The Head Start National Reporting System: A Critique

Samuel J. Meisels and Sally Atkins-Burnett

Initiated in the fall of 2003, a high-stakes achievement test is being administered to all four- and five-year-olds in Head Start. Two times a year, 35,000 teachers or surrogates will administer the 15- to 20-minute test to more than a half-million children, at a cost in excess of $16 million. The purpose of the test, as described by the Head Start Bureau, is threefold: (1) to enhance local aggregation of child outcome data and local program self-assessment efforts; (2) to enable the Head Start Bureau and Administration for Children and Families (ACF) Regional Offices to plan training and technical assistance efforts; and (3) to incorporate child outcome information into future Head Start program monitoring reviews. Never before in the history of this nation have so many young children been exposed to a standardized achievement test.

Unfortunately, this test, called the National Reporting System (NRS), includes items that are rife with class prejudice and are developmentally inappropriate. This is particularly troubling because the test is used by Head Start officials as a quality assurance system. In fact, the idea that a narrow test of young children’s skills in literacy and math can represent a quality indicator of a holistic program like Head Start shows a stunning lack of appreciation for the comprehensive goals of the 38-year-old program. Moreover, program quality cannot be evaluated by student outcomes alone, since this approach does not take into account differences among children and programs.

A problematic approach

The approach to program evaluation that the Head Start Bureau has chosen is extremely problematic. Researchers have demonstrated repeatedly that testing children using the design of the NRS—an isolated evaluation of children’s skills at two time points—is highly inaccurate and renders very poor predictions of future achievement (LaParo & Pianta 2000). Young children’s skills are in flux; this lack of stability renders the conclusions based on con-
ventional pre/post tests potentially misleading. Moreover, the Head Start population, composed primarily of children whose families have incomes below the national poverty line (approximately $18,500 a year for a family of four in 2003), is very diverse in terms of ethnicity, languages spoken at home, parental education, and concomitant opportunities for children to learn prior to entering Head Start. Both the National Education Goals Panel (Shepard, Kagan, & Wurtz 1998) and the National Research Council of the National Academies (Heubert & Hauser 1999) urge that children below age eight not be administered the kinds of tests that are represented by the NRS. Even the No Child Left Behind Act does not mandate testing until third grade. For many negative reasons, the NRS breaks new ground.

**Description of the test**

The NRS consists of five subtests: two measure English-language competence, one evaluates receptive vocabulary knowledge, one focuses on letter names, and the final subtest addresses mathematics.

The test begins with two subtests that are intended to determine if a child has sufficient mastery of the English language to take the rest of the test. If the child does not demonstrate this mastery, and is a Spanish speaker, a Spanish-language version of the test is administered. However, before the decision is made to administer the test in Spanish, the child must miss more than 70 percent of the items on these two subtests. This means the child experiences at least 15 failures. If the tester knows that the child comes from a Spanish-speaking family, but the child passes the language screen, the child is administered both language versions. If the child speaks any of the roughly one hundred other languages represented in Head Start, the child is not tested, because the test is available only in Spanish and English.

The vocabulary subtest is adapted from the Peabody Picture Vocabulary Test-III (PPVT-III), a test of receptive language skills such as listening comprehension. Children are shown a page with four pictures; the examiner says a word and asks the child to point to the picture that “best shows what the word means.” The NRS version contains 24 items. Although the measure includes items that appropriately ask children to identify body parts, animals, and actions, some items are less appropriate, including a number that are very class-biased. For example, for the word *ease* the incorrect illustrations, or test foils, include items that could actually all be used as vases.

An item that requires children to select the facial expression for *horrified* from an array of four faces is highly inappropriate for four-year-olds. The facial features of all four images are Caucasian, ignoring the fact that facial expressions differ in different cultural and ethnic groups. Also, by including this emotion, the test ignores extensive research indicating that the depiction of the facial expression for *horrified* (which researchers note is often confused with the expression of anger or rage) is not recognized by most children until much later in life (Izard 1977, 1991; DePaulo & Rosenthal 1982). This is the only emotional expression that children are asked to identify in the assessment.

When the entire PPVT-III is administered, differences in the general knowledge and experience of children taking the test are balanced by the broad range of vocabulary items on the test. In a subset of about a third of the items for the age ranges selected for the NRS (2½ to 9), such differences are highlighted; and these items clearly place children from impoverished backgrounds at a significant disadvantage. Further, the items are administered in an order that differs from the way they are presented on the original PPVT-III, which may result in a further loss of validity for this group of items. The evidence from the field test indicated lower reliability for this subset of items (below .80 for interrater reliability, or consistency, and for the Item Response Theory [IRT]-based statistical reliability when examined by age) than for the test as a whole.
learn—and provides limited information even about that—while ignoring research about the importance of phoneme and sound awareness. The math subtest is the most problematic, as it is more a test of language competence than quantitative skills. Approximately 25 percent of this subtest involves naming a number or pointing to a shape that is named. It also contains several if-then statements and terms like longer than and other comparatives. Further, the items are very poorly designed. For example, children are asked to count grapes, but some of the grapes are shown in a cluster, some individually. The point is to assess mathematical skills, not general knowledge.

The test does not provide a correct answer to one of the questions. It asks which coin from an array of four crayons is longer than a brush, but the brush that is shown is shorter than all the crayons. If the item had asked for a comparison of the crayons with the paintbrush, this question could have been answered unambiguously, but this is not what the item asks for. It requires a child to know that the word brush refers to the handle and brush as a single item. Children who are unfamiliar with English shorthand may be confused by the question and not be able to answer correctly. The item could be improved by using a pencil or something else with only one part.

Another anomalous item shows several coins and asks, “Which coin is smaller in size than the penny?” This question is extremely difficult because it requires that children make a distinction between the physical size of a coin and its monetary value, be familiar with five different coins (one of which is the relatively rare 50-cent piece), and know the meaning of which.

In general, the assessment of number concepts is very limited. Only two items require a child to have a concept of any number greater than three. There are no items that assess comprehension of pattern, number constancy, matching, classification, or estimation; spatial reasoning; or recognition of more and less, all of which are central to the development of young children’s mathematical thinking.

The if-then construction used in three questions (for example, “If you gave a friend one of these books, how many would you have left?”) is difficult for children to understand before elementary school and is particularly language driven. Most children are at least school age before they are competent in understanding sentences that are joined with such conjunctions as if, because, and unless (Gummersall & Strong 1999; Hulit & Howard 2002). The complexity of the sentence construction distacts from the assessment of mathematical thinking. The test’s measurement item that refers to inches is also not considered appropriate before second or third grade, according to the National Council of Teachers of Mathematics (2000) and other researchers (Kribs-Zaleta & Bradshaw 2003). The final item, which requires mastery of one-to-one correspondence while counting, is so difficult to score that reliability among examiners is likely to be very low.

Conclusion

In short, this test teaches us very little about young children’s preschool skills. It provides no authentic literacy evaluation and little information about math skills. Entire areas of development, such as social-emotional growth, physical development, science, social studies, the arts, and most of literacy and even phonemic awareness, are omitted. Basic concepts are included only in the initial language mastery items, and not only are these very limited, it is not clear that this subtest will be reported in terms of anything other than English-language mastery.

The vocabulary test, which contains many items that are culturally specific, may have secondary negative effects. Because so many Head Start teachers have restricted or narrow training (fewer than 30 percent hold a B.A.), they may consider the items on this test to be appropriate for all children to know and this may greatly distort their curriculum. As one trainer in a Head Start agency said, “This is the first time the Head Start Bureau has directed us to do something that can really do damage to children.”

The NRS is biased toward children from families with high socioeconomic status and Caucasian culture. Not only is it class- and race-oriented, it also reflects a singular pedagogical approach. It has few expressive and no creative, constructive, or active elements; it encourages a passive transmission model of education. It offers no problem-solving items, does not call for complex or higher-order thinking, and allows for few if any alternative answers. The lessons it teaches Head Start staff are negative regarding children’s potential, and it does not recognize the richness of the backgrounds that Head Start children bring with them. Any conclusions regarding the quality of Head Start that are based on data from this test will be tenuous at best.

Because the NRS will be used in making decisions about the continuation of Head Start, it is an example of a test in which indicators of learning can overwhelm learning itself, since potentially successful programs may be classified as failures based on an invalid and poorly constructed test. After all the experience this nation has had regarding the impact of measurement-driven instruction in the K–12 arena, it is baffling that a test like this would be visited on young children from poor households.

This test is not good early education practice. It is not good psychometric practice. It is not good public policy. And it is certainly not good for young children.

References


Copyright © 2004 by the National Association for the Education of Young Children. See Permissions and Reprints online at www.naeyc.org/resources/journal.
The Head Start Family and Child Experiences Survey (FACES), 1997–2022, is a longitudinal study of Head Start programs around the country. It describes the characteristics, well-being, and development of children and families at the beginning of the Head Start program year. A national portrait of the population Head Start serves provides information to consider when assessing needs and setting national goals for the program. A National Portrait of Head Start Children and Families: FACES 2014 (Feb. HSNRS is the acronym for Head Start National Reporting System. Share this. Have you found the page useful? Please use the following to spread the word: About | Contact Us Link to Us iOS app | Android Popular Abbreviations Popular Categories. Business Medical Abbreviations Military Abbreviations Technology Slang Terms. Types of Acronyms Acronym Generator. © 2005-2021 All Acronyms | Terms | Privacy & Cookies Policy. Search for abbreviation meaning, word to abbreviate, or category. Shortcuts for power users - examples. Head Start's National Reporting System: a work in progress. Pediatrics. 2004 Sep;114(3):858-9. doi: 10.1542/peds.2004-0707. State(s) of Head Start is the first report to describe and analyze in detail Head Start enrollment, funding, quality, and duration, state-by-state. The report focuses on the 2014-2015 program year but also provides longitudinal data beginning with the 2006-2007 program year. Despite the fact that Head Start is a federally funded, national program, the report reveals that access to Head Start programs, funding per child, teacher education, quality of teaching, and duration of services all vary widely by state. The authors call for an independent bipartisan national commission to study the issues raised in this report and develop an action plan to ensure every eligible child in every state has an equal opportunity to benefit from Head Start. Report Contents. Head Start's national numbers. $9,838,693,013. FY18 Federal Funding in the United States. “Evidence suggests Head Start not only helps kids develop and allows parents to go to work, but it may also help at-risk kids from ending up in the foster care system.” Head Start parents have steeper increases in their educational attainment by the time their children are six years old. State-by-State Fact Sheets. Access the state-by-state fact sheets. View citations and relevant notes here. National Fact Sheets. National Fact Sheet. Migrant and Seasonal Head Start Fact Sheet. Become a Member. Become part of the national Head Start community. Join or Renew now! New to Head Start?