



HARVARD EDUCATION LETTER

Testing Goes to Preschool

Will state and federal testing programs advance the goal of school readiness for all children?

by Robert Rothman

Since the fall of 2003, teachers in Head Start preschool programs around the country have been sitting down with their students and having conversations like this:

I want you to look at some pictures with me. I'm going to say some words. For each word I say, point to the picture that best shows what the word means.

Let's try one. Put your finger on "ball."

Good! Let's try another one. Put your finger on "dog."

Using exercises like the one above, Head Start teachers across the country have been asking their students to demonstrate their understanding of certain words, to identify letters, to recognize geometric shapes, and to solve simple addition and subtraction problems as part of the National Reporting System (NRS), a standardized assessment for Head Start instituted by the federal government.

The NRS reflects a nationwide movement toward testing in preschools that is gathering force state by state. In recent years, state-funded preschool programs have grown rapidly and are now in 40 of the 50 states, according to statistics by the Education Commission of the States (ECS). The number of children in state-funded pre-kindergarten programs rose from about 300,000 in 1990 to more than 700,000 in 2001, and spending more than doubled, from \$700 million to \$1.7 billion, during roughly this same period. Four other states provide supplemental funding for federal Head Start programs, and six states have no state-funded prekindergarten programs, ECS statistics show.

Along with increased enrollment in and funding for preschool at the state level came a call for greater accountability. "There was a sincere desire on the part of people investing public funds in early childhood education to see whether there are results," says Marilou Hyson, deputy executive director of the National Association for the Education of Young Children (NAEYC).

As a result, in recent years states have increasingly mandated assessments of preschool programs to provide information on children's abilities and track the progress of the programs. And, at the federal level, the NRS is by far the largest effort of this kind, mandating the evaluation of nearly every one of the 450,000 four- and five-year-olds in the Head Start program.

Editors' Note

Educators, researchers, and policymakers are all becoming increasingly aware of the role that high-quality early childhood education plays in the success of students not only in the elementary grades, but throughout their lives. With this in mind, the *Harvard Education Letter* is pleased to introduce a series of articles on education in preschool through grade three, made possible by the generous support of the Foundation for Child Development (FCD).

This month's cover feature by Robert Rothman on the rapid growth of testing in preschool is the first article in this new series, which will invite readers to consider far-reaching questions about the ways in which young children are educated in the United States. Our July/August issue will continue the series with a special issue devoted exclusively to preschool and early elementary education. And this summer, watch for the launch of a new website, supported by FCD, that will include all the articles in our series as they are published and other valuable resources on schooling in these critical early years.

Caroline Chauncey
Michael Sadowski

I N S I D E

a closer look

**The Three R's of One Model
Afterschool Program** 5

research analysis

**Can Brain-Based Software
Improve Reading?** 8



Harvard Education Letter

PUBLISHER
Douglas Clayton

EDITORS
Caroline Chaucey
Michael Sadowski

PRODUCTION MANAGER
Dody Riggs

ASSISTANT EDITOR
Elizabeth Barrett Kidder

FACULTY EDITOR
Richard F. Elmore

EDITORIAL ADVISORY BOARD
Katherine C. Boles, Lecturer, HGSE; Laura A. Cooper, Asst. Superintendent, Evanston Township (Ill.) High School; Linda Darling-Hammond, Professor, Stanford University; Sally Dias, Vice President for Programs and Partnerships in Education, Emmanuel College, Boston; Susan Moore Johnson, Professor, HGSE; Robert Kegan, Professor, HGSE; Peggy Kemp, Head of School, Fenway High School, Boston; Marya Levenson, Director, Teacher Education, Brandeis University; Deborah Meier, Co-Principal, Mission Hill School, Boston; John Merrow, President, The Merrow Report; Jerome T. Murphy, Professor, HGSE; Milli Pierce, President, Fund for Educational Excellence; Arthur J. Rosenthal, Publishing Consultant; Catherine Snow, Professor, HGSE; Jay Sugarman, Teacher, Runkle School, Brookline, Mass.; Ariadne Valsamis, Director of Program and Resource Development, John F. Kennedy Library Foundation.

Harvard Education Letter (ISSN 8755-3716) is published bimonthly by the Harvard Graduate School of Education, 8 Story St., Cambridge, MA 02138-3752. Second-class postage paid at Boston, MA, and additional mailing offices. Postmaster: Send address change(s) to Harvard Education Letter, 8 Story St., Cambridge, MA 02138-3752.

Signed articles in the *Harvard Education Letter* represent the views of the authors.

©2005 by the President and Fellows of Harvard College. Published as a non-profit service. All rights reserved. Special permission is required to reproduce in any manner, in whole or in part, the material herein contained.

HOW TO SUBSCRIBE
Send \$38 for individuals, \$49 for institutions (\$52 for Canada/Mexico, \$63 other foreign, in U.S. funds only). Subscription prices subject to change without notice. Single copies, \$7.00. Back issues and bulk subscriptions available at special reduced rates.

Address all correspondence to Harvard Education Letter, 8 Story St., Cambridge, MA 02138-3752; phone 617-495-3432 in Massachusetts, 800-513-0763 outside Mass.; fax 617-496-3584; email: editor@edletter.org; web: www.edletter.org.

Like the standards movement in elementary and secondary education, the use of standardized tests in pre-schools has generated heated debate among specialists in early childhood education. First, critics of preschool standardized testing have charged that the tests may not provide accurate information on children's abilities. Second, many worry about the impact of testing on program quality. By emphasizing such skills as letter and word recognition, opponents warn that tests like the NRS could direct teachers away from other essential areas of learning readiness, such as motivation and learning how to learn. And other important aspects of early childhood education, like health and social and emotional development, may not be addressed at all, since they are not included on many assessments.

"Policymakers may need help getting a better understanding of the components of program quality," Hyson says. "You can't use one—even one good—assessment to say whether a program is good or not."

Multiple Choices

Although the impetus for testing and accountability may be similar in K-12 and prekindergarten education, the two sectors vary substantially. Nationally, preschool education is highly diverse, with wide variation among programs, curricula, and assessments. For one thing, the private sector makes up a much larger proportion of the field in prekindergarten than it does in K-12. Of the 2.2 million three- to five-year-olds who are enrolled in center-based programs, about half are in private programs, which tend to be fairly small. These programs are generally exempt from testing requirements, according to Sharon Lynn Kagan, a professor of early childhood and family policy at Teachers College, Columbia University. "The vast majority of early childhood programs are not under public purview," she says. "These folks are barely accountable at all."

However, Kagan notes that private programs generally do assess students for instructional purposes. And she points out that an increasing number of private programs use packaged curricula, which often include assessments addressing some of the skills early childhood educators advocate for preschool programs, such as social and emotional learning.

States generally use two types of assessments, according to W. Steven Barnett, director of the National Institute for Early Education Research at Rutgers University. Some states, such as New Jersey, use standardized tests that provide general information on language and mathematics skills but do not provide fine-grained information that teachers can use to inform instruction. "It's like taking a temperature," Barnett says. "But state people are realizing it doesn't give a very precise picture."

At the same time, states like Maryland use portfolios and classroom observations as assessments to provide teachers with detailed information about children's devel-

opment. These assessments, while useful for instruction, often lack the technical qualities that enable the results to be aggregated to measure program quality, Barnett notes.

Yet even the states that use standardized tests have been reluctant to attach consequences to the results because many educators doubt that programs can be held accountable for children's performance, Barnett says. "To attribute the source of kids' gains to a program, as opposed to kids' families and communities—people are skeptical of that."

Holding Head Start Accountable

Issues of accountability are at the heart of the debate over the NRS. The Head Start assessment is aimed at providing comparative information about the performance of the 1,300 programs operated by the federal agency, according to Wade F. Horn, assistant secretary for children and families at the Department of Health

and Human Services (HHS), the agency that administers Head Start. The NRS grew out of 1998 legislation that set standards for student outcomes in the program. Originally, program administrators could decide for themselves how to assess these outcomes, but the Bush administration determined that the assessments may not have been reliable or valid and sought to measure outcomes against a common yardstick. HHS officials say that they are confident that the NRS will ultimately yield valuable information that will help them

identify the areas in which Head Start students and programs need additional help.

"The purposes of the system are, first, to help with educational planning, and second, to identify which programs may need additional training and technical assistance to achieve good outcomes for children," writes Horn in a department publication. "If a particular program is not achieving the kinds of results we all want for children enrolled in Head Start, the response will not be to de-fund the grantee, but to provide intensive assistance designed to increase the capacity of that program to help children achieve good outcomes."

Horn stresses that test results will not be the only factor in evaluating Head Start programs. "What if," he says, "we find a program where kids show no progress in an area we know is highly predictive of school success, like

"You can't use one assessment—even a good one—to say whether a program is good or not."

Respond to the Harvard Education Letter Reader Survey . . . for a FREE gift subscription!

Just fill out the reader survey enclosed with this issue and return it in the postage-paid envelope by April 15. To express our thanks, we will give a free one-year subscription to the individual or institution of your choice. Thank you!

A Decades-Old Battle

The debate over Head Start's National Reporting System is the latest skirmish in an often heated battle over testing young children. In the 1980s, as part of the education reforms enacted in the wake of the landmark report *A Nation at Risk*, states and school districts increasingly implemented testing programs for children as soon as kindergarten, or earlier. These programs included screening tests to determine whether children were ready to enter kindergarten or first grade and achievement tests for first and second graders, modeled on tests for older children, which were intended to hold schools accountable for student performance.

Early childhood educators reacted strongly to this trend. In a 1987 policy statement, the National Association for the Education of Young Children (NAEYC) cautioned against most forms of testing before age eight. The statement noted that tests are appropriate if they provide information that can contribute to improving outcomes for children, but warned that most standardized tests fail to meet this standard. Because of the wide variations in children's development, tests seldom yield valid results, the statement said: "Rather than use tests of doubtful validity, it is better not to test, because false labels that come from tests may cause educators or parents to alter inappropriately their treatment of children."

In place of standardized tests, early childhood educators argued for "developmentally appropriate" practices, such as informal assessments, including teacher observations and portfolios, that they consider sensitive to the way young children grow and learn.

The NAEYC statement, and the strong sentiment behind it, proved enormously influential. By 1996, according to a report by the National Education Goals Panel, "almost all state-mandated standardized testing for purposes of school accountability had been eliminated for children below grade 3." And states increasingly adopted informal assessments for preschool programs, according to research by the Erikson Institute, a Chicago-based graduate school in child development. However, the use of tests for accountability in early education began to reemerge in the late 1990s, in large part because of the influence of the standards-based reform movement in elementary and secondary schools. ■

For Further Information

C. Horton and B.T. Bowman. *Child Assessment at the Preprimary Level: Expert Opinion and State Trends*. Chicago: Erikson Institute, 2002.

National Association for the Education of Young Children. *Standardized Testing of Young Children 3 Through 8 Years of Age*. Washington, DC: Author, 1987.

vocabulary? We don't go in with guns blazing and say, 'You have a crappy program.' We'll try to understand the context. What is the curriculum related to vocabulary? Are the teachers well trained? When they read to kids, do they use that opportunity to expand their vocabulary? The NRS is a piece of information that allows us to ask more questions of a program."

The question of whether accountability will improve results as "underperforming" preschools work harder to raise students' test scores is deeply controversial, however. Early childhood specialists who are critical of the NRS insist that they are not opposed to testing or accountability. But like their counterparts in K-12 education, they are worried about using a narrow set of criteria to evaluate the quality of preschool programs.

"We worry a lot about the unintended negative consequences of choosing a set of narrow tools," says Hyson. "We know what happened in K-12: The curriculum is driven by assessment. If teachers in early childhood know children will be assessed on these dimensions, teachers, despite their best intentions, will teach to the test."

Teaching to the Test?

While few results from the NRS are currently available (the first year's administration of the tests was considered experimental), there are some preliminary indications that "teaching to the test" may already be occurring, according to Linda Espinosa, an associate professor of education at the University of Missouri who serves on an NRS technical advisory panel. Among Spanish-speaking children in English-language Head Start programs, Espinosa notes, understanding of English improved dramatically between fall 2003 and spring 2004. But Spanish vocabulary declined over this same time period, suggesting that many programs may have been focusing on English and neglecting the development of children's home language. This emphasis could have serious academic consequences for English-language learners over the long term, she notes.

"The first language has to continue to develop," Espinosa says. Without such development, she notes, students will miss out on key content knowledge that helps lay the groundwork for later learning: "By third, fourth, or fifth grade, you'll see an academic decline."

For Further Information



Education Commission of the States, 700 Broadway, Ste. 1200, Denver, CO 80203-3460; tel: 303-299-3600. www.ecs.org

W.F. Horn. "Improving Head Start: A Common Cause." *Head Start Bulletin* 76, 2003.

C. Horton and B.T. Bowman. *Child Assessment at the Preprimary Level: Expert Opinion and State Trends*. Chicago: Erikson Institute, 2002.

National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education. *Early Childhood Curriculum, Assessment, and Program Evaluation: Joint Position Statement*, November 2003. NAEYC, 1509 16th St. NW, Washington, DC 20036; tel: 202-232-8777. Available online at www.naeyc.org/about/positions/pdf/p scape.pdf

R. Rothstein. "Too Young to Test: Why We Need a Better Means of Evaluating Our Nation's Youngest Children." *The American Prospect* 15, no. 11 (2004): A12-A13. Available online at www.prospect.org/web/page.ww?section=root&name=ViewPrint&articleId=8774

Espinosa also says that assessments like the NRS, which focus largely on discrete pieces of knowledge such as vocabulary words, may encourage teachers to use inappropriate practices to raise test scores. She notes, for example, that the vocabulary tested on the NRS includes words like "penguin" and "knight" that are not part of the vocabulary of a typical preschooler, so teachers may feel compelled to drill students on those words. "Drilling children on words is not going to give them the underlying [literacy] competence," she says.

Former *New York Times* education columnist Richard Rothstein, who has reviewed research on early childhood testing, has similar concerns. In a recent issue of the journal *The American Prospect*, Rothstein notes that research does in fact show that preschoolers who have better letter naming and recognition skills tend to become better readers later on, but that these skills are best developed through "natural literacy activities," not drill and memorization: "[R]esearch showing that letter recognition predicts reading success is based on assessing children who learned letters through natural literacy activities, like having stories read to them or playing with picture books," Rothstein writes. "There is no evidence that memorizing alphabet letters out of context predicts later reading skill. But the test will lead teachers to spend more time on alphabet drills and less on reading—just the opposite of what Head Start needs."

Horn responds that the purpose of Head Start is to ensure that low-income children develop language skills, including vocabulary, so that they will be ready to learn when they enter kindergarten. Although drilling students on the words on the test would be "wrong," Horn says, vocabulary is a vital skill and should be taught in Head Start: "Some tests are appropriate to teach to."

Horn adds that he does not believe the NRS encourages teachers to neglect social and emotional development and other aspects of early childhood education that are not currently included on the assessment. In fact, this spring the Head Start administration is field-testing measures in social and emotional development for possible inclusion in future versions of the tests. But Horn points out that the measures will only be included if they prove valid and reliable. "We'll include them in a thoughtful way, based on what I believe are high scientific standards," Horn says. "The only thing worse than no data is garbage data."

Toward Comprehensive Assessment

In response to the recent burgeoning of preschool tests, early childhood educators have begun to articulate what they believe does and does not constitute the appropriate testing of preschoolers. Advocates like Hyson see the need for comprehensive measures that can provide more detailed information on children's school readiness and program quality. "We can very reliably assess whether kids know the alphabet and whether children can match letters and sounds," she says. "But that's not the whole readiness story. We still lack good assessments in the social and emotional arena."

In November 2003, the NAEYC and the National Association of Early Childhood Specialists in State Departments of Education released a statement urging the use of "ethical, appropriate, valid, and reliable assessment." The statement urges assessments that are "developmentally appropriate, culturally and linguistically responsive, tied to children's daily activities, supported by professional development, inclusive of families, and connected to specific, beneficial purposes" such as instructional improvement and the identification of special needs.

"We tried not to take a negative, oppositional attitude against any assessment," says Hyson. "Assessment is an essential part of high-quality programs."

Hyson also warns that the emphasis on testing could divert resources and attention from other improvements in preschool education, such as expanding access and training and recruiting highly skilled teachers: "You don't create high-quality programs simply by testing children." ■

Robert Rothman is a principal associate at the Annenberg Institute for School Reform at Brown University and the editor of Voices in Urban Education.

Maryland's Work Sampling System

A 2003 statement by the National Association for the Education of Young Children (NAEYC) called for a comprehensive approach to early childhood testing that takes into account a variety of children's developmental, cultural, and linguistic needs (see main article). Although the statement does not identify exemplary programs, one that appears to match many of the NAEYC criteria is Maryland's. There, state officials began by identifying a set of outcomes for young children and developed intensive professional development around those outcomes. The state then selected an assessment, known as the Work Sampling System, that was aligned to the outcomes.

The Work Sampling System assesses children's development on a range of dimensions, including language and literacy, mathematics, science, social studies, the arts, social and personal development, and physical development. Unlike traditional standardized tests, it relies on teachers' documentation of children's performance and behavior and portfolios of children's work. These components can be aggregated to provide information to policymakers and the public, and they provide useful information for teachers, notes Rolf Grafwallner, coordinator of the early learning office in the Maryland State Department of Education.

"If you look at the research on early childhood education, you look for children's performance in their natural learning situation," he says. "Over time, you have multiple points of information on children's learning. That's helpful to teachers." ■