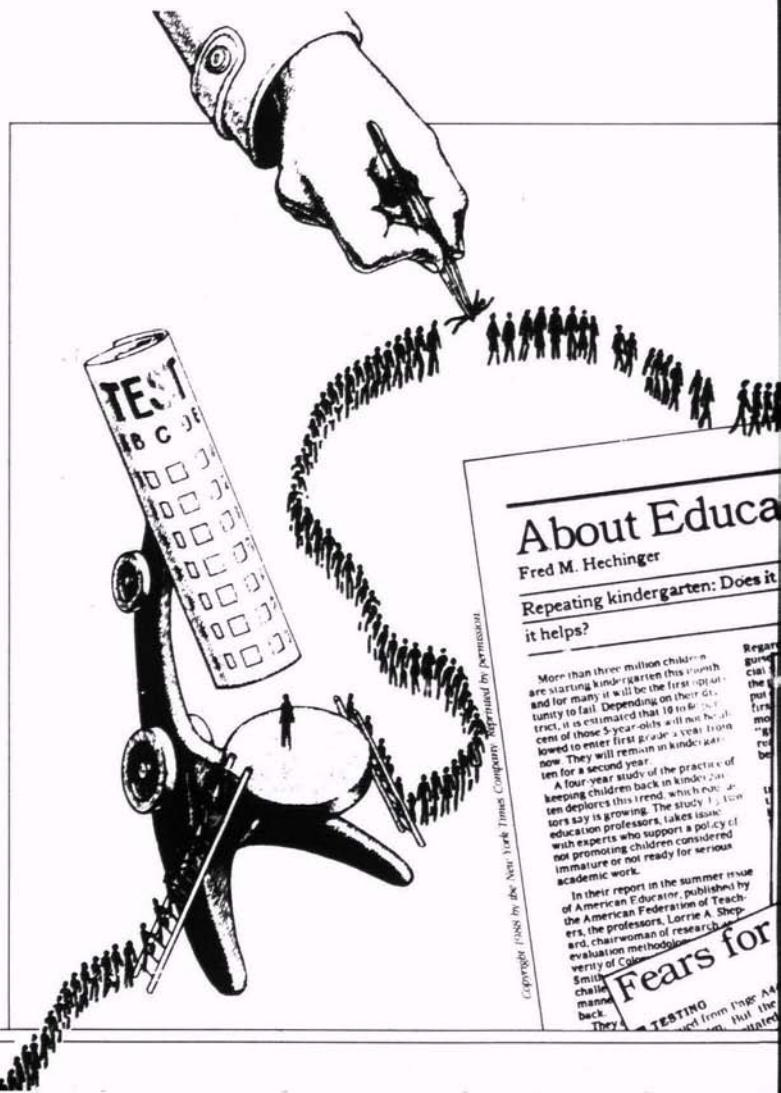
High-Stakes Testing

The use of readiness tests to classify, promote, and retain children is inappropriate and costly; we can help young students begin school more successfully by reformulating curriculum and classroom organization.



In increasing numbers, school districts are adding an additional year at the outset of children's school careers, instituting extensive policies of kindergarten retention, and establishing prekindergarten "readiness" programs and pre-first grade "transition" programs for children deemed "not ready" for traditional school-entry programs. Typically, the decisions to place children in these programs are based on the inappropriate use of tests. Here I examine the relationship between testing and extra-year early childhood programs and propose alternative strategies to help young children begin school more productively.

Measurement-Driven Instruction

Previously, when teachers sought to evaluate children in child-centered programs, readiness and standardized tests were criticized as irrelevant and unhelpful (see Bryk et al. 1979, Carini 1975, Hein 1979). But now testing has become much more prevalent in public schools generally and in kindergarten programs in particular (Riley et al. 1988, Meisels et al. 1988, and E. Fiske 1988).

It is the schools, not the children, that have changed, partly in response to demands for accountability. The pressure for teachers at each grade level to be held accountable has resulted in "academic trickle-down," a major influence on teachers in earlier grades (Cunningham 1988). Their decisions about what and how to teach

n Kindergarten

Artwork by Marlee Ilic copyright 1988 by The New York Times Company. Reprinted by permission.

Copyright 1988 by the Time, Inc., Magazine Company. Reprinted by permission.

Education

Can Kids Flunk Kindergarten?

Yes, sir—especially where the law mandates tests for first grade

As the crisis season for college acceptances climaxed across the country last week, 102,000 Georgia youngsters sat through a multiple-choice test to determine whether they too would qualify for a higher level of education. In a series of sessions totaling 90 minutes, they scribbled their answers. Then they went home, where their families are sweating out the mid-May announcement of whether they would gain admittance—to the first grade.

Yes, first grade. This year Georgia became the first state to require a standardized written exam as part of a "readiness assessment" that determines who passes and who fails Kindergarten. Testing of various kinds is prevalent in three-fourths of the other states for evaluating aspiring first-graders. In Minneapolis, for example, kindergartners must answer a 56 verbal questions for exam-

Copyright 1988 by the Boston Globe. Reprinted by permission.

ndergarten: Who is ready?

By Phyllis Coons
Globe Staff

RANDOLPH – Are some 5 year old children too young for kindergarten? Or, putting it another way, has kindergarten become too hard for them?

At the 26th annual Lesley College ew England Kindergarten Confer- e earlier this month, two early- dhood specialists argued this cur- controversial issue. The spe- held two differing views.

hildren who think readiness kindergarten may be the need it the most and should muel J. Meisels, professor on at the University of

ung son

Fears for a son going into a test-crazy world

COMMENTARY

By Robert P. Yagelski
Special to the Globe

Copyright 1988 by Robert P. Yagelski

are strongly influenced by the need for their students to perform well in the next grade level, as indicated in part by test results. In other words, teachers are very likely to shape their instruction to match a test's specific focus (see Darling-Hammond and Wise 1985).

This phenomenon, known as "measurement-driven instruction" (Madaus 1988), transforms testing programs, ideally servants of educational programs, into masters of the educational process. The results are a narrowing of the curriculum, a concentration on those skills most amenable to testing, a constraint on the creativity and flexibility of teachers, and a demeaning of teachers' professional judgment. These outcomes represent a vast alteration in educational policy, aided and abetted by the inappropriate use of tests; they are creating an emerging crisis in public early childhood education.

High-Stakes Testing

The use of readiness or achievement tests for the classification, retention, or promotion of students qualifies them as high-stakes tests: "Those whose results are seen—rightly or wrongly—by students, teachers, administrators, parents, or the general public, as being used to make important decisions that immediately and directly affect them" (Madaus 1988, p. 87). Three specific characteristics of high-stakes tests can be described.

Perceptions. First, because of the importance ascribed to high-stakes tests, their original purposes are often blurred, and their results may assume greater significance than is warranted. The SATs are the best example of this phenomenon. That is, a test intended to provide supplementary information for decisions regarding college admission has become an absolute criterion for many of those decisions and, what's more, a barometer of the entire nation's educational status.

Instruction. The corollary to this phenomenon is that high-stakes tests

heavily influence teachers' behaviors and their instructional decisions. If teachers know their pupils will be tested on certain skills or subject areas, and if the results of the examinations will be made public, then what they teach will likely reflect test-specific items or formats. High-stakes achievement tests invariably narrow instruction and learning, focusing the curriculum on the content that will be included on a test.

Decision making. Third, high-stakes tests transfer control over the curriculum to the agency that designs or controls the exam (Madaus 1988). Although there are many ways to learn to read, for example, if a school district adopts a test structured around a particular approach, teachers may feel enjoined to teach that approach. Hence, instructional decision making is removed from the arena of teacher-child interaction and supplanted by the instructional approach implicit in the test.

Examples of High-Stakes Testing

In early childhood programs, many tests have achieved high-stakes status. Here I briefly review two tests and one state testing program to illustrate the impact of such testing in the early childhood years.

The Gesell School Readiness Test. Madaus (1988) suggests that the power of high-stakes testing is "a perceptual phenomenon: if students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false—the effect is produced by what individuals perceive to be the case" (p. 88). This principle is clearly embodied in the widespread adoption of the Gesell test. Problems with the Gesell test (Haines et al. 1980) are extensive and have been described at length elsewhere.[1] The test's principal fault lies in the discrepancies between its stated purposes and the empirical evidence available to support those statements. Its developers claim that it can identify children who are at high-risk for school failure and that it can be used to determine when children should begin school, which children should be promoted, and which should be retained in grade. Clearly, the Gesell is a high-stakes test.

Unfortunately, there are no data to support these assertions. In a study that paradoxically claims to validate the Gesellian concept of developmental age, Wood, Powell, and Knight (1984) found that more than half of those kindergarten-age children considered "ready" by the Gesell did not have successful kindergarten experiences, as reported by their classroom teachers. A second study by May and Welch (1984b) also revealed major problems with the Gesell's accuracy and found no support for the effectiveness of an extra-year program recommended on the basis of Gesell test results. (Other studies with similar results are reviewed in the references cited in endnote 1.)

Yet the test continues to be widely used—perhaps because of the unfounded perception that it is efficacious and because it provides a means for teachers to cope with "academic trickle-down." In other words, if, as the Gesell theorists claim, their test measures "developmental age," which is maturationally driven and genetically derived, a child who cannot cope with an academically oriented curriculum does not necessarily represent a failure on the part of the child, teacher, or parent. Rather, the child is simply not "ready," and no amount of instruction, intervention, or effort can be expected to have an effect.

This position assumes that readiness is an absolute concept, not a relative one. But Bruner (1966) notes that the idea of readiness is a "mischievous half-truth . . . largely because it turns out that one *teaches* readiness or provides opportunities for its nurture, one does not simply wait for it" (p. 29). Readiness, in other words, is *modifiable*. When the Gesell test is used to define readiness, the concept of readiness is misrepresented; and the stakes are very high indeed.

The Georgia plan for kindergarten testing and retention is the *reductio ad absurdum* of high-stakes testing, where an entire state has transferred control over its early education program to a single group-administered paper-and-pencil test.

The Brigance K and 1 Screen. Consistent with Bruner's perspective, the purpose of readiness tests is to evaluate a child's *relative* preparedness to profit from a specific curriculum (see Meisels 1985, 1986). Most readiness tests are *criterion-referenced*, where a particular score indicates a specific level of concurrent performance mastery. In contrast, norm-referenced tests are interpreted on the basis of a child's standing in relationship to a larger population or group of children (see Angoff and Anderson 1967, Barnes 1982). Predictions about future performance can be made based on this standing. Thus, the basic purpose of criterion-referenced tests is to measure current achievement, not to predict future performance. Therefore, the use of criterion-referenced readiness tests for purposes of classification, retention, and promotion is unjustified.

The misuse of the Brigance K and 1 Screen (Brigance 1982) exemplifies this error. The Brigance is a brief criterion-referenced readiness test designed to provide a general picture of a young child's language develop-

About Education

Fred M. Hechinger

Repeating kindergarten: Does it hurt more than it helps?

More than three million children are starting kindergarten this month and for many it will be the first opportunity to fail. Depending on their district, it is estimated that 10 to [illegible] percent of those 5-year-olds will not be allowed to enter first grade a year from now. They will remain in kindergarten for a second year.

A four-year study of the practice of keeping children back in kindergarten deplores this trend, which educators say is growing. The study, by two education professors, takes issue with experts who support a policy of not promoting children considered immature or not ready for serious academic work.

In their report in the summer issue of American Educator, published by the American Federation of Teachers, the professors, Lorrie A. Shepard, chairwoman of research and evaluation methodology at the University of Colorado, and Mary Lee Smith, of Arizona State University, challenge the reasons for and the manner of holding kindergarteners back.

They concede that the policy is a well-intentioned one and say its advocates believe that they have the children's best interests at heart. The advocates also want to prevent future failure in school.

But Professors Shepard and Smith say their research has found that keeping children back in kindergarten "does nothing to boost their subsequent academic achievement."

Regardless of [illegible] the decision is disguised, they [illegible] it "creates a social stigma. [illegible] addition, they say, the pr[illegible] encourages the schools [illegible] put [illegible] more academic pressure [illegible] f[illegible] ers. After all, more and [illegible] youngsters are expected to "graduate" from kindergarten with reading and writing skills that used [illegible] be taught in first grad[illegible]

The issue is a cha[illegible] traditional child ex[illegible] traumatic for child[illegible] into first grade bef[illegible] But Professors [illegible] [illegible] k wh[illegible] r it is [illegible] atic [illegible] as the[illegible] adv[illegible] are als[illegible] princip... teachers, to p[illegible] ma[illegible] sters from f[illegible] ing ki[illegible] put excessive pre[illegible] teners to advance [illegible] the schools' terms.

No matter how the extra year in kindergarten is explained and how carefully school officials choose the euphemisms for being held back, parents find it hard to say on the cocktail-party circuit or in supermarket conversation why their child was held back. In an atmosphere of competitiveness, the danger is that kindergarten will become more and more like first grade.

"The current fad to flunk children in kindergarten is the product of an inappropriate curriculum," the report says. "Over the past 20 years there has been a persistent escalation of academic demand in kindergarten [illegible] were formerly [illegible] are shoved [illegible] rade."

[illegible] at a child [illegible] mates [illegible] back was [illegible] school later. [illegible] chers ac[illegible] ar programs [illegible] children [illegible] flexible, [illegible]," the re[illegible] adde[illegible]

They warn th[illegible] ore and more "unready" chi[illegible] kept out, first [illegible] becau[illegible] for 7-year-[illegible] ll of [illegible] all children [illegible] skills are [illegible] ng and

[illegible] by in[illegible] competi[illegible] ir children en[illegible] indergarten, [illegible] ir future [illegible] Aptitude Test and [illegible] college. They want the[illegible] be pushed academically.

Such a[illegible] atched by national de[illegible] turn to basics" [illegible] to the Japanese [illegible] ll and crammi[illegible]

There [illegible] memory of what [illegible] know[illegible] promo[illegible] ich mov[illegible] from o[illegible] to the next regardless of their academic achievement [illegible] was a policy that made it possible for students to get into, and graduate from, high school without being able to read and write.

When this policy was abolished in the 1950's, it was often replaced by

ment, motor ability, number skills, body awareness, and auditory and visual discrimination. The inventory consists of several traits, skills, and behaviors that children demonstrate at different ages. Nevertheless, the Brigance is in wide use nationally to "rank or group children who are high, average, or lower than their local reference group in order to contribute to readiness decisions, to make placement decisions, and to serve as an indicator for more comprehensive evaluation or referral for special services" (Boehm 1985, p. 224). To fulfill these purposes, a test must be norm-referenced and valid. However, no reliability, validity, or standardization data are available for the Brigance. To assume that an unstandardized collection of criterion-referenced items gives a definitive picture of a child's future ability is highly questionable.

Furthermore, high-stakes testing carries consequences for the tester as well as the child. Lack of standardization data for the Brigance suggests that "any school system that formally and systematically uses the Brigance inventories without going through a local validation effort is placing itself at risk legally" (Robinson and Kovacevich 1984, p. 98).

Given this background, the use of the Brigance for instructional decision making is also questionable. Indeed, because most achievement and readiness tests assess a restricted range of instructional objectives, omit major adaptive and socio-emotional behavior, or do little more than confirm what the teacher already knew about the child, they are of limited relevance to teachers (Durkin 1987, Kelleghan et al. 1982, Salmon-Cox 1981).

What is missing in these cases is a match between the test and the teacher's curriculum. In low-stakes internal testing programs, tests are rarely perceived as particularly useful because of this lack of fit with the teacher's goals; and they usually do not have a major impact on instruction. Yet when the same tests are transformed into high-stakes tests, they can and do influence instruction, though clearly not for the right reasons.

The Georgia Experience. The final attribute of high-stakes testing to be discussed here is the subtle transfer of control over the curriculum to the test developer. Nowhere is this abrogation of instructional authority better exemplified than in the testing program implemented by the state of Georgia in 1988. In 1986 the state passed the Quality Basic Education (QBE) Act requiring all children seeking to enter 1st and 4th grades to pass an academic readiness test. (The kindergarten teacher also completes an independent assessment of the child's readiness.)

The test selected for 1st grade entry by the Georgia Department of Education is the California Achievement Test (CAT), level 10 (CTB/McGraw-Hill 1988). In the "Georgia Edition" of the CAT, however, only 64 of the 146 items (44 percent) are administered. The stated purpose of the Georgia CAT is to measure achievement in the basic skills and to provide specific information about students' instructional needs. The manual states that the CAT items "may be used to establish reference points for beginning instruction in kindergarten and to predict 1st grade reading achievement" (CTB/McGraw-Hill 1988, p. 1). However, no predictive validity data are given, nor is the reading test that was used as a criterion identified. Nevertheless, the Georgia CAT is used as a high-stakes test: it is designed to render decisions about student classification, retention, and promotion; it is intended to guide instructional decisions; and it is perceived as carrying out the state's mandate to establish quality education programs.

Unfortunately, the test and the testing program fall far short of achieving these goals. First, the test was modified without any empirical validation, although it is a psychometric axiom that subsets of items do not share the psychometric properties of the core test from which the items were drawn (APA/AERA/NCME 1985). Second, the Georgia CAT includes only the three subtests of visual recognition, sound recognition, and mathematical concepts. There is no assessment of the child's attention, motivation, expressive language, motor development, rate of learning, and so on. A third flaw of the test is the consequence of failing it, which is retention in grade; however, the evidence on kindergarten retention does not support it as a strategy for improving academic achievement (Plummer et al. 1986; Shepard and Smith 1987, 1988). More-

over, age/grade status is the single most sensitive indicator of high school dropout potential in urban school districts, and retained students are at least one year overage for grade.

A fourth problem concerns the test's cutoff for failure. Initial results indicate that 8 percent of the children who took the CAT in 1988 failed to score above the Georgia cutoff, that is, the 10th percentile. In some districts the failure ratio was as low as 1 percent, but in others it was as high as 26 percent (Cunningham 1988). No data are available concerning the racial, ethnic, geographic, and socioeconomic composition of this group of children. It is possible that poor and minority children are overrepresented among these "failures" and that the lack of cultural sensitivity of the test may have contributed to this problem. In any event, if as many as 8 percent of the children were unable to perform above the 10th percentile on this test, it is clear that the school districts had not previously identified those children who were at high risk for school failure. Use of a validated developmental screening test at the outset of kindergarten (see Meisels 1985) could have resulted in most of these children being identified before they experienced a year of kindergarten failure.

The action of the Georgia legislature in promulgating the QBE should serve as a warning to parents, professionals, and lawmakers throughout the nation. Administrators and teachers in local school districts have been told that their performance will be evaluated based on the gains made by their students on the CAT in succeeding years. Teachers have begun to alter their programs and their teaching styles so that children will have a better chance to do well on the test. Private firms have begun to offer preparatory classes to kindergartners ("CAT Academies"), and national companies are marketing kindergarten test-taking skills programs. The Georgia plan is the *reductio ad absurdum* of high-stakes testing, where an entire state (and so far the only state) has transferred control over its early education program to a single group-administered paper-and-pencil test.

There are presently no readiness or achievement tests sufficiently accurate to serve the high-stakes functions they are being asked to perform.

Does High-Stakes Testing Have a Place?

Before using a high-stakes test with young children, school districts should carefully consider: (1) the purpose for testing, (2) the selection of a valid test related to the reason for testing, and (3) the program options to which children who fail the test can be referred.

Purpose. An alternative to the present dilemma is to view readiness tests as a *first step* in the diagnostic-prescriptive process. From this perspective, the purpose of assessing a child's readiness is to understand better what the child does or does not know and can or cannot do, in order to design an instructional program that can increase and enhance the child's knowledge and skills. Using this low stakes model, one assesses readiness to enhance the benefits of instruction, not to prohibit a child from participating in a program. Readiness testing of this kind could result in major positive educational outcomes and could serve accountability needs with criterion-referenced documentation of student progress.

Tests. There are presently no readiness or achievement tests sufficiently accurate to serve the high-stakes functions they are being asked to perform. Rather, we are witnessing a major misuse of tests in the early childhood years. Consequently, a number of national organizations[2] have adopted the maxim that tests should be used only for their intended purposes to obtain the best and most appropriate services for the greatest number of children. Since readiness tests generally do not have predictive validity, and since they are often used to place children in extra-year programs that do not have demonstrable efficacy, their use for predicting future school functioning should be halted.

The critical issue for parents, professionals, and policymakers is to understand the importance of selecting the right test for the right child at the right time. Reliable and valid developmental screening tests, when administered by trained testers to individual children before they enter school, can identify children who are at high risk for school failure (see Meisels 1988, 1989). Children so identified would then move on to a more comprehensive diagnostic process to determine conclusively the nature of their problems and, subsequently, to obtain appropriate interventions. In addition, well-developed criterion-referenced tests, when used on multiple occasions, can be used to assess child progress. In other words, the problem is not tests *per se*, but the appropriate and inappropriate uses of tests in specific situations by specific individuals.

Programs. Children who are placed in readiness or transition classes are assumed to be the beneficiaries of more appropriate programming than would be possible in regular classrooms. However, the research on grade retention, transition classes, and even the relative achievements of younger versus older children in the same classrooms does not support this assumption (Gredler 1978; Plummer et al. 1986; Shepard and Smith 1986, 1988; Smith and Shepard 1987). Rather, extra-year programs often translate into retention programs that provide little more than an additional year of schooling, with possible negative consequences. When children are retained, it is essential that they experience a challenging program with the potential for helping them rejoin their age-mates within a year, and in a context of high parental, teacher, and administrator expectations. Currently, few transition programs meet these criteria.

Alternative Strategies

Schools have been sharply criticized for the practice of "social promotion," that is, keeping children with their age groups regardless of academic performance or achievement. The shift to high-stakes testing is a reaction to the perceived absence of standards implicit in promoting children this way. But, as we have seen, high-stakes testing is a highly flawed policy. In its place we need a reformulation of early childhood curriculum and classroom organization.

We must consider a number of alternatives, beginning with limiting class size, making more classroom aides available, and providing tutors whenever they are needed, to give students ample adult attention. We can also plan to offer more small-group instructional programs and more cross-age tutoring and instructional grouping in early childhood. Further, we need to increase the variety and quality of materials and resources used in early childhood programs. And, if we adopt a unified developmental elementary curriculum, we can broaden our approaches to teaching academic skills and feel comfortable supporting teachers' innovations. Last but not least, we must continue to promote effective home-school communication during the first four years of school.

These alternative strategies are consistent with a model of *low-stakes* testing whereby the information teachers obtain can benefit their students. This information is enhanced by the use of valid developmental screening tests to identify those children at highest risk for school failure and child-focused readiness tests to help teachers diagnose educational needs. Also needed are criterion-referenced assessments to document children's progress. Educators can use such information internally to fashion responsive educational programs and externally to document children's progress for purposes of planning and accountability.

These innovations are labor-intensive and expensive. But what is taking place in the name of readiness testing and retention is already extremely costly in terms of both economics and deferred human potential. Indeed, one study of 161 extra-year early childhood programs in a single state reported an annual cost of more than $3.4 million (Michigan Department of Education 1984). Nationally, the cost of implementing extra-year programs is probably in the hundreds of millions of dollars. It is time to transfer some of these funds to more worthwhile ends. The highest-stakes test of all is our ability to help children realize their full potential.□

1. See Bear and Modlin 1987; Bradley 1985; Kaufman 1985; May and Campbell 1981; May and Welch 1984a, 1984b; Meisels 1987; Naglieri 1985; Shepard and Smith 1985, 1986, 1987; Smith and Shepard 1987.

2. National Association for the Education of Young Children 1988, National Association of Early Childhood Specialists in State Departments of Education 1987, and National Association of State Boards of Education 1988.

References

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Tests*. Washington, D.C.: American Psychological Association.

Angoff, W.H., and S.B. Anderson. (1967). "The Standardization of Educational Psychological Tests." In *Educational and Psychological Measurement*, edited by D.A. Payne and R.F. McMorris. Waltham, Mass.: Blaisdell.

Barnes, K.E. (1982). *Preschool Screening: The Measurement and Prediction of Children At-Risk*. Springfield, Ill.: Charles C Thomas.

Bear, G.G., and P.D. Modlin. (1987). "Gesell's Developmental Testing: What Purpose Does it Serve?" *Psychology in the Schools* 24: 40-44.

Boehm, A.E. (1985). "Review of Brigance K and 1 Screen." In *The Ninth Mental Measurements Yearbook*, vol. 1, edited by J. Mitchell, Jr. Lincoln, Nebr.: University of Nebraska Press.

Bradley, R.H. (1985). "Review of Gesell School Readiness Tests." In *The Ninth Mental Measurements Yearbook*, vol. 1, edited by J. Mitchell, Jr. Lincoln, Nebr.: The University of Nebraska Press.

Brigance, A.H. (1982). *Brigance K and 1 Screen for Kindergarten and First Grade*. North Billerica, Mass.: Curriculum Associates, Inc.

Bruner, J.S. (1966). *Towards a Theory of Instruction*. Cambridge, Mass.: Harvard University Press.

Bryk, A.S., S.J. Meisels, and M.T. Markowitz. (1979). "Assessing the Effectiveness of Open Classrooms on Children with Special Needs." In *Special Education and Development: Perspectives on Young Children with Special Needs*, edited by S. J. Meisels. Baltimore: University Park Press.

Carini, P.F. (1975). *Observation and Description: An Alternative Methodology for the Investigation of Human Phenomena*. Grand Forks: University of North Dakota.

Coons, P. (November 29, 1987). "Kindergarten: Who Is Ready?" *The Boston Sunday Globe*, pp. B-77, B-79.

CTB/McGraw-Hill (1988). *California Achievement Test, Grade K*. Georgia Edition. Monterey, Calif.: CTB/McGraw-Hill.

Cunningham, A.E. (1988). *Eeny, Meeny, Miny, Moe: Testing Policy and Practice in Early Childhood*. Berkeley, Calif.: National Commission on Testing and Public Policy.

Darling-Hammond, L., and A.E. Wise. (1985). "Beyond Standardization: State Standards and School Improvement." *The Elementary School Journal* 85: 315-336.

Durkin, D. (1987). "Testing in the Kindergarten." *The Reading Teacher* 40: 766-770.

Fiske, E.B. (April 10, 1988). "America's Test Mania." *The New York Times Spring Education Supplement*, pp. 16-20.

Fiske, J. (May 8, 1988). "Kindergarten: The Rules Have Changed." *The Press-Enterprise* (Riverside, Calif.), pp. B-1, B-3.

Gredler, G.R. (1978). "A Look at Some Important Factors in Assessing Readiness for School." *Journal of Learning Disabilities* 11: 284-290.

Haines, J., L.B. Ames, and C. Gillespie. (1980). *The Gesell Preschool Test Manual*. Lumberville, Pa.: Modern Learning Press.

Hechinger, F.M. (September 14, 1988). "Repeating Kindergarten: Does It Hurt More Than It Helps?" *The New York Times*, p. 24.

Hein, G.E. (1979). "Evaluation in Open Education: Emergence of a Qualitative Methodology." In *Special Education and Development: Perspectives on Young Children with Special Needs*, edited by S.J. Meisels. Baltimore: University Park Press.

Kaufman, N.L. (1985). "Review of Gesell

Preschool Test." In *The Ninth Mental Measurements Yearbook*, vol. 1., edited by J. Mitchell, Jr. Lincoln: The University of Nebraska Press.

Kelleghan, T., G.F. Madaus, and P.W. Airasian. (1982). *The Effects of Standardized Testing*. Boston: Kluwer-Nijhoff Publishing.

Madaus, G.F. (1988). "The Influence of Testing on the Curriculum." In *Critical Issues in Curriculum: 87th Yearbook of the National Society for the Study of Education*, edited by L.N. Tanner. Chicago: University of Chicago Press.

May, C.R., and R.M. Campbell. (1981). "Readiness for Learning: Assumptions and Realities." *Theory Into Practice* 20: 130-134.

May, D.C., and E.L. Welch. (1984a). "The Effects of Developmental Placement and Early Retention on Children's Later Scores on Standardized Tests." *Psychology in the Schools* 21: 381-385.

May, D.C., and E.L. Welch. (1984b). "Developmental Placement: Does It Prevent Future Learning Problems?" *Journal of Learning Disabilities* 17: 338-341.

Meisels, S.J. (1985). *Developmental Screening in Early Childhood: A Guide*. Rev. ed. Washington, D.C.: National Association for the Education of Young Children.

Meisels, S.J. (November 1986). "Testing Four- and Five-Year-Olds." *Educational Leadership* 44: 90-92.

Meisels, S.J. (1987). "Uses and Abuses of Developmental Screening and School Readiness Testing." *Young Children* 42: 4-6, 68-73.

Meisels, S.J. (1988). "Developmental Screening in Early Childhood: The Interaction of Research and Social Policy." In *Annual Review of Public Health*, vol. 9, edited by L. Breslow, J.E. Fielding, and L.B. Lave. Palo Alto, Calif.: Annual Reviews.

Meisels, S.J. (1989). "Can Developmental Screening Tests Identify Children Who Are Developmentally at Risk?" *Pediatrics* 83.

Meisels, S.J., G. Harbin, K. Modigliani, and K. Olson. (1988). "Formulating Optimal State Early Childhood Intervention Policies." *Exceptional Children* 55: 159-165.

Michigan Department of Education. (1984). *Superintendent's Study Group on Early Childhood Education*. Lansing: Michigan Department of Education.

Naglieri, J.A. (1985). "Review of Gesell Preschool Tests." In *The Ninth Mental Measurement Yearbook*, vol. 1, edited by J. Mitchell, Jr. Lincoln: The University of Nebraska Press.

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.

National Association of Early Childhood Specialists in State Departments of Education. (1987). *Unacceptable Trends in Kindergarten Entry and Placement: A Position Statement*. Lincoln, Nebr.: NAECSSDE.

National Association of State Boards of Education. (1988). *Right From the Start: The Report of the NASBE Task Force on Early Childhood Education*. Alexandria, Va.: NASBE.

Plummer, D.L., M.H. Lineberger, and W.G. Graziano. (1986). "The Academic and Social Consequences of Grade Retention: A Convergent Analysis." In *Current Topics in Early Childhood Education*, vol. 6, edited by L.G. Katz. Norwood, N.J.: Ablex Publishing Co.

Riley, S., P. Carter, C. Cummings, J. Firestone, C. Flynn, S. Javid, and D. Ruiter. (September 1988). "Survey Results: Early Childhood Programming." Paper presented at State Kindergarten Conference, Flint, Mich.

Robinson, J.H., and D.A. Kovacevich. (1984). "The Brigance Inventories." In *Test Critiques*, vol. 1, edited by D.J. Keyser and R.C. Sweetland. Kansas City, Mo.: Test Corporation of America.

Salmon-Cox, L. (1981). "Teachers and Standardized Achievement Tests: What's Really Happening?" *Phi Delta Kappan* 62: 631-633.

Shepard, L.A., and M.L. Smith. (1985). *Boulder Valley Kindergarten Study: Retention Practices and Retention Effects*. Boulder, Colo.: Boulder Valley Public Schools.

Shepard, L.A., and M.L. Smith. (November 1986). "Synthesis of Research on School Readiness and Kindergarten Retention." *Educational Leadership* 44: 78-86.

Shepard, L.A., and M.L. Smith. (1987). "Effects of Kindergarten Retention at the End of First Grade." *Psychology in the Schools* 24: 346-357.

Shepard, L.A., and M.L. Smith. (1988). "Escalating Academic Demand in Kindergarten: Counterproductive Policies." *Elementary School Journal* 89: 135-145.

Smith, M.L., and L.A. Shepard. (1987). "What Doesn't Work: Explaining Policies of Retention in the Early Grades." *Phi Delta Kappan* 69: 129-134.

Wood, C., S. Powell, and R.C. Knight. (1984). "Predicting School Readiness: The Validity of Developmental Age." *Journal of Learning Disabilities* 17: 8-11.

Author's note: I am grateful to Nancy Klein, Nick Anastasiow, Steve Sternberg, and Judy Jefferson for extremely helpful comments on earlier drafts of this paper.

Samuel J. Meisels is Professor, School of Education, and Research Scientist, Center for Human Growth and Development, University of Michigan, 300 N. Ingalls, Ann Arbor, MI 48109.