 4 Spring 2007 data collection, interviews with program directors and lead trainers

Note: N=40 programs. Note that not all programs provided information.

[a] Does not total 40 because some programs provided more than one response.

Making Comparisons with National Data

Seventeen percent of programs reported using reports to gain a national perspective of their performance as compared with Head Start programs across the country. Programs reported that while much of the information on program strengths and weaknesses was not new to them, having national data available added an important dimension that they could not obtain from local assessments.

Changes in Classroom Practices as a Result of the NRS

	Percentage of Programs
Any changes to classroom practices	65
Increased focus on:	
Literacy	27
Math	12
Materials purchased	
Books, literacy materials	10
New curriculum	12
Encouraged parents to supplement classroom learning at home	12

N = 40 Head Start programs

CHANGES MADE TO CLASSROOM PRACTICES

Sixty-five percent of Head Start programs reported modifying their classroom practices or making curriculum choices in response to the NRS (see Box this page; some programs indicated more than one change). For the most part, programs have placed greater emphasis on skill areas included on the NRS assessment during classroom instruction. Over a quarter of the programs mentioned focusing more attention on cultivating language and literacy skills, such as alphabet knowledge, phonemic awareness, building children's vocabulary, and reading to children. A smaller subset of program staff reported expanding the type of vocabulary they used to include the vocabulary words on that section of the cognitive assessment. Twelve percent of programs also have placed greater importance both on developing math skills, for example, introducing children to graphing, addition, and subtraction. Another twelve percent of programs reported purchasing a new curriculum in response to low performance on the NRS cognitive assessment and their own local assessment. Ten percent of programs have used the NRS as a guide for purchasing classroom materials. Books, letter stencils, and literacy tool kits to support literacy development were most common. To supplement the curricular changes taking place at their Head Start centers, 12 percent of programs asked parents to become more involved in nurturing their child's literacy and numeracy skills outside of the classroom. These activities included developing an onsite lending library, developing literacy tool kits with activities for parents to use with children, and introducing a daily message about literacy for children and parents.

In addition to program-wide changes to classroom practice and curricula, staff in thirty-seven percent of programs (15 programs) reported using information about the performance of individual children on the NRS to inform classroom practice. Programs that assigned teachers to assess their own children were more likely to report this use of the data than programs that used no teachers to assess children (42 percent of the programs that used some or all teachers as assessors, compared with 28 percent of the programs that used no teachers). Frequently, these changes were not program-wide initiatives, but were changes instituted at the classroom or center level, which could be approximated through aggregation

of individual child-level data The Office of Head Start has consistently indicated that use of individual child-level assessment data is not recommended.

How Programs Plan to Use the NRS Results

	Percentage of Programs
Intend to use NRS results	57
No intention of using NRS results	23
Data unavailable	20
Specific plans:	
Provide self assessment	25
Compare to local assessment results	15
Present to stakeholders	7
Inform changes in practice	7
Other	3

N = 40 Head Start programs

FUTURE PLANS FOR USING THE NRS REPORTS

While visiting Head Start programs in spring 2007, 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). Programs reported that they plan to use results (1) to assess program improvement and overall performance; (2) to compare to local assessment results: (3) to present to stakeholders and include in grant applications; and (4) to inform changes to classroom practices and staff training. Another 9 programs said they had no plans to use the NRS, and 8 programs provided no information.

SUMMARY

The majority of Head Start staff who participated in the site visits found the 2005-2006 Growth Report easy to understand. Program 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). A majority of the 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. Some programs reported purchasing a new curriculum. Others mentioned enhancing parent involvement to support classroom activities. Because the NRS reports did not provide center-, classroom- or child-level information, some programs reported tracking the item responses of individual children; this practice was more likely in programs which used teachers as assessors than those which did not. Programs mentioned using NRS data to make comparisons with national data or with data on comparable programs, and to make comparisons of their data across multiple years.

With regard to using the NRS results in the future, more than half of the 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 intended to continue incorporating NRS outcomes as one source of information for program-planning efforts.

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 fourth 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. Staff perspectives also can 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 2007 site visits. We then describe their overarching concerns and the implications these have for future directions that Head Start might take. Next, we articulate the main suggestions of program staff for improving communication and planning, and improving specific aspects of the assessment and its procedures. Finally, we highlight the contributions, concerns, and suggestions identified by staff in the tribal programs we visited. In some cases, although comments may come from only a small number of programs, we have included them here to provide illustrations of the perspectives of program staff.

CONTRIBUTIONS OF THE NRS

When asked to list the contributions the NRS had made to their programs, staff members at programs described a number of benefits of the NRS. Primarily, the staff-defined contributions were that the NRS (1) increases program accountability by providing national comparisons, (2) helps to improve classroom practice and inform teacher training, (3) provides measures of children's baseline or growth in knowledge, (4) demonstrates Head Start's effectiveness, and (5) raises staff awareness about assessments and/or OHS

expectations (see Box, next page). A sizable share of programs also reported that the NRS helped them validate local assessment results.

Most programs (55 percent) reported that the NRS increased program accountability, and comparisons to other programs were a source of motivation to perform better. Others felt that growth data can or does show that Head Start children are learning and that the program is working. One director said that the NRS lends credibility to Head Start at the national level, and helps to combat the misconception that the program is "just day care for poor kids." Program staff also saw value in the ability to compare their own performance to that of other programs serving the same population, one of the key goals of the initiative. Staff at a large share of programs (43 percent) credited the NRS with helping them to improve instructional practice in the classroom by showing areas of strength and weakness. For example, some programs specifically mentioned they had increased emphasis on letter naming (many programs said their curriculum had previously focused on letter sounds but not letter naming), number recognition, and more difficult and varied vocabulary. However, this also represents a decrease in perceived contributions to classroom practice from spring 2006, when nearly two thirds of programs told us that the NRS made contributions in the classroom. This decrease could be attributed to programs already having made changes to classroom practices. Staff reactions to NRS reports and results for their respective programs are described in detail in Chapter VI.

Contributions of the NRS

	Percentage of Programs
Accountability and national comparability	55
Helped improve classroom practice	43
Measured what children know	33
Proved Head Start works	28
Raised staff awareness of assessments and expectations	28
No contributions	33

N = 40 Head Start programs

Approximately equivalent proportions of programs noted contributions in three additional areas: they liked that the NRS (1) provides measures of what children know (33 percent), (2) "proves" that Head Start works (28 percent), and (3) raises staff awareness about assessments and OHS expectations (28 percent). Staff identifying the positive role of the NRS as an additional development measure were pleased to note child progress, and glad for a tool (in the form of assessment reports) that they could share with parents, the Policy Council, governing board, and other stakeholders to point to program effectiveness. Some programs reporting that the NRS proves the efficacy of Head Start indicated they saw NRS results as a vehicle to help raise funds for the program from the community. In some other programs staff interpreted fall to spring growth as "proving" Head Start is effective, thus boosting staff confidence in the program. When discussing the contribution of the NRS to illustrating OHS expectations and raising staff awareness, many programs noted that the NRS does a good job of alerting staff to the importance of assessment. Staff from other

programs stated that the NRS helped them understand what the federal government's priorities are and that a national assessment raised awareness about the program.

A smaller portion of programs had staff reporting NRS contributions in other areas. In about one-fifth of programs, staff said they enjoyed and even looked forward to the opportunity for one-on-one time with the children (20 percent). Others appreciated the NRS providing new information on children that was not collected otherwise (18 percent), and that it provided a useful opportunity for children to experience a testing environment that would prepare them for expectations in kindergarten (13 percent).

Staff in about one-third of programs in spring 2007 said it had made no contribution (in spring 2006, 34 percent of programs said the NRS had made no contribution). At 15 percent of programs in spring 2007, all staff that we interviewed reported that the NRS made no contribution to their program. At another 18 percent of programs, some staff thought the NRS made no contribution while other program staff thought that the assessment contributed additional data, raised teacher awareness about instruction, or contributed to the program in another way.

CONCERNS ABOUT THE NATIONAL REPORTING SYSTEM

During site visits, we asked respondents the open-ended question, "What are your strongest concerns about the NRS?" Responses clustered around four primary issues: (1) lack of clarity about the purpose of the NRS and how the results will be used, (2) accuracy with which results reflect children's abilities and programs' performance, (3) concerns about Spanish-language assessments, and (4) the staff time and financial resources devoted to the NRS. These concerns are similar to those reported during previous rounds of site visits, except that programs rarely mentioned interpretation of NRS results as a concern in this round. It was clear, however, that program staff varied widely in their familiarity with child assessment and many could benefit from more information. Details of their concerns are described below.

Purpose of the NRS and Use of Assessment Results

In more than one-third of the programs, staff expressed concerns about the purpose of the NRS (see Box this page). It is notable that this attitude in spring 2007 is about half as prevalent as in 2006. Thirty percent of programs did not understand how to use the program-level NRS results they receive. Lack of understanding of the NRS' purpose led some program staff to voice their fears about the ways the data might be used at the national level and what they saw as

Concerns About Purpose and Use of Results

	Percentage of Programs
Do not understand purpose	35
Unclear about how to locally use data presented at program-wide level	30
NRS growth scores may affect future funding	22

N = 40 Head Start programs

potential consequences for programs if children did not perform well on the assessment. Uncertainty about the purpose of the NRS also appeared to fuel other concerns. For example, in the absence of understanding the purpose of the NRS, staff in about one-fifth of programs speculated that growth scores could be used for future program level funding decisions, with under-performing programs defunded and shut down. Some program staff speculated that the true purpose of the NRS was to harm, undermine, discredit, or dismantle Head Start. In a few programs, staff speculations concerned federal use of the NRS to evaluate individual teachers or even to track outcomes for individual students.

Accuracy of Portrayal of Program Performance or Child Abilities

Perhaps in part because of their concerns about the consequences of the NRS results for their programs, most local programs (88 percent) raised a number of concerns about whether the NRS is a reliable tool for measuring Head Start program performance. Many of the concerns staff raised concerns about the validity of the NRS (measuring what each section is supposed to measure). In this section we make special note of issues with items that may affect validity differently 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.

Staff in half of the programs expressed concern that a standardized test such as the NRS was not a developmentally appropriate way to assess preschool-age children. They were particularly concerned about the format of NRS administration, because it is a direct assessment in which the child is removed from the classroom and possibly assessed by an adult they do not know. Rather, they felt observation over time would be a better choice for children in this age group. Many staff were concerned that preschool children were not accustomed to sitting for assessments or listening to what staff characterized as inappropriately complex instructions.

Concerns About the Accuracy of Portrayal of Program Performance or Child Abilities

	Percentage of Programs
Any concern about accuracy	88
Standardized testing not developmentally appropriate for prekindergarteners	50
Scripting too formal, stiff, or unnatural	35
Vocabulary items contain cultural, regional, or socioeconomic bias	35
Validity of assessment/item development	28
Fall to spring comparisons not valid and/or confusing	23
Data otherwise unreliable	5

Staff also reported uncertainty about how much flexibility they had to depart from the script and redirect children when necessary. (Chapters II and III present detailed discussions of staff members' experiences administering the NRS and their requests for additional training). Related to standardization procedures, staff in more than one-third of programs reported that the scripting felt "stiff," "unnatural," and "rigid" and that this was confusing to children who were used to interacting with adults in a much warmer manner. Some staff

members asserted that children's performance would be negatively affected in such a setting. Additional information or training to clarify that assessors can use a warm tone while staying on script may be useful for administering this or other standardized assessments.

In more than one-quarter (28 percent) of programs staff worried that the NRS was not valid either because implementation varied too much across sites, or they were lacking information about how items were developed to ensure that comparisons were valid. At one program, the lead trainer expressed concern about whether it was valid to compare children assessed by different staff (inter-rater reliability).[56] At another program, an assessor was uncertain whether their children, who had taken a somewhat similar school district assessment in preparation for kindergarten the day before, might have exhibited some practice effects. She said that additional guidance from OHS about the timing of the NRS around other similar local assessments could improve cross-site validity. Finally, at several programs staff members were concerned or confused about how the NRS was developed and therefore had reservations about the validity of assessment items. Further information about basic testing and assessment procedures might help to alleviate some staff concerns.

Local program staff also expressed other concerns about what they termed the validity of NRS data. Again, in some cases, these concerns reflected lack of information about valid assessment practices, which may be addressed with additional training or information. For example, staff in nearly one quarter (23 percent) of the programs felt that the NRS could not validly measure growth from fall to spring since the assessment battery was changed between those administrations. A substantial number of programs also worried that the NRS does not accurately measure child knowledge, either because it represents only two short snapshots or because they believe that children become nervous or frightened during assessments. Finally, staff at two programs stated that they do not believe that the results they receive in their NRS report reflect the score sheets they submitted, and that their reports have fewer children than they had assessed.[57]

Staff in most programs had concerns about items in specific portions of the assessment. Mainly, these concerns concentrated on individual items in the receptive vocabulary section (PPVT), with nearly 90 percent of programs noting concerns with those items. In some instances, staff noted that areas of concern had improved over time (especially the Early Math section). We characterize concerns about the receptive vocabulary portion of the NRS

[56] Notably, concern about the effect of variation in administration on the validity of assessment results is at odds with staff concerns that the NRS is too scripted and uncomfortable to administer. While staff felt frustrated by the highly-scripted assessment, they also saw the importance of collecting data that are comparable across sites.

[57] At one program, whose overall reaction to the NRS was fairly negative, the lead trainer said that she is tempted to keep copies of the score sheets for later comparison with the data that comes back to the program, but has not done so. Another program's director reported an ongoing problem with receiving a report based on far fewer (80 to 250 missing in a program with nearly 1,000 assessments) children than the number for whom they submitted reports, without receiving any information about why those other score sheets were rejected from the Scantron process. ACF reports that scoring sheets are sometimes rejected because of incorrect identification numbers or errors in marking responses on the sheet.

as (1) concerning perceived bias against particular subgroups, and (2) general concerns about items.

Staff in more than one quarter of programs (28 percent) mentioned concerns about bias in receptive vocabulary (biases reported were against rural, urban, nonwhite, and those from low-income families). For example, staff in some programs indicated that rural children would not be familiar with "fountain." Program staff also noted that the types of fences and cows in some areas of the country do not look like the ones pictured on the PPVT. Other items raising differential concerns for program staff included: sanding, knight, and selecting. Additional item level concerns are described in detail in Chapter II.

As described in Chapter II, many programs had general concerns about the vocabulary section, which were not likely to affect subgroups differently. Most general concerns about the section came from staff who felt the items were too difficult or unfamiliar to children, the section too long, the pictures poorly rendered, or the distracters on the plates unfairly similar. In total, staff in nearly three quarters (73 percent) of programs described their concerns about the PPVT section of the test as being universal to all children in this age and/or sociodemographic group. Twelve separate PPVT items raised staff concerns, with the most common items being knight, sanding, surprised, and vase. Staff in many programs said that it was difficult to keep children engaged during the vocabulary section, due to its length and because the pictures, as black-and-white line drawings, were not visually interesting or engaging. Staff suggested simpler pictures, color drawings, photographs, and fewer items as a way to improve the PPVT. Other concerns ranged from words in general being too difficult for children of this age group to distracters on the plates that assessors felt were also valid answers.

Other sections of the NRS concerned staff as well, including Letter Naming and Early Math. The primary concerns in Letter Naming 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. Assessors at several grantees suggested that they be permitted to prompt children for letters they had skipped so that the assessment would more correctly assess the number of letters a child could name. Staff also disliked the focus on letter naming, rather than asking children if they knew letter sounds. This concern was pronounced at programs where staff reported that classroom practices focus on letter sounds rather than letter naming; these programs believed that the design of the NRS letter naming section does not allow their children to demonstrate their letter knowledge. The form or content of the Early Math section was the source of concern for a small number of programs (see Chapter II). These programs felt that concepts such as subtraction and reading graphs were too advanced for this age group and that not offering children manipulatives made it difficult for them to formulate correct answers.

Concerns About Spanish-Language Assessments

In programs that conducted Spanish-language assessments (29 of 40 programs)[58], 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) other scripting differences between the Spanish and English versions of the NRS. In the discussion below we note where the concerns raised relate to the validity of the assessment in Spanish. For concerns that make the test difficult to administer 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 for the English assessment: staff stated 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. For example, the scripted word for knock (as in "knock on the table") is golpea, which means hit, beat, punch. Staff reported that children became confused because they are always instructed not to hit and here the word is used out of context. According to assessors the equivalent phrasing of "knock on the door" would be "toca la puerta," translated as "touch the door." In another Simon Says example, assessors said some 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." Assessors at one site pointed out that "toma" could be translated as "to drink," which confuses children who are instructed to "toma el papel" ("pick up the paper"). Although these comments about dialect differences were usually made by just a few staff, they may be indicative of a need for further revision of the instrument, or training or guidance of assessors.

Acceptable word alternatives and the wording of the instructions were also of concern. 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), that in some dialects would be "cubierto." Again, to the extent that children's understanding of instructions is regional (that is, if Spanish speaking children from different regions use different dialects) then these concerns could have implications for validity. At a minimum it appears that some of the items in the Spanish-language assessment deserve closer scrutiny, with an eye toward dialectical or regional differences in meaning.

As in the English assessment, the Vocabulary section of the Spanish-language assessment was the source of many concerns. The quality of the drawings was an issue for staff, and they also noted that the black-and-white line drawings were not only visually uninteresting to children, but that the renderings were difficult to identify even when children knew the word. A common example was that the drawing of the skirt in the Pre-

[58] We sampled children for Spanish assessment observations at 17 programs, but a total of 29 of our 40 programs in spring 2007 offered the Spanish assessment. Note that Chapter II describes the experiences of the 32 programs assessing English Language Learners, and that in 29 of these 32 programs the ELL children speak Spanish.

LAS section 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.

Potentially serious problems concern the meaning of words in different Spanish dialects. For example, according to staff in Puerto Rico, "juala" (cage) 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. Puerto Rico-based staff raised this same issue in the spring 2006 round of site visits. A second vocabulary concern focused on the snake item in the TVIP section. For this item, assessors at several sites said that their children responded with "víbora," rather than "culebra" and in two of these sites responses were confined to children of Mexican descent. Finally, staff in one program that assesses children from a Dominican background commented that the word for "ball" in that dialect is "vijiga," an unacceptable response on the NRS. In contrast, assessors reported that some variation in vocabulary choices among dialects was useful on other items. Staff in several programs that assess children from Mexican backgrounds noted that "bote" means basket or garbage can, rather than boat. These staff reported adjusting the assessment to use the word "barco" when administering the boat item, (the easel makes this an option for assessors). This illustrates one way that assessors have been able to adapt the assessment, using an appropriate dialect variation approved by OHS, to accommodate the linguistic background of children in their program. Allowing a similar word choice for other items might alleviate some assessor concerns about how dialect affects child performance on the problematic vocabulary items above.

Other concerns about the Spanish assessment included different items on the Spanish and English versions of the test. Some, 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 was not a part of the Spanish-language assessment. Concerns in the Letter Naming section focused on (1) inclusion of four Spanish letters (ch, ll, ñ, rr)[59], and (2) why children had to name the letters in Spanish (for example, 'ah' as opposed to 'a'), when they are usually taught to say letters in English in the classroom (see Chapter II). The concerns raised by staff interviewed in spring 2007 are similar to those raised in prior years of the Quality Assurance Study.

[59] Although program staff mentioned that three of these letters were no longer employed in Spanish usage, OHS has determined that because only "rr" is actually out of common usage, it is being eliminated from the assessments.

Staff Time and Financial Resources Dedicated to the NRS

In the majority of the programs, staff expressed concern about the amount of time and financial resources dedicated to the NRS (see Box on this page). Similar concerns had been raised in previous rounds of site visits, but in spring 2007, we saw a sharp increase to 70 percent of programs reporting problems with staff time: 40 percent reported this problem in 2006, and 60 percent did so in 2005. Directors and lead trainers complained that time dedicated to the NRS takes teachers away from their classrooms.

Concerns About Staff Time and Financial Resources

	Percentage of Programs
Staff time devoted to the NRS	70
Funds spent on the NRS	20
Funds spent printing easels	20
N = 40 Head Start programs	

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, One fifth of programs reported an increase in direct costs associated with the NRS. Many other programs reported that the NRS uses program resources (in the form of staff time, mileage reimbursements, and training dollars) that are covered by their OHS grant, but that would otherwise be spent on other program activities. These programs described absorbing the costs into their regular operations. 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. One-fifth of programs specifically cited the costs of printing new assessment easels for each round when only a few items typically change. In fact, 13 percent of programs made the specific suggestion that sending replacement pages would a better approach. These concerns are similar to those raised in prior rounds of the study.

Other Concerns

Staff members also shared their concerns about two other aspects of the NRS in spring 2007: (1) overtesting children and (2) assessing ELLs. At 18 percent of programs, staff said that they are alarmed by the number of assessments and tests administered to very young children, and that the NRS contributes to the problem. For example, one program director described her children as being bombarded with tests and screenings throughout the year, which come from both their local curriculum as well as state and federal requirements. Two programs (5 percent) that enrolled mostly children speaking languages other than English or Spanish were disappointed that children could not be assessed in their home language, creating substantial confusion for the children and assessors. Children were uncertain about which language to use for their answers, assessors were uncertain about scoring answers that would translate correctly, and staff at both programs expressed a strong wish that the NRS have assessment options that were more appropriate for these children. Additional information on how to handle assessing children speaking a third language may be useful to these programs.

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 offered a broad range of suggestions for improvement that cluster in five main categories: (1) communication and planning, (2) reporting results, (3) adding domains to the assessment battery, (4) improving the Spanish-language version of the assessment, and (5) improving assessment procedures. Many of these suggestions naturally follow from concerns expressed above, and many have already been described in previous sections of the report. We briefly describe them here. The suggestions staff shared this year were similar to those in previous years of the study, except that this year there were more suggestions for how results are reported and in previous years we heard more suggestions for training and guidance about administering the assessment.

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 38 percent of the programs either stated that they did not understand why they do the NRS or suggested that OHS provide clarification about the purpose of the NRS and more-specific information about how the results will be used (see Box on this page). In 20 percent of programs, staff expressed uncertainty about how they could use NRS results locally. About one quarter of programs requested this guidance in previous years of the NRS quality assurance study.

Suggestions for Improving Communication and Planning

	Percentage of Programs
Provide more information on purpose (or concern about not understanding purpose)	38
Clarify suggested or appropriate use of data	20
Provide materials for stakeholders	18

N = 40 Head Start programs

Nearly one fifth of programs asked for written materials they could use to describe the NRS and its purpose to parents, Policy Council, Board members, and other program stakeholders. Staff at various programs suggested a form letter from OHS, some text programs could adapt for inclusion in their handbook, and a video geared towards parents as additional materials that would be helpful to them. Programs made similar requests for these materials in 2004 through 2006.

Late shipment of materials has been an ongoing challenge of NRS implementation, but program concerns about materials were somewhat less frequent in spring 2007 than we have observed in prior years (see Chapter III for more discussion of this issue). While assessment and training materials did arrive earlier in spring 2007, programs would like to have some prior warning about shipment so that they can better plan the training and assessment schedule. For example, staff at one program said that they "feel blind" about when materials will arrive, and this contributes to an already high stress level about the amount of work they must complete in a short time period. Program staff are apparently not aware of

the rolling dates for sending materials based upon programs' start dates and therefore have trouble planning.

Reporting the Results

More than one third (38 percent) of programs requested that OHS send outcome reports for each wave of assessments and the growth report on children's progress from fall to spring more promptly after each wave of assessments. 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.) Many programs explained that the timing of the distribution of results makes the NRS of little use for them locally, despite the new policy of providing interim reports online over the summer.

At some programs, staff made suggestions about how the format and content of the NRS reports could increase their usability. For example, one program director suggested that OHS provide a report that compares results across years for each program so that any trends over time are evident. Three programs made suggestions about improving report layout. Two of these suggested changing the font and coloring of the reports so that they could be more readable and photocopied more clearly for distribution to program staff. At another program, the lead trainer suggested that reports place the national and program means in adjacent columns rather than across the page from one another. She also suggested that "Mean Score" be re-labeled as "Your Program's Mean Score," and to define whether the row of percentages under each skill level pertains to the grantee or to the nation. Staff in nearly half (45 percent) of programs expressed a desire for NRS results at a more disaggregated level–by county, center, or classroom. This request echoes feedback from programs in previous years. Although not feasible because these data are not collected, at several programs (18 percent) staff described that reports could be more useful if the assessment could be extended to other populations (three-year-olds, other non-Head Start pre-kindergarten children, kindergarteners) so that they could observe growth over time.

Expanding the Assessment Battery

As in spring 2005 and 2006, staff members were split on the question of whether new domains should be added to the NRS. Staff seemed to be torn between desires 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 a large majority

Expanding the Assessment Battery

	Percentage of Programs
Do not add new domains/ lengthen assessment	70
Add at least one domain	38
• Science	
• Gross/fine motor skills	
• Approaches to Learning	
• Physical Health	
• Mental Health	
• Domain not specified	

N = 40 Head Start programs. Numbers do not sum to 100 percent due to multiple opinions expressed within single programs.

(70 percent of programs) said the length of the assessment concerns them, and/or that new domains should not be added. This is a large increase from spring 2006, in which 37 percent of programs said that no more domains should be added. In 2007, the SED rating form was added to the NRS and this may have satisfied program staff members' wishes for the NRS to measure additional domains. At the same time, staff in more than one third of programs (38 percent) indicated that the assessment was too restricted and suggested adding at least one new domain to the NRS. Domains that staff suggested included science, gross and fine motor skills, approaches to learning, physical health, and mental health. Concerns about too narrow a focus appear to have increased slightly over the last three years, but are still less prevalent than in the first year of the NRS, when 60 percent of programs thought the assessment's scope was too narrow.

Improving the Spanish-Language Version

Programs made a number of recommendations about improving the Spanish language version of the assessment. The issue of dialect differences often came up and staff suggested more flexibility to substitute words in the directions to account for the 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 the Quality Assurance Study. Beginning in fall 2004, OHS allowed for discretionary use of some terms, such as replacing "señala" (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; feedback from others indicated that they are not aware of the change.

The concerns section already notes that some program staff raised issues about the four letters (ch, ll, rr, ñ) in the Spanish Letter Naming task, and problems with specific words in the receptive Vocabulary section. At one program, staff made the suggestion that all acceptable dialect differences in questions and responses be placed on the easel itself for easy reference. Finally, staff in a few programs said that the items and pictures in the Spanish-language 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 (a psychometric issue that cannot be assessed based on face validity of the items, but that can affect staff attitudes and perceptions of the instrument).

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, 15 percent of the programs suggested combining the NRS and the local assessment. At 30 percent of programs, staff reported that the NRS is redundant and duplicates information they already collect in their local assessment. Staff at one program suggested that the NRS be administered in separate five-minute modules as a method of addressing concerns about child fatigue and age-appropriate assessment methods. At

another program, staff suggested that the OHS introduce some kind of permanent mechanism for staff to provide feedback on the content and administration of the NRS. At three other programs, assessors suggested easel changes: modifying the easel to show a greater contrast in font and formatting between instructions and script, placing tabs on easel pages to facilitate turning one page at a time, and orienting the numbers on the stimulus quadrants towards the assessor (though the pictures in the quadrants should still face the child). Local program staff also made a broad range of suggestions to improve specific sections of the NRS assessment, noted above and described fully in Chapter II.

CONTRIBUTIONS, CONCERNS AND SUGGESTIONS SPECIFIC TO TRIBAL PROGRAMS

In spring 2007, we intentionally oversampled tribal Head Start programs to participate in our site visits. At the five tribal programs we visited, we asked some specific questions that were not covered at other sites, including questions about any contributions, concerns, or suggestions related to the NRS that were specific to tribal programs. At two of these programs, staff appeared to take exception to the idea that tribal Head Start programs would be any different from non-tribal programs. In the others, program staff that noted differences were sometimes reluctant to use the experience of their own tribe to speak for tribal programs in general.

When asked about contributions of the NRS, tribal program staff unanimously reported that contributions are not unique to tribal programs (though some cited the contributions mentioned by other programs, such as a national comparison and preparation for future testing in kindergarten). In four of these five programs, staff had no concerns about the NRS that they considered unique to their tribal status. The fifth program had some concerns about items and language of administration (most children at this program speak a tribal language at home) but staff noted that these concerns were due to geography and the local language and could also apply to non-tribal programs. Similarly, programs had no suggestions for improving the NRS in ways that would uniquely benefit tribal programs, although one program did recommend that the PPVT include more items that would be culturally relevant to tribal children.

SUMMARY

Program staff cited a number of contributions the NRS has made to their programs, notably the national comparison it provides, an indication that programs have begun to appreciate the primary aim of the NRS and to use the data as intended. Staff also reported that the NRS contributed to their program by raising program accountability, improving classroom practice and teacher training, measuring child knowledge, proving Head Start is effective, and raising staff awareness about assessments and OHS expectations. In all, staff at 85 percent of programs reported that the NRS made some contribution.

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 OHS share more information about how the results would be used, send training materials and outcome reports to programs more promptly, provide additional 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 materials about the NRS for parents and other program stakeholders.

Regarding the assessment battery, programs were split on whether new domains should be added. Most thought that the assessment needed no additional domains, but approximately one third suggested adding at least one more. As in previous rounds of NRS site visits, program staff recommended 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.

In our experience talking with staff at selected tribal programs, no concerns, contributions, or suggestions were identified as being unique to tribal programs. These staff shared reactions to the assessment similar to those observed in other programs.

CHAPTER VII

THE SOCIAL-EMOTIONAL DEVELOPMENT RATING FORM

The spring 2007 implementation of the SED rating form marks the first time that the same social-emotional development rating tool has been administered to all children enrolled in Head Start during the year prior to kindergarten. The form is completed in both the fall and the spring. ACF instructed local programs that information recorded on the fall SED form should be based on behavior observed during the first four weeks that the child is enrolled in Head Start to ensure that staff get to know children before completing the form. The form is intended to be completed by the staff person or persons who know the child best. In addition, unlike the NRS cognitive child assessment, ACF does not require special training or certification of staff who fill out the form. However, completion and interpretation of the SED rating forms in a consistent manner across all Head Start programs is crucial for maintaining the integrity of the results. As such, we consider the possibility that different standards for assessor training, coupled with differences in timing of completion of the rating forms, and standards for who completes them, could influence the quality of information and therefore the usefulness of the data on a national level.

In this chapter, we describe the approaches to training and implementation of the new SED rating form in the 40 programs we visited in spring 2007. Our program-level data are derived from individual interviews with the program director and the lead trainer, and a focus group with the staff members who completed SED rating forms. Specifically, we describe the national training materials, programs' approaches to local training, and staff feedback on these resources and experiences; implementation of the SED rating form and coordination of activities; tracking the completion of the rating forms; and communicating with parents, Policy Council members, tribal leaders/elders, and other stakeholders about the new instrument. We then present feedback from local programs on the content of the rating form and staff plans to use this information for program improvement efforts. Finally, we describe the overall impressions of the rating form, including the main contributions and concerns that local program staff identified during the spring 2007 site visits.

TRAINING AND TECHNICAL ASSISTANCE

To prepare local programs to administer the new SED assessment tool as part of the overall NRS initiative, ACF prepared a set of instructions addressed to both the lead NRS trainer and those staff who would complete the rating forms. In addition, to introduce Head Start programs across the country to the SED rating form, ACF produced a 15-minute webcast to provide more details about the form and guidance for its implementation, which local staff could download and view at any time.[60] In this section, we first discuss local programs' responses to the training materials. We then discuss the local approaches to preparing staff to complete the SED rating form and staff feedback on local training efforts.

National Training Materials and Resources

All programs felt that the SED-related instructions and materials received from ACF were easy to understand and sufficiently demonstrated how to implement this new component of the NRS. Lead trainers did not report receiving questions from staff members who would be responsible for completing the forms, and none reported calling the NRS helpline for technical assistance. Staff in the majority of programs even noted that the clarity of the materials, coupled with similarities of the SED rating form with their own local assessment, made conducting a formal local training on the form unnecessary.

Staff in many programs could not recall when they received training materials in the fall. However, programs reported beginning completion of the forms in September or October, often immediately after receiving them from OHS. In the spring, programs reported receiving SED materials around the same time as the NRS cognitive child assessment materials. While programs often cited competing program tasks and activities occurring within the first few and last few months of the program year, none reported that the timing of the arrival of materials—in either fall or spring—created a challenge in terms of meeting the deadline for completing the SED form.

Approaches to Preparing Local Staff to Complete the Rating Forms

The majority of programs (85 percent) did not hold a formal training on the SED rating form; no training materials are provided by OHS, and no formal training is recommended.

[60] The webcast explained the rationale for introducing the new social-emotional development tool, described the pilot phase of the development of the rating form, and gave an overview of the three scales. It also outlined which staff should complete the form, which children should be rated, and identified the types of information the SED rating form is intended to capture and *not* capture. None of the staff in the sampled programs visited in spring 2007 reported watching the webcast. As a result, in this chapter, we do not discuss staff members' feedback on this aspect of training. However, in fall 2006 follow-up phone interviews with lead trainers from the spring 2006 sample of programs, we learned that staff from one-third of programs had watched the webcast. Five programs (15 percent of the sample) reported that they appreciated learning about the motivation for the SED and the results of the field test. Three programs did not feel the information was necessary because written instructions were so straightforward, and three programs did not comment on the usefulness of the webcast. Program staff members who did not watch had a variety of reasons why, ranging from not realizing that it was available at more than one time to having limited Internet access.

In these programs, lead trainers distributed the instruction sheet and rating forms to staff members who would be responsible for completing them, briefly informed them about the new component of the NRS (either through a memo, at a staff meeting, or as part of the NRS cognitive child assessment training), and instructed raters to review the materials and to contact the lead trainer with any questions or concerns. Aside from minimal instructions (e.g., filling in the bubble forms correctly, filling out forms for NRS-eligible children only), no other directions were provided to staff in these programs. Although these programs did not hold a formal training to inform staff about the new SED rating form, about one-third of them held a separate meeting on the SED rating form, either at the end of the regular refresher training for the NRS cognitive child assessment, or as part of a staff training or in-service day. None of these lead trainers reported that staff had any questions.

Lead trainers in only six programs (15 percent) implemented a more methodical approach, which included reading staff the memo they received from ACF and distributing blank rating forms, reviewing the rating form item-by-item, or facilitating a question-and-answer session. In some cases, lead trainers focused on specific items on the form that staff had questions about rather than reviewing the entire form.

Feedback on the Local Training

Lead NRS trainers overwhelmingly reported that the content and depth of the materials provided by ACF for the SED rating form were adequate. Most of them agreed that nothing additional beyond the national explanatory materials was needed to prepare staff to complete the forms, and staff reported overall that the form was easy to complete. As noted above, many lead trainers did not feel that a formal training session was necessary and instead simply provided the rating forms and instruction sheet to staff. Most programs (90 percent) did not offer any suggestions on ways to improve the training process for the SED rating form. However, two lead trainers (5 percent) would have liked more direction from ACF for training. One lead trainer desired more clarity on the selection, purpose, and origin of items included on the SED rating form so that she could be better prepared for any questions staff had. Another suggested providing more training for lead trainers, particularly on the rating categories, to better prepare them to assist staff in completing the forms since the rating forms were new to all Head Start staff, including lead trainers. This lead trainer suggested that ACF describe the SED rating form and scales on a training video.[61]

At the same time, SED raters also provided ideas for improving their preparation for completing the SED rating forms. Staff at two programs (5 percent) would have liked additional guidance during training on how to interpret the meaning of the four rating categories—"never," "sometimes," "often," and "very often"—and how to define frequencies when rating children. Raters felt that they did not know what length of time or

[61] As noted earlier, prior to its implementation in fall 2006, ACF produced a webcast to provide more details on the rating form and guidance for its implementation. These concerns expressed by lead trainers suggest some staff were unaware of this webcast and its content.

how many incidents would define a behavior as "sometimes" versus "often."[62] Some also noted uncertainty about how to code a behavior that they had not observed, reporting that they felt uncomfortable indicating "never" just because they had not observed the behavior. Staff noted that assessors often struggled with deciding between two options, and noted the interpretation of these categories could likely vary from assessor to assessor. Accordingly, additional guidance on how raters should define and interpret the frequency categories would be useful to these programs. Staff in one program that did not provide formal training on the rating form indicated that more formal guidance on completing the SED forms would have been helpful. This suggests that staff in some programs would benefit from a more formalized training, regardless of whether lead trainers felt the rating tool and accompanying materials were self-explanatory.

Overall, local Head Start programs did not report any challenges with the training and technical assistance provided in support of the rating form. Nearly all of them intended to use the same approach to local training in the future.

IMPLEMENTING THE SOCIAL-EMOTIONAL DEVELOPMENT RATING FORM

This section describes the approach of the 40 sampled programs to coordinating SED rating form activities to support smooth implementation and timely completion of the rating forms. We focus on programs' approaches to four coordination issues: (1) assigning staff to complete the rating forms, (2) timing of completion of the SED rating forms, (3) scheduling logistics and completing the forms, and (4) how programs track progress in completing the forms.

Programs' Main Approaches to Staffing the SED Rating Forms

	Percentage of Programs
Children in center-based option	
Teachers/assistant teachers	100
N = 40 Head Start programs	
Children in home-based option	
Home visitors	45
Classroom teachers	45
Home-based teachers	9
N = 11 Head Start programs	

Assigning Staff to Complete the Rating Forms

Among programs that described their approach to determining who would complete the SED rating forms, almost one-half said that management-level staff, such as lead trainers or directors, assigned the staff members who were most familiar with the children. A similar percentage of these programs reported that they assigned staff in accordance with the instructions provided by ACF, often suggesting that raters should be classroom teachers or a similar staff person who knows the child well.

[62] As discussed in a later section of this chapter, other staff also voiced concerns about the interpretation of the rating categories as it relates to completing the forms consistently and reliably across raters and programs. Staff did not draw links between these concerns and implications for local training; instead, they reflect staff concerns about the rating form itself and interpretation of items.

The remaining programs indicated that they assigned certain staff because they were also responsible for completing other assessment and screening tools for the program locally.

All programs used classroom teachers to complete the rating forms for children enrolled in the center-based option. Among the 11 programs (28 percent) with home-based services, almost half used home visitors, 9 percent used home-based teachers, and the remainder reported using classroom teachers who saw the child during group socializations. Notably, staffing approaches for rating home-based children may affect staff members' ability to rate children accurately. For example, home visitors in one program indicated that it was difficult to rate some of the items because it was important to see a child in a classroom setting. Classroom teachers responsible for rating home-based children in one program indicated that they had difficulty rating children at the beginning of the program year because they had only seen them during group socializations on a few occasions. Although these concerns are rare, they may indicate the need for adjustments in procedures for rating home-based children.

Head Start programs used a range of strategies for assigning responsibility for the SED rating forms among staff. At 23 percent of the programs, only one staff member—either the lead teacher in a classroom or a home visitor, home-based teacher, or classroom teacher for a given home-based caseload—filled out the form for an individual child. On the other hand, in about three-quarters (78 percent) of the programs, at least two staff members worked together to complete each form. In more than one-quarter of these cases, the teaching assistants or aides worked collaboratively with the lead teacher to complete all forms for the classroom. In the other programs, lead teachers sometimes consulted their teaching assistants if they wanted a second opinion (for example, if the teacher was not sure how to characterize a child regarding a specific item), or if the assistant or aide was more familiar with the child. One program reported conducting a large meeting in which all staff discussed the rating items jointly as they individually completed the forms for each child.

Across all programs, twenty-five percent used a hybrid approach to completing the forms. For example, in some classrooms, the lead teacher and teaching assistant collaboratively filled out the forms, while in others the lead teacher filled out all forms without any additional input. In four other programs, the lead teachers and teaching assistants divided all SED rating forms for the classroom and filled them out independently as a way to complete the forms more efficiently.[63]

[63] Our data do not allow us to indicate whether the split was random or based on which staff member knew the child better.

There was considerable variation in the extent to which programs used the same staff members to complete the SED rating forms and to conduct the cognitive child assessments. Sixty percent of programs (see Box) reported crossover in staff used to conduct the cognitive child assessments and the staff used to complete the rating forms. For example, teaching and non-teaching staff conducted the cognitive child assessments but only teachers filled out the rating forms, or lead teachers were involved in both assessments while the teaching aides only played a role on the social-emotional development tool. In contrast, 10 percent of the programs used the same staff to administer the NRS assessments and complete the SED rating forms. About one-quarter of the programs (28 percent) used different staff members to conduct the cognitive child assessments, usually non-teaching, management-level staff members, such as center supervisors, family service specialists, and program directors.

The overwhelming majority of lead trainers did not have plans to change their staffing approach for the SED rating forms in future rounds. One program will continue to use a similar staffing approach but will also consider asking the mental health consultants to help with the forms, since these consultants assist with similar information gathered locally.

Programs' Approaches to Staffing the SED Rating Form and the Cognitive Child Assessments

	Percentage of Programs
Some SED and NRS staff crossover	60
All SED and NRS staff differ	28
All SED and NRS staff are the same	10

N = 40 Head Start programs

Timing of Completion of the SED Rating Form

The fall SED rating form aims to capture a baseline measure of some key components of children's social-emotional well-being. As such, it is important that staff have had enough time to become familiar with children in order to evaluate their attitudes, behaviors, and interactions with their peers and adults. Head Start programs are instructed to wait at least four weeks before they complete the rating forms, which are based upon observed behaviors in the classroom or home environment.

Staff members from center-based programs were divided on whether they thought that four weeks was an adequate amount of time to get to know the children so that they could fill out the SED rating forms. Thirteen programs (33 percent) thought that four weeks was enough, noting that staff members also complete the program's own social-emotional ratings within the same time frame.

However, 50 percent of programs did not think that four weeks gave staff enough time. Several SED raters noted that the beginning of the program year is a period in which children are adjusting to a new environment and not yet exhibiting "typical" behavior. Another program reported that while four weeks is enough time to rate children enrolled in its full-day, full-time centers, staff from the part-day, part-time centers did not feel they can accurately assess children in their classrooms after four weeks, especially if the child's

attendance is irregular. Staff in 18 percent of the programs thought four weeks was an adequate amount of time for children who are second-year Head Start participants because staff members are already familiar with them.

Among the 20 programs that did not think four weeks was long enough, 11 of them (28 percent of our total sample) recommended waiting between six and eight weeks before completing the SED rating forms. Other programs did not give a specific time frame or were unsure how much time was enough time.

Staff who completed forms on children enrolled in the home-based option also had mixed views on the adequacy of the time frame. One-third of these programs thought that four weeks was enough time to become familiar with children's attitudes and behaviors. In contrast, the remaining two-thirds reported that staff often struggled with feeling that they did not know the children well enough. These programs noted that at the time that staff needed to complete the SED rating forms, they had only observed children a few times in the home or during one or two group socializations, which did not provide sufficient opportunity to assess how children interact with their peers, get along with others, and so forth. Staff in one of these programs suggested a guideline of "number of visits with the child" rather than "weeks enrolled" to determine when a rating form should be filled out. This program also suggested involving parents of home-based children in the rating process to make it more accurate. Although these comments were rare, they may indicate the need to specify different procedures for rating children in home-based programs.

Scheduling Logistics and Completing the Forms

Regardless of their opinions on whether or not four weeks provided an adequate amount of time for staff to become familiar with the children, nearly all programs (98 percent) reported waiting at least four weeks prior to completing the forms. Raters in one program reported waiting only two weeks after enrollment, so that the forms were completed before the local assessment; these raters also indicated that this was not sufficient time to get to know children to accurately complete the forms.

Among programs that described their approach to scheduling and completing the forms, most (91 percent) gave teachers and home visitors a deadline to complete forms, and asked that they do so according to their schedules and work duties. Two programs gave raters a specific window of time to complete all SED forms. One program postponed a large staff meeting so that staff could complete all rating forms at the same time. Across programs, staff reported filling out the rating forms whenever their schedules allowed, including during naptime, after children went home, and during planning periods. All programs were able to submit the SED rating forms to the Head Start NRS Processing Center by the deadline of November 10, 2006, and all anticipated being able to submit them by the spring deadline of June 29, 2007.

Programs were mixed in their impressions of how burdensome the SED rating form was for staff to complete. No programs formally tracked how long each form took to complete, though staff gave average estimates that ranged from 2 to 20 minutes per child.

They cautioned, however, that times would vary among staff members due to reasons such as differing abilities to recall children's behaviors or differences in the behaviors of children themselves. For example, some staff suggested that it took longer to rate children who exhibited more problem behaviors, who were more introverted, or whose behavior varied across occasions. Staff in two programs (5 percent) indicated that it could take up to thirty minutes to complete some forms if they have to consult with other staff or reference information from other program records and staff notes about the child and his/her behavior. Some staff also indicated that completing the forms in the spring is a much quicker process than in the fall, because they are more familiar with the children and their typical behaviors at that time.

Most of the programs (93 percent) did not find the SED rating form task to be time consuming, and most did not say it was burdensome for staff (83 percent). Even though most programs appreciated that the form is short and takes little time to complete, they repeatedly noted that the start and end of the program year are busy, stressful times for Head Start. At the start of the year, staff must conduct other screenings and assessments, parent-teacher conferences, and home visits, along with facilitating transitional activities, establishing routines, and bonding with the children. Similarly, the end of the year is marked by other assessments and end-of-year activities. As a result, one-fifth of the programs noted that they had delayed completion of the rating forms until after the local assessments and other activities, in an effort to reduce burden on staff.

Programs did not reveal any information to suggest that local Head Start staff may not be following the appropriate protocols for implementing the SED rating forms. Notably, a few programs reported using resources other than fellow staff members to complete the forms. In nine programs (23 percent), staff indicated that they used anecdotal or classroom notes, home visit forms, and information from other screenings or records to "jog their memories" and complete the rating forms for individual children.

TRACKING PROGRESS IN COMPLETING THE RATING FORMS

All but one lead NRS child assessment trainer assumed responsibility for overseeing and tracking completion of the SED rating forms.[64] Typically, classroom teachers or home visitors forwarded completed forms directly to the lead trainer upon completion.

According to programs, staff completed SED rating forms on most children who were eligible. To be eligible, children had to have been enrolled in Head Start for at least four weeks prior to September 25, 2006. Fifteen percent of the programs did not submit forms for a small number of children, for reasons such as late enrollment, prolonged absences, and dropping out of Head Start. One of these programs did not complete an SED rating form

[64] The lead NRS trainer in this program was a member of the grantee staff and did not have oversight responsibility for the SED rating forms; instead, another staff member—the education coordinator—was responsible for overseeing completion of the forms.

for a child with a diagnosed disability.[65] Staff in one program suggested that they had completed rating forms for children who were chronically absent, even though completion of these forms was admittedly very difficult.

COMMUNICATING WITH PARENTS, POLICY COUNCILS, TRIBAL LEADERS/ELDERS, AND OTHER STAKEHOLDERS ABOUT THE SED RATING FORM

More than half of programs (58 percent) reported sharing at least some information with parents, Policy Council members, or other stakeholders about the new SED rating form, and 28 percent shared information with more than one group. Staff in the remaining programs (14 percent) indicated that they did not discuss or inform parents or the Policy Council about the SED rating form separately from the NRS cognitive assessment. Instead, the NRS was discussed broadly, without explicit mention of the new SED rating form. Staff indicated that they generally inform parents about the new instrument at enrollment or through written communication sent to the child's home. Ten percent of programs opted to discuss the SED form during one-on-one parent conferences with classroom teachers or home visitors or during larger parent committee meetings. Programs shared information on the form with Policy Council or Parent Committee members at their regularly scheduled meetings, and one program also shared information with their board of directors.

Staff from most programs (29 out of 31 reporting) felt adequately prepared to discuss the rating form with parents and stakeholders and to address their questions or concerns. Two programs, however, expressed some reservations. One of these programs indicated that staff would have felt more confident about responding to questions if they had more information on the results and how they would be reported. A staff member in the other program said that she would not be able to explain the purpose and background of the rating form or how the results will be used (although as noted, staff did not watch the webcast on this topic). Although these exceptions are rare, they may indicate the need for providing further training or guidance to some program staff, or ensuring that they are aware of information that has been provided.

Parental Consent Forms

None of the sampled programs asked that parents sign a special consent form allowing their child to be rated on the new social-emotional development form. In contrast, among those reporting, the majority of programs either mentioned using a blanket parental consent form for all assessments (91 percent) or considered the SED form as being covered by a special consent form used for the NRS cognitive child assessment (8 percent). Because none of the programs required special consent for the SED rating form, no programs reported having parents refuse consent to have their child rated.

[65] Neither the webcast nor the instructions to NRS lead trainers and classroom teachers directed programs not to complete an SED rating form for children with disabilities.

Concerns Raised by Parents and Stakeholders

Parents rarely raised concerns about the SED rating form. As staff in a few programs explained, parents are aware of (and sometimes involved with) the program's local assessment and are thus familiar with the purpose and usefulness of social-emotional development evaluations. Of the 58 percent of programs that shared information about the SED ratings with parents, only one program received questions. A group of parents from this program wanted to know more about what the SED rating form included and how it would be used. Similarly, only two Policy Councils (out of 17 stakeholder groups that received information on the rating form) raised any concern. They worried about the overall number of assessments in which children were participating and whether the rating form was redundant with the local social-emotional rating tools.

CONTENT-SPECIFIC FEEDBACK ON THE RATING FORM

During interviews and focus group discussions, Head Start staff were asked to provide feedback on the three scales found in the SED rating form—Scale A (Approaches to Learning), Scale B (Cooperative Classroom Behavior and Relations with Other Children), and Scale C (Behavior Problems)—along with the rating categories ("never," "sometimes," "often," and "very often"). Several programs also commented on specific items within each rating scale. This section describes their opinions of the form's content and the rating categories.

Content of Scales and Items

One-half of the programs were satisfied with the content of the rating form and did not raise concerns with any items. They thought that the instrument's items were relevant, appropriate, and comprehensive. Only four programs (10 percent) thought any additional scales should be added to the instrument. Staff in these programs suggested adding scales or items that captured children's independence and self-care skills, for example, tying shoelaces, toilet learning, zipping up pants (two programs), self-regulation (one program), and ability to solve conflict with peers (one program).

Twenty-five percent of programs had concerns with multiple items across scales, or else took issue with an entire scale. For example, 15 percent of programs thought that the tone and wording of the form was too negative, with more of the items on the form focusing on children's problem behaviors rather than their social skills, and with many of the items using very strong language (e.g., "feels worthless or inferior," "hits or fights with others"). Approximately 20 percent of programs also raised the issue of reliability or subjectivity with certain items, especially those that lacked observable behavioral anchors. For example, staff asserted that two teachers could have rated the same child completely differently for certain items based on their interpretation of the item, tolerance for specific behaviors, or the abstractness of the behavior (e.g., "feels worthless or inferior," "worries about things for a long time"). Five percent of programs also indicated that the wording and ordering of items sometimes made rating difficult. For example, items containing more than one descriptor (e.g., "is easily confused, seems to be in a fog," "is nervous, high strung, or

tense") were difficult to rate when at least one of the descriptors did not apply to the child's behavior. A similar percentage of programs questioned the translation of the Spanish-language form. For example, staff felt that the translation was not always clear or an accurate reflection of the English-language form. Staff in one tribal program (3 percent) said that the instrument does not take cultural differences in perspectives on children into account, making it problematic, but did not specify items.

While the remaining programs that cited issues generally liked the form's content, they offered feedback on a small number of items. One concern was whether the items were developmentally appropriate for preschool-age children (18 percent). For example, staff maintained that preschool children tend to be impulsive (item A5), do not respond well to teasing (item A6), and are not particularly organized (item A2). In addition, 8 percent of programs said that items were difficult to answer because they were worded using double negatives (i.e., item is phrased negatively or identifies a negative behavior and the assessor should respond in the negative if the child does not engage in the behavior, for example "lacks confidence"). Fifteen percent of programs also mentioned that items such as C8 ("feels worthless or inferior") focus on children's inner thoughts and feelings, which they felt cannot be objectively or easily observed in children, particularly those of this age. In some instances, staff indicated that they did not feel qualified to answer such items.

Rating Categories

Slightly fewer than half of programs (43 percent) thought that the rating categories ("never," "sometimes," "often," "very often") allowed them to record scores accurately for each item, but the majority of the programs (58 percent) offered a wide range of suggestions to improve the categories. Several suggestions centered on the number of categories. Four programs (10 percent) said three categories would be sufficient and would make rating easier, because it would offer a "middle" category. In contrast, 10 programs (25 percent) would like to see one or two categories added to give staff more options in characterizing attitudes and behaviors. Seven programs (18 percent) would like to see an "N/A" or "not observed" category added to account for those items not observed during the first four to six weeks of school. Staff in these programs felt this would be useful when rating all children, but particularly for those of some home-based children. Three programs (8 percent) thought that "rarely" should replace "never" since they felt it is a more realistic metric than an absolute. Finally, staff at six programs (15 percent) struggled with defining the rating categories (for example, does "sometimes" mean once a week or three times a week?). These staff often felt that having a suggested time frame for reference would make rating easier and less subjective across raters.

USING DATA FROM THE SOCIAL-EMOTIONAL RATING FORM

ACF intends that the results of the SED rating form, like the NRS cognitive child assessment, will help the Office of Head Start better understand the growth in social-emotional development outcomes of Head Start children as they prepare to enter kindergarten. The tool is designed to help provide a national perspective on children's growth and development in the areas of approaches to learning, social-emotional skills, and

behaviors, with the intention of informing training and technical assistance decisions, as well as program improvement efforts. This section reports on programs' views on the usefulness of the data to local programs, the preferred format for results reported back to local Head Start staff, and their plans to use data from the SED rating forms.

Perceived Usefulness of the Data at the Local Level

Sixty percent of the programs thought that the SED rating form duplicated local assessments and screenings that staff members complete for individual children. In their opinion, for example, the new tool is not as comprehensive as local instruments, does not capture any new information not already collected through the local assessment process, and does not generate any "value added" for staff members who complete the rating forms or for the program locally.[66]

In contrast, some programs (28 percent) observed differences between the SED rating form and local assessment tools and in the information it provided to program staff. According to 18 percent of programs, the new rating form is more detailed and complements their local instruments by capturing some data that had not been collected before. For example, some of these programs felt that the items on the SED rating form were more specific and clear than those covered by their local assessment tool, making it more informative. Eight percent of programs noted that the process of filling out the form helped staff think more about individual children and the behavior of the classroom as a whole. Three percent of programs indicated that their local assessment only includes two of the three scales included on the SED rating form, and another indicated that the form contains more items than their local assessment, making the new form more informative to staff. All remaining programs indicated that the SED did not provide new insight for program staff, not necessarily because it duplicated the local assessment but because staff members were already familiar with children's behavior from their daily interactions.

During focus groups and interviews, a few of the programs offered feedback on what could be added to the SED rating form or changed about the rating process to make it more informative and accurate. Seven programs (18 percent) suggested additions to the rating form for providing context to children's ratings, including adding a notes section for documenting children's disability status (which was collected in the CBRS), number of absences, whether the child had recently experienced some sort of major change or trauma, and other sorts of explanations for why assessors picked a certain response. Two programs also offered suggestions on making the rating process more informative by including parents or a home observational component in the rating process. Finally, one program suggested receiving the rating forms earlier in the program year (or having a checklist reminder at other points in the year) so that staff could anticipate behaviors to be observed in children before filling out the rating forms.

[66] During focus groups and interviews, program staff members were not asked about perceived usefulness of the data at the federal level.

Preferred Format for the Baseline and Growth Report

Like the growth reports that have been distributed each year for the NRS cognitive child assessment, ACF planned to compile and disseminate social-emotional development growth reports at the program, regional, and national levels.[67] Staff in half of the programs provided feedback on what could be done to make the SED baseline and growth report more useful to local Head Start programs. Most commonly, programs that provided feedback reported (75 percent) that results would be beneficial to programs if they could be provided at the individual, classroom, and/or center level. Seven of the programs providing feedback (35 percent) would like to receive the results sooner to inform program planning for the following year, or to respond to program-wide "red flags" that emerge at baseline. In addition, 10 percent[68] of the programs that provided feedback suggested that the results should be accompanied by suggestions from ACF on how to improve social-emotional development outcomes of children at home and in the classroom, particularly for areas in which children have elevated problem behavior scores. Such guidance would make the rating form more useful at the local level.

Programs' Plans to Use SED Results

Most programs (83 percent) did not offer concrete plans for using the data report or results from the social-emotional development data. Head Start staff in seven programs (18 percent) offered concrete plans for using results in the future. Three of these programs reported that SED results will help local staff identify issues of concern for children served by the program and inform mental health consultants who provide technical assistance to staff; one program had already shared rating forms and the baseline report with a specialist who then crafted training for teachers to meet program needs. Two programs (5 percent) indicated that they were considering the possibility of using the SED rating form as

Contributions of the SED Rating Form

	Percentage of Programs
Promotes issues of social-emotional development among Head Start staff and federal government	35
Helps identify key issues and goal setting for individual children	25
Serves as resource for discussions with parents	10
Informs program planning and training needs	8
Complements local assessment tools	8
No contributions	30

N = 40 Head Start programs

[67] At the time of our spring 2007 site visits, staff in many programs had not yet received their baseline report or had only recently received the report but had not yet had time to review it. (ACF indicated that the data for these reports were being analyzed for reliability before being sent to programs). As a result, our discussions centered on what programs felt would be useful in a baseline or growth report for program planning and improvement.

[68] Percentages do not add to 100 because respondents could provide more than one suggestion for improving the format of the baseline and growth reports.

a screening tool or as a replacement for one of their screeners. Staff in one of these programs argued for using the rating form in this capacity since a screening tool is already required for Head Start programs and the SED is already validated and provides national comparisons.[69] Finally, two programs (5 percent) planned to use results from the rating form in their discussions with parents about children's social-emotional outcomes.

CONTRIBUTIONS OF THE SED RATING FORM

Although the majority of programs did not have concrete plans for using the data, when asked to list the contributions the SED had made or could make to their programs, more than two-thirds of programs (70 percent) offered contributions that the SED rating form could make at the child, program/agency, and national levels (see Box on the previous page). The contributions that staff listed included that the SED (1) promotes issues of social-emotional development as important to staff and the federal government; (2) helps staff identify strengths and issues of concern for specific children; (3) serves as a resource for communicating with parents; (4) informs program planning and training needs; and (5) complements local assessment results.

About one-third of programs said the form demonstrated for staff the importance of children's social development and taking a holistic approach to their learning and development (i.e., cognitive and non-cognitive). Staff in one-quarter of programs suggested that results from the rating form could be used at the child level for goal setting purposes and planning for individual children.[70] Staff in these programs suggested that the SED rating form could help local staff identify issues of concern for specific children, enable them to track individual children's progress, and inform lesson plans. They also suggested that the form had made teachers more perceptive and conscientious about specific children's strengths and needs, and had encouraged them to pay more attention to behaviors exhibited in the classroom.

A few programs (10 percent) saw a use for the data in their discussions with parents about children's social-emotional problems. For example, these programs suggested that results could be used to inform discussions about activities parents can do to promote children's social skills or to address problem areas. Only eight percent felt the form was helpful in confirming and/or complementing other information gathered by the program locally. In some cases this was because the rating form augmented the local assessment and forced staff to consider additional domains of children's behaviors that were not addressed by the local assessment tool; in other instances, staff indicated that the SED provided useful information that validated local assessments. Essentially, it served as a "double check" with local assessment results.

---

[69] From ACF's perspective, the SED rating form is not intended to be used as a screening tool and should not be used to identify potential social or emotional challenges for individual children.

[70] Although ACF stressed during its webcast that the SED rating tool should not be used to identify children with behavior problems or to guide individualized interventions, these programs indicated that they would use data at the individual child level.

Eight percent of programs said that the data could be useful at the program/grantee level to identify areas of improvement and inform program planning and staff trainings. For example, programs may use the outcomes to modify the curriculum across all centers, to decide to place more emphasis on mental health and behavior issues for professional development, or to inform the technical assistance that mental health consultants provide to staff. These programs also mentioned using the results to identify "red flags" that could be addressed by teachers in the classroom or by referrals to mental health services.

The remaining thirty percent of programs did not cite any contributions the SED rating form could make to local Head Start improvement efforts. They remained satisfied with their own local assessments and screening tools and felt they were most effective in planning for individual children and in meeting their program planning and improvement needs. For example, four programs indicated that the SED rating form is not useful locally unless immediate feedback to teachers and guidance on the results is provided.

CONCERNS ABOUT THE SED RATING FORM

During the spring 2007 site visits, programs discussed concerns about the new SED rating form and presented suggestions for improving it. Some programs (35 percent) did not raise any concerns. However, staff from the majority of programs (65 percent) raised various issues and cited features of the new tool that they did not like. Responses clustered around seven primary concerns (see Box). These were: (1) reliability and content of the rating form; (2) provision of results in a non-useful or timely fashion; (3) duplication of local efforts; (4) burden on staff time; (5) issues related to specific subgroups, including ELLs, children from certain cultural backgrounds, and children with disabilities; (6) timing of the fall ratings; and (7) lack of clarity about purpose, rating categories, and context for individual ratings.

Concerns Cited by Head Start Staff About SED Rating Form

	Percentage of Programs
Rating form is too subjective	23
Do not receive results in a timely or locally useful (i.e., child, classroom level) fashion	20
Duplicates local efforts	18
Places a burden on staff time	13
Use with certain populations is not always appropriate	10
Occurs too early in the program year	10
Do not understand purpose or use of data	8
Cannot provide contextual support for ratings	5
Rating categories are unclear	5

N = 26 Head Start programs

Twenty-three percent of the programs expressed concern about the instrument's subjectivity. According to local staff, ratings can differ dramatically across raters. For example, certain teachers might judge a behavior one way, whereas another teacher would assess the same child differently. Staff in two programs suggested that other factors, such as knowing children for a longer period of time or interacting with children and families outside of the program, made it difficult for them to

rate children objectively. Moreover, programs noted that a number of items on the SED rating form ask staff to surmise what a child is thinking or feeling, as opposed to something that can be more objectively observed (for example, "worries about things for a long time" or "is sensitive to feelings of others").

Local programs (20 percent) also had concerns about the nature of the results and the lack of timely feedback on the SED ratings. Staff in these programs were concerned that results were not provided at the child level, arrived too late in the program year to address problem areas, and were not accompanied by guidance from ACF on handling social-emotional issues. Thirteen percent of the programs noted that the rating form places a burden on local staff because it is conducted at a busy, stressful time of year during which they juggle numerous assessments and screenings, transitions for children, and parent meetings. This timing challenge is coupled with the fact that several programs (18 percent) found the form to be duplicative without offering any added value for local Head Start programs, primarily because it cannot be immediately used to inform classroom practice.

Ten percent of the programs felt that the rating occurred too early in the program year to be an accurate reflection of children's true behavior. A similar percent of programs expressed concerns related to use of the rating form with special populations, including children with disabilities, English language learners, and children in tribal programs. For example, one tribal program had concerns about the cultural sensitivity of the rating form.[71] Another program was concerned that children learning English, who are often more reserved and less vocal at the beginning of the year, might have their ratings detrimentally affected. Two programs (five percent) expressed difficulty in rating some children with disabilities, given their unique needs as compared to other children. For example, children with certain disabilities are more dependent on adults for their needs, but assessors were uncertain about whether the child should be classified as being "too" dependent on adults. Staff members in one program (three percent) were concerned that the results could be used to label children at a young age with behavioral or mental health problems. Two programs (five percent) questioned the wording of certain items, and in some cases were concerned that the tone was too negative (five percent). The same percent also felt the rating categories were unclear. Finally, a small percentage of programs expressed concern about the lack of clarity regarding the uses of the SED data (eight percent), and the inability to provide explanations and other contextual information (e.g., IEP status, behavior changes related to recent stressors or transitions) on the rating form (five percent). Although these concerns were expressed by relatively few programs, they may indicate the need for providing further guidance to some program staff.

SUMMARY

Like the NRS cognitive child assessment, if the SED rating form is to be useful at the local level, local Head Start programs must be able to implement it with fidelity to the model,

[71] In general, however, staff at tribal program did not express concerns about the SED rating form that they felt uniquely affected tribal populations or programs.

accept it as a valid and reliable instrument, and understand how it can lead to local program improvements. Unlike the NRS cognitive child assessment, ACF does not require special training or certification for staff to fill out the SED rating form. Among the programs visited in spring 2007, many lead trainers felt the rating form and accompanying materials were straightforward and required little training. As a result, most programs did not provide a formal training for staff, instead distributing rating forms and the instruction sheet to raters. Most programs reported that the rating form took little to no preparation and was easy for staff to complete. Only SED raters in one program and lead trainers in two programs expressed a desire for formalized training or more instruction from ACF. While most programs felt the forms were relatively easy to fill out and not burdensome, staff often stressed the time constraints at the start and end of the program year when the rating forms need to be completed.

As with the NRS cognitive assessments, program directors assigned a lead trainer the responsibility of overseeing implementation, including training raters, scheduling and tracking the completion of rating forms, and submitting rating forms by the deadline set by the Office of Head Start. In all but one program, the lead NRS trainer was also responsible for overseeing implementation of the SED rating form. Among the 40 Head Start programs we visited in spring 2007, all programs used classroom teachers to complete the rating forms for children enrolled in the center-based option; programs used home visitors, home-based teachers, and classroom teachers to complete the forms for home-based children. Programs adopted these staffing approaches because they felt assigned staff were most familiar with the children, they wanted to be responsive to instructions provided by ACF, or selected staff were also responsible for completing other local assessments and screenings. More than three-quarters of programs used at least two staff members to complete forms for individual children, with teachers often consulting teaching assistants or working collaboratively with them (or other staff) to fill out forms.

More than half of programs communicated with parents, Policy Councils, tribal leaders/elders, and other stakeholders about the SED rating form in some respect. Remaining programs indicated that they did not discuss or inform these groups about the SED separately from the NRS cognitive assessment. While the means of communication varied, programs made an effort to inform parents at the start of the program year. None of the programs sought to obtain parents' written consent for the SED rating form alone; instead, the majority of programs considered the rating form as being covered by a blanket consent form for all assessments or a separate consent for the NRS cognitive assessment. No programs had any parent refusals, and parents (and stakeholders) rarely raised concerns.

Most programs did not offer concrete plans for using the SED data in the future. However, more than half of programs felt that the SED could make a contribution at the child, program, or national level. Some programs thought that the SED rating form complemented local assessments and screenings, and some also expressed appreciation that a social-emotional development component has been incorporated into the NRS. Moreover, some programs expected the rating form to inform local training, technical assistance, and program improvement efforts. Many programs felt that the items included in the rating form were comprehensive and well targeted.

On the other hand, local program staff raised a number of concerns about the SED rating form and results. Many staff said that programs should receive SED results in a more timely fashion and at the child level in order to use them for program planning and improvement. In addition, some programs felt that guidance from ACF on how to address problem areas would be helpful. Local staff also expressed concern about whether the rating form imposed a burden on staff at the beginning and end of the program year; duplicated local assessment efforts; and contained items that were perceived by some staff as being too negative, subjective, or difficult to quantify based on the rating categories. Staff in a few programs also expressed concerns about the rating form for children in tribal programs, English language learners, and children with disabilities.

CHAPTER VIII

IMPLICATIONS FOR SYSTEMIC 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 reports effectively for program improvement. Accuracy of administration, validity and reliability are also important to meeting 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 four 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 other English Language Learners, and children with disabilities), (4) feedback on the new SED rating form, and (5) ideas for changes to the assessment battery. Over the course of six 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 four 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, contributing to decisions on funding or employment that are made 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 one-third of programs were still expressing concerns or confusion about the purpose in spring 2007, down from two-thirds in spring 2006. This continued 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. One fifth of programs were concerned that growth scores could affect future funding, and a small number of programs were concerned about program consequences linked to child performance. Notably, concern about the assessment's purpose is half as prevalent in 2007 as in 2006, suggesting substantial improvement in communication efforts by OHS. Ongoing efforts to help programs understand the specific ways the OHS will use NRS results can help to alleviate program concerns.

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 have brought in successive rounds. For example, in spring 2007, 88 percent of programs raised concerns about assessment validity. 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 may need further improvements. Programs are, on the whole, constructive in the suggestions they provide, many of which have been implemented in successive rounds

Concerns about validity fall into two categories: first, program staff have expressed concerns that directly administered, on-demand tests may not be valid for preschool children developmentally (for fifty percent of programs this year). Other validity concerns (for 28 percent of programs this year) 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.[72] 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. Finally, staff in nearly one-quarter of programs expressed doubts about the validity of comparing fall and spring test

[72] See Administration on Children, Youth, and Families. "Design of the Head Start National Reporting System Child Assessment," Washington, DC: U.S. Department of Health and Human Services, 2007; Administration on Children, Youth, and Families. "Head Start National Reporting System: Year One Implementation," Washington, DC. U.S. Department of Health and Human Services, 2007.

scores because the tests use different items, a concern that could be potentially alleviated by further staff education on the test development and by communicating assessment validity data.

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. During the spring 2007 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. In fact, this Quality Assurance Study is itself an attempt to gather feedback from the 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. In 2007 nearly one-fifth of programs specifically requested that OHS begin to provide them with NRS descriptive materials specifically intended for a parent audience. Programs made similar requests for guidance in the 2004–2006 rounds.

ACCESS TO AND USEFULNESS OF ASSESSMENT RESULTS FOR LOCAL PROGRAMS

Each round of data collection shows that more programs 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. The 2006-2007 program year introduced the SED rating form, and with it a report for program use.[73] As in spring 2006, 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 2006 assessment results and on children's progress from fall to spring (the 2005–2006 Growth Report) in October 2006, several months earlier than in past rounds. The Office of Head Start has moved 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 including national comparisons within a few months. In 2006, the first interim report was available online in July, and the final report in August. Programs did not always seem aware of these interim reports, and continued to complain about not having them before the start of the new program year. Some had hoped to use the spring 2006 results to

[73] The SED report had not yet been received at the time of spring 2007 site visits.

plan for the 2006–2007 program year and about 30 percent reported that they were disappointed that they did not receive the results in time for the planning process. Instead, programs developed technical assistance and staff training plans without the benefit of NRS growth results from the previous year.

Consider providing results disaggregated to lower levels and allow for easier comparisons with national, regional, and local data, and clarify appropriate uses of child assessment data. Programs varied in the level of reporting they preferred, but nearly half wanted results at least at the center or classroom level. Nearly a quarter requested results at the child level. In addition, programs requested reporting formats that would allow them to incorporate their program's outcome data into presentations and reports to stakeholders. Nearly all programs had shared their results beyond the directors and lead trainers responsible for NRS oversight. Nearly half of programs reported using the results to target their staff development efforts. 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: Nearly one-third of the programs used Creative Curriculum's assessment. About a quarter of programs used the Ages and Stages Questionnaire (which is intended for use as a screening tool), twenty percent used a "locally designed" assessment, and another seventeen percent used the High/Scope COR. 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.

Programs that made comparisons between NRS and local assessment results generally found the results aligned, although some programs decided not to compare, citing different formats and emphases. 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 majority of programs reported making changes to classroom practices in response to NRS results. Over a quarter increased their emphasis on literacy instruction, and ten percent purchased classroom supplies in order to bolster instruction in literacy. In addition, twelve percent reported having purchased a

new curriculum based on low performance on the NRS child cognitive assessment and their own local assessment.

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 when needed to plan for and schedule assessments. During prior rounds of site visits, late receipt of training and assessment materials was mentioned as a significant problem for many programs, but programs had somewhat fewer concerns about this in 2007. Problems with timely delivery continued for 13 percent of programs in spring 2007, characterized by delays in receipt that caused delays or interfered with planned training sessions. This is a large decrease from 2006, in which half of programs described delivery delays. And, in spring 2007, an additional 13 percent of programs reported that material deliveries were noticeably better than in prior years for the NRS. Nonetheless, 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 2007 training materials and information were mixed. At least one staff member at 70 percent of programs had watched one of the past NRS broadcasts, and at more than one quarter of programs someone on staff had seen 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 twice as high in spring 2007 than in spring 2006. However, some programs did not comply with suggested methodologies, either because materials did not arrive in time for planned training, assessors were considered already experienced and knowledgeable, and/or assessors lacked time for training. One quarter of programs were required to train and certify new assessors in spring 2007, although the training conducted in most programs certifying new staff was shorter than recommended by OHS, and in rare cases, new assessors were not formally certified.

For the second time, we conducted analyses that demonstrated a link between offering assessor training sessions and assessor performance on the assessment. Assessors in the three programs that failed to offer refresher training were significantly more likely to make administration errors, such as gesturing, non-neutral encouragement, and coaching. This year, we also compared administration errors in programs that offered comprehensive training to those in programs that offered less training than the OHS requirement (typically, omitting the role play activity).Assessors in programs with less than the recommended training were significantly more likely to make gesturing and scoring errors and to provide non-neutral encouragement than those in programs which offered comprehensive training.

Assessors from those lower training intensity programs also received significantly lower (by about 3 points) certification scores on the English assessment–another reason why recertification might be useful to correct administration problems before assessment begins (see recommendation below).

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. However, given that several programs specifically requested that training be shortened or limited to new staff only, or limited to the fall only (see Chapter VII), programs may resist new requirements that they perceive to increase the burden of the assessments.

Provide more guidance on how to interact appropriately with children during the assessment. According to program staff, assessments went more smoothly in spring than in the fall round, as children had gained skills, confidence with the assessment process, and comfort working with adults. Three quarters of the programs reported having some children with challenging behaviors, such as being bored/distracted, overly chatty, nervous, or anxious about their performance. Because some staff 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. Specifically, some assessors requested guidance on how to maintain a natural tone while staying on script. Practice and familiarity with the script is one way to increase comfort.

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-nine out of 40 programs assessed children in Spanish.[74] Spanish assessment certification scores have declined since spring 2006. The mean certification score was 89 (compared with 97 percent in 2006), and 60 percent of assessments (compared with 96 percent) were completed by assessors who had a certification score over 85. Among

[74] Three additional programs reported having at least one NRS-eligible child who spoke Spanish at home, but did not assess these children in Spanish either because of the lack of a Spanish assessor or because they deemed the child proficient enough in English to take only the English assessment.

Spanish assessors, the most frequent errors were in scoring, straying from the script, and gesturing.

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 adhering to the script, specific guidelines on appropriate substitutions may be needed, both in terms of the assessor's script and the child's responses. Incorporating these acceptable terms into the easel materials (rather than only in the separate Assessors' Guide) would be most helpful and would help to standardize acceptable regional responses.

Clarify expectations for growth in Spanish skills. Some assessors questioned why children were asked to respond in Spanish when English is usually the language of instruction in their programs (except in Puerto Rico).[75] 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 ENGLISH LANGUAGE LEARNERS

Clarify the best procedures for assessing children who speak a home language other than Spanish or English. Two programs (5 percent) that enrolled mostly children speaking languages other than English or Spanish reported problems with assessing these children, for whom neither assessment tool really captured their cognitive abilities. Children were uncertain about which language to use for their answers, assessors were uncertain about scoring answers that would translate correctly, and staff at both programs expressed a strong wish that the NRS have assessment options that were more appropriate for these children. Additional information on how to handle assessing children speaking a third language may be useful to these programs, but the NRS system was not designed to assess children unless they could respond in English or Spanish.

[75] ACF reports that OHS requires that programs help children make progress toward English acquisition, while being respectful of home language. ACF reports that different programs may implement this requirement differently, and many programs use multiple languages for instruction.

GUIDANCE ON ASSESSING CHILDREN WITH DISABILITIES

All but three programs in the sample assessed some children with identified disabilities during spring 2007. The majority of disabilities were speech and language delays. Forty-six out of 315 English assessments and 9 out of 98 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, programs were not always 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 children with autism, for example, or children who were nonverbal or had speech impairments. While many programs felt that they had received adequate training and guidance, staff from two programs complained about the inadequacy of information about assessing children with disabilities (a substantial decrease from 2006, when 13 programs made this complaint). 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 (though this was less prevalent in 2007 than in previous rounds of the QA study). For example, one program invited a speech therapist to be present to assist the assessor with understanding the responses of a speech-impaired child, and in another program assessors altered the script for a child with a behavior disorder. Programs need additional guidance on accommodations that conform to NRS rules of administration.

FEEDBACK ON THE SOCIAL EMOTIONAL DEVELOPMENT RATING FORM

Most programs did not offer concrete plans for using the SED data in the future. However, more than half of programs felt that the SED could make a contribution at the child, program, or national level. Some programs thought that the SED rating form complemented local assessments and screenings, and some also expressed appreciation that a social-emotional development component has been incorporated into the NRS. Moreover, some programs expected the rating form to inform local training, technical assistance, and program improvement efforts. Many programs felt that the items included in the rating form were comprehensive and well targeted.

Consider program concerns about the timing and burden of the SED rating form. Despite their positive overall reaction to the new form, local program staff raised a number of concerns. Many staff said that programs should receive SED results in a more

timely fashion and at the child level in order to use them for program planning and improvement.[76] In addition, some programs felt that guidance from ACF on how to address problem areas would be helpful. Local staff also expressed concern about whether the rating form imposed a burden on staff at the beginning and end of the program year; duplicated local assessment efforts; and contained items that were perceived by some staff as being too negative, subjective, or difficult to quantify based on the rating categories. Staff in a few programs also expressed concerns about the appropriateness of the rating form for children in tribal programs, ELLs, and children with disabilities.

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, we make the following suggestions for amending the assessment:

Consider amending the Language Screener and Vocabulary tasks and carrying out analyses to allay concerns about bias. For 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. On the Vocabulary task, 15 programs raised the issue of cultural appropriateness and regional bias of items on the test. 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 alternative approaches to the Letter Naming task. Twenty-eight percent of the programs in spring 2007 found the current format and directions for the Letter Naming task to be time-consuming and frustrating for the children, a decrease from previous rounds in which this was a concern for nearly half of programs. 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, which staff at some programs suggested including as acceptable responses. 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 in Spring 2007, we have suggested 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

[76] According to ACF, OHS told programs that reporting on SED would be delayed this first year due to the need to examine the data for reliability and potential irregularities before reporting out.

children assessed in Spanish) and (2) helping programs understand how to use and interpret data that they collect from regularly administered local assessments. Given that, as of the time this report is being finalized, the NRS has ended, we continue to underscore the importance of planfully collecting, interpreting, and using locally collected data on children's performance to improve practice. Through the ongoing Learning from Assessment task in the Quality Assurance Study, MPR continues to work with the Office of Head Start and expert consultants including Head Start program staff, to develop materials that will help programs use their own locally collected data more effectively.

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.

Administration on Children, Youth, and Families. "Design of the Head Start National Reporting System Child Assessment." Washington, DC: U.S. Department of Health and Human Services, 2007.

Administration on Children, Youth, and Families. "Head Start National Reporting System: Year One Implementation." Washington, DC. U.S. Department of Health and Human Services, 2007.

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].

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.

Vogel, Cheri, Renée Nogales, Nikki Aikens, and Louisa Tarullo. "Implementation of the Head Start National Reporting System: Spring 2006." Princeton, NJ: Mathematica Policy Research, Inc., February 2008.

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

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