

Tracking Progress Toward the School Readiness Goal

It's time to design new forms of school readiness assessment, forms that do not encourage tracking of students, narrowing the curriculum, or kindergarten retention.

"By the year 2000 all children will start school ready to learn."¹

The National Governors' Association (NGA) set this national school readiness goal at its meeting in February 1990. At first glance, the statement may appear to be a call for a nationwide school admissions test for 1st grade, but if you know what lies behind this sentence, you'll see that nothing could be farther from the truth. In fact, this statement helped focus public attention on the compelling arguments *against* group-administered school readiness tests.

Widely publicized in the literature, these arguments have been debated in numerous forums and forged into policy statements by at least a dozen education organizations. Three general premises emerge: group-administered pencil-and-paper readiness tests are inappropriate for preschoolers and lack sufficient validity for making school entry decisions (Meisels 1989, Meisels et al. 1989, NAEYC 1988); their use often has the effect of narrowing the preschool curriculum and making it excessively academic (Shepard and Smith 1988, Bredekamp and Shepard 1989); and test scores, when used to deny school entry, contribute to the

practice of kindergarten retention, which is counterproductive public policy (Shepard and Smith 1989).

At the NGA meeting, the President and the governors warned that an assessment for school readiness should *not* be developed for purposes of measuring progress toward the readiness goal, because of the danger that it could be wrongfully used to determine when a child should start school. "Other current indicators of readiness may serve as proxies," they noted, "and still others need to be developed."²

The governors suggested that new methods be developed using teachers' cumulative observations of children.³ At a minimum, this would yield better information to improve learning. It could possibly produce meaningful data for policymakers as well.

The Limitations of Proxy Measures

Figure 1 lists specific objectives toward meeting the readiness goal. These objectives concern provision of quality preschool programs, prenatal nutrition, and health care, as well as

Fig. 1. National Goals For Education Readiness

Readiness Goal 1: By the year 2000, all children in America will start school ready to learn.

Objectives:

- All disadvantaged and disabled children will have access to high-quality and developmentally appropriate preschool programs that help prepare children for school.
- Every parent in America will be a child's first teacher and devote time each day helping his or her preschool child learn; parents will have access to the training and support they need.
- Children will receive the nutrition and health care needed to arrive at school with healthy minds and bodies, and the number of low birthweight babies will be significantly reduced through enhanced prenatal health systems.

the involvement of parents as their children's teachers. Policymakers will use "proxies," defined as indicators that do not measure children directly but that indirectly measure factors positively associated with readiness, in order to track national progress toward these objectives.

The National Center for Education Statistics (NCES) is now assembling these proxy data from various government-sponsored national surveys. These include national statistics on prenatal care; low birthweight babies; child nutrition; the percent of eligibles served by subsidized preschool programs; the supply, demand, and quality of preschool services; the kinds of learning experiences provided; the extent of parental involvement; and child retention in early grades. NCES will rely heavily on the National Household Education Survey, a planned triennial telephone survey to a national sample of households, for the ongoing tracking of the education-related information, including data on parents as their children's teachers.⁴

Proxies, however, are not adequate readiness measures. While appropriate now as an interim method for tracking progress toward the readiness goal, they are nonetheless insufficient for this purpose. They provide an important piece of the puzzle by periodically informing us about factors, including essential services to children, that enhance early learning. But they do not tell us the extent to which our children have the requisite skills, attitudes, or behaviors they need and, therefore, what we need to know to help them succeed in school. Knowing the percentage of low birthweight babies each year is important, but this won't tell us why children aren't learning to read. Making high-quality preschool programs available to all needy children is unquestionably desirable; but paradoxically, to obtain good "proxy" data on the overall quality of such a program, you need an outcome measure of its effect on children.

The Office of Educational Research and Improvement (OERI) acknowledges "the proxy measures based on preschool enrollment, parental involvement, and characteristics of pre-

Knowing the percentage of low birthweight babies each year is important, but this won't tell us why children aren't learning to read.

school programs do not suffice to make inferences about actual readiness to begin school."⁵

With good, and more direct, measures of children's behavior, we will know better, for example, if certain kindergarten teaching strategies are differentially effective for youngsters with limited English proficiency, a particular type of disability, or a certain learning style. If we increase funding or redesign a program, we need to know the effect of such changes. With only proxy data to report, we can say we doubled the budget and we're serving twice as many, but we can't say if the program made a difference in children's learning.

Real improvement in a child's readiness to learn and a teacher's readiness to teach is enhanced, not by having macro-level data on the national percent of eligible participants in a program, but by understanding behavior within the personal microcosm of that teacher with that child, as they interact with parents and other children.

The Fear and the Need

In discouraging the development of a national readiness test, the governors did the right thing. Their decision reflects concern in the professional

community that national use of an inappropriate assessment, even if administered to only a sample of children, could lead to a highly academic preschool curriculum and ultimately to denying children the right to enter school with their age-mates.

Their decision is, in fact, only one of several recent behavioral and policy changes that discourage the use of school readiness tests. Several states have discontinued mandated kindergarten or early grade testing, and pre-K through 1st grade programs were recently granted a statutory exemption from Chapter 1 testing requirements. Further, informal, off-the-record communications with publishers reveal that sales of group-administered tests for the early and preschool grades have either remained flat or started to decline.

Assessment of young children, however, hasn't gone away, and it probably won't, because assessment—*good assessment*—is needed more than ever. In spite of the decrease in legislative requirements and a broad awareness of the dangers of misusing readiness tests, the pressure for "results" from states and districts for accountability purposes is increasing.

The Southern Regional Education Board (SREB), for example, now reporting annually on the progress of its 15 member states toward Year 2000 education goals, includes results of readiness assessments among the indicators of progress toward its school readiness goal.⁶ And as one measure of the quality of programs, SREB includes "the use of assessments or tests of readiness for young children."⁷ Legislation introduced in the U.S. Senate in 1990 requires that all relevant data on school readiness be considered for inclusion in a national report card on the educational goals.⁸ Good readiness measures are needed not only for accountability purposes, but also for diagnosing learning needs, planning appropriate interventions, and evaluating programs.

A Call for Reform

Because of what Sharon Kagan, Yale University's Associate Director of the Bush Center in Child Development

and Social Policy, calls a "fortuitous collision of events," a unique opportunity is now present for a major national reform in the assessment of young children. Our increased need for information to help children begin to learn has combined with our heightened awareness of the failings of readiness tests, so that "the seeds are right for change."⁹ Now is the time to invest in a reform of this country's school readiness assessment.

The reform should begin with an effort to separate school entry from readiness. Many other nations have specific school starting points. In Sweden, children start school at age seven; in Japan, West Germany, and Switzerland at six; in New Zealand at five; and in Britain and Australia as early as four. The concept of readiness, wrongly applied, can be used to keep the "unready" out and thus deny those who need it most an opportunity to engage in learning at school. The dangers of wrong school entry decisions, whether on the basis of test scores or not, and the increasing practice of voluntary retention by overzealous parents, would be greatly reduced if, at the established age, all children entered school, "ready" or not. As Meisels says, "If you're alive, you're ready to learn, no matter what the tests say."¹⁰

The next step of assessment reform should be to reach consensus on the expectations we have for young children. We need to know what we want to "ready" children and teachers for. That means defining the skills, behaviors, and attitudes children should learn and teachers should teach in the 1st grade of school. Perhaps some of our nation's best minds, most experienced teachers, and most knowledgeable parents—those who understand child development and its many dimensions, its wide variations and its spurts and starts—could reach consensus.¹¹ Perhaps the encroachment of increasingly higher levels of academic demands into the preschool and kindergarten could be reversed by publicly stating just what we expect *and do not expect* of 1st graders. The precursors to these expectations then could become our readiness indicators.

A Kinder, Gentler Way

We need to develop better means of assessing children's readiness than group-administered readiness tests. The governors, in fact, suggested several strategies for states to use to improve assessments:

"Develop assessment systems for young children that reflect the ultimate goals of producing independent, creative, and critical thinkers. Train teachers to observe and assess children's work in different content areas, using methods such as portfolio systems, observational checklists, and cumulative sampling of children's work. Develop models to use teacher assessments of student proficiencies for reporting to parents and the public."¹²

We know a good deal about assessing readiness from work already begun (Anderson 1987, NAEYC 1988, Meisels 1989, Meisels et al. 1989). This work tells us that, ideally, readiness assessments should:

- encompass the multiple dimensions of readiness, including cognitive, social/emotional, attitudinal, and physical/motor behaviors;
- be an ongoing process of observing a child, rather than a one-time snapshot;
- gather information on behaviors that children have had an opportunity to develop;
- provide data useful for instructional improvement—to help the teacher get ready for the child;
- be indirect measures of children—that is, recorded by an adult, rather than directly by a child on an answer sheet;
- be conducted in a natural setting that is comfortable, familiar, and nonthreatening to the child;
- be administered to individuals, one-on-one, or to very small clusters of children, but not to large groups;
- be designed so that children can respond by pointing, acting, doing, or manipulating;
- be conducted by someone who is properly trained and who can relate well to children;
- be scored to yield a profile along the various readiness dimensions;
- be used in a nonpunitive way—that is, not for sorting, tracking, or denial of school entry. This last point is most important.

Promising Prospects

In this section, I want to highlight three examples of developmental work in progress that appear to offer particular promise: one is from the State of Georgia, one from a test publisher (CTB Macmillan/McGraw-Hill), and one from research scientist Samuel Meisels.

The State of Georgia has now produced a developmentally appropriate method for evaluating readiness called the Georgia Kindergarten Assessment Program (GKAP). This program represents a positive directional change from Georgia's 1988 group-administered, machine-scored, norm-referenced test used for making school entry decisions (which probably did more to advance readiness assessment reform in this country than all other causes combined).

GKAP is a homegrown individually-administered readiness assessment on which three times a year, teachers record children's behavior across five capability areas. A series of "structured assessment activities" involving children's use of manipulatives gives additional information on two of the dimensions. A videotape illustrates the expected behavior for standardization of teacher ratings. The process relies heavily on teacher judgment and yields good diagnostic information; however, its value for accountability purposes is not yet clear. A mechanism has been designed, but not totally implemented, for aggregating data from classrooms, to schools, to districts, to the state. Teachers record *yes/no* data, but not a total score, for each child on each of the five capabilities, on scannable forms for automatic generation of totals. The new system has been very costly, and the price of maintaining it is not yet known. Further refinements need to be made and analyses conducted.

CTB Macmillan/McGraw-Hill has recently produced a "Developing Skills Checklist" for children aged 4 to 6, which is individually administered and is packaged in a kit with attractive manipulatives. For each child, teachers can generate a criterion-referenced diagnostic profile of performance in eight areas. These checklists produce

normative scores on mathematical concepts and operations, language, memory, auditory skills, print concepts, and a prereading composite. National percentiles, stanines, and normal curve equivalents are provided for each scale for four time periods (spring of pre-K and fall, winter, and spring of kindergarten). Included in the process are a method for recording multiple observations throughout the year on seven clusters of social/emotional behavior and a mechanism for compiling group data. So far, researchers have found high internal consistency reliability, and a two-year predictive validity study is now under way.

Samuel Meisels, Professor of Education and Research Scientist at the University of Michigan, is currently developing and piloting in nine school districts a new three-part readiness assessment process consisting of (1) a comprehensive criterion-referenced checklist of developmentally-based classroom learning, completed by the kindergarten teacher for each child on three different occasions throughout the school year; (2) the compilation of a portfolio of samples of the progression of each child's work; and (3) a summative teacher report form for providing year-end comparative, and possibly scaled and aggregate, data.

These three efforts suggest that supportive, child-friendly, and learning-enhancing measures can also produce aggregate data for accountability to the public. Since the governors recommended that a readiness assessment *not* be developed to measure progress toward the national goal, the government will presumably seek methods for indirectly compiling information gathered by others, such as teachers or parents, who have observed the behavior of children. The efforts I have described suggest the kinds of data that might be compiled.

Educators need more information, however, to ensure that the data compiled are valid, reliable, and useful and that they are collected in a uniform and equitable manner. It is essential to assure that observation criteria are clearly specified, that representative samples of all categories of behaviors are observed, and that raters are consistent. Test developers must create

effective methods for recording, compiling, analyzing, scaling, and reporting data. They must conduct field tests and validity studies to determine whether performance on a readiness assessment predicts success in school, and evaluate the usefulness of aggregate scores reported to policymakers.

Any new assessment method must be monitored for possible inappropriate uses, such as retaining preschoolers, tracking, narrowing, further academic loading of the curriculum, or "teaching to the test," and evaluated for practical considerations such as costs and time. Finally, test users should compare the overall value of a new assessment method with existing practices.

Having been instructed by the White House to develop "other" readiness measures, OERI should lead the way toward better school readiness assessment by underwriting the consensus-building process and sponsoring needed research and development. Whether or not a method better than proxies is ever created for national monitoring, readiness assessment in this country needs to be reformed. We need a kinder, gentler way to measure. □

¹National Governors' Association, *National Education Goals*. A statement adopted by the members of the NGA in Charlottesville, Virginia, on February 25, 1990, p. 3.

²The White House, *National Education Goals*. Press release from the Office of the Press Secretary, January 31, 1990, p. 3.

³National Governors' Association, *Educating America: State Strategies for Achieving the National Goals*. Report of the Task Force on Education. (Washington, D.C.: NGA, 1990), p. 15.

⁴See B.J. Turnbull, (April 1990), "Readiness for School—Issue Brief." Washington, D.C.: Policy Studies Associates, and Office of Educational Research and Improvement (OERI), "Education Summit Follow-up, Special Set, 6/190," internal document, for a more complete delineation of all the proxy data sources mentioned here.

⁵OERI, *Ibid.*, p. 4.

⁶J. D. Creech, (1990), *Educational Benchmarks 1990* (Atlanta: Southern Regional Education Board), p. 6.

⁷Southern Regional Education Board, (1989), *Reaching the Goal of Readiness for School* (Atlanta: SREB), pp 11–12.

⁸S 3095, The National Academic Report Card Act of 1990, introduced by Senator

Jeff Bingaman on September 24, 1990.

⁹S. L. Kagan, (July 13, 1990), personal communication, The Bush Center in Child Development and Social Policy, Yale University.

¹⁰S. Meisels, (July 5, 1990), personal communication, Center for Human Growth and Development, University of Michigan.

¹¹See National Association for the Education of Young Children, (1990), "Position Statement on Readiness," for a good discussion of what is meant by readiness.

¹²National Governors' Association, *Educating America: State Strategies for Achieving the National Education Goals*, *Ibid.*

References

- Anderson, L. W. (1987). "Comments on a National Assessment of School Readiness." In *The Assessment of Readiness for School: Implications for a Statistical Program—Report of a Planning Conference*. Washington, D. C.: Center for Education Statistics, pp. 27–36.
- Bredenkamp, S., and L. Shepard. (1989). "How Best to Protect Children from Inappropriate School Expectations, Practices, and Policies." *Young Children* 44, 3: 14–24.
- Meisels, S. J. (1989). "High-Stakes Testing in Kindergarten." *Educational Leadership* 46, 7: 16–22.
- Meisels, S. J., D. Steele, and K. Quinn. (1989). "Testing, Tracking, and Retaining Young Children: An Analysis of Research and Social Policy." A commissioned paper prepared for the National Center for Education Statistics, December 1989.
- National Association for the Education of Young Children. (1988). "Position Statement on Standardized Testing of Young Children 3 through 8 years of age." *Young Children* 43: 42–47.
- Shepard, L. A., and M. L. Smith. (1988). "Escalating Academic Demand in Kindergarten: Counterproductive Policies." *The Elementary School Journal* 89: 135–145.
- Shepard, L. A., and M. L. Smith. (1989). "Academic and Emotional Effects of Kindergarten Retention." In *Flunking Grades: Research and Policies on Retention*, edited by L.A. Shepard and M.L. Smith. Philadelphia: The Falmer Press, pp. 79–107.

Penelope Engel is a Professional Associate of the Educational Testing Service, Suite 475, 1825 Eye St., N.W., Washington, D.C. 20006.

Copyright © 1991 by the Association for Supervision and Curriculum Development. All rights reserved.