DOCUMENT RESUME

ED 416 033 PS 026 296

AUTHOR Shepard, Lorrie, Ed.; Kagan, Sharon Lynn, Ed.; Wurtz, Emily, Ed.

TITLE Principles and Recommendations for Early Childhood Assessments.
INSTITUTION National Education Goals Panel, Washington, DC.
PUB DATE 1998-02-00
NOTE 45p.
AVAILABLE FROM National Education Goals Panel, 1255 22nd Street, N.W., Suite 502, Washington, DC 20037; phone: 202-724-0015; fax: 202-632-0957; www: http://www.negp.gov; e-mail: NEGP@goalline.org
PUB TYPE Guides - Non-Classroom (055)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Early Childhood Education; *Educational Objectives; Evaluation Criteria; Evaluation Methods; *Learning

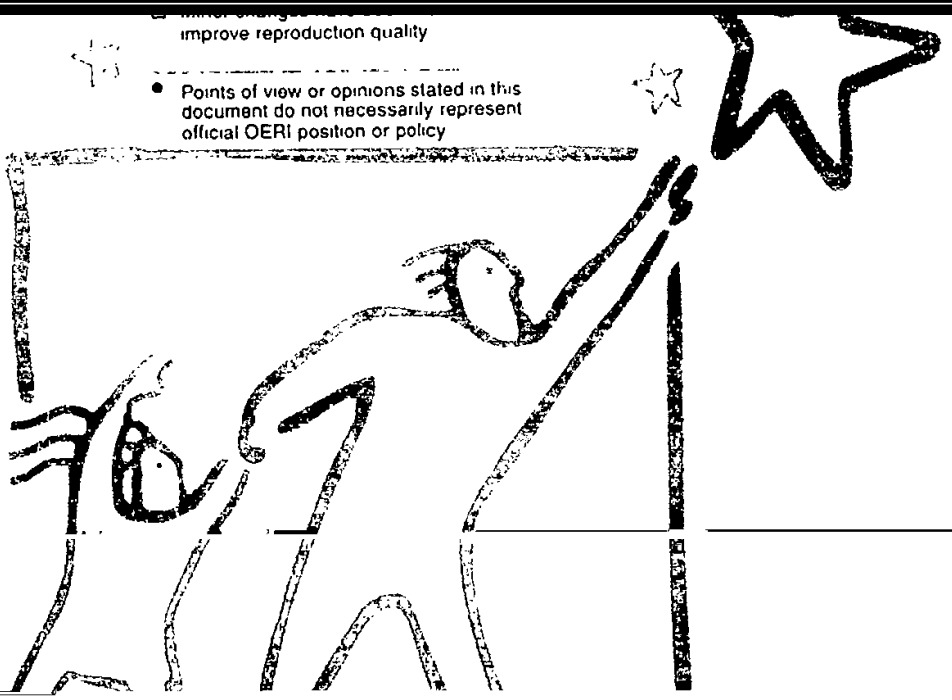
ABSTRACT

The first of the National Education Goals states that by the year 2000 all children in America will start school ready to learn. Pressed

demands for clear and useful information. Following a look at recent

purpose, the definition, audience, technical requirements and age continuum are listed. Recommendations for policymakers are also presented for each purpose, and a chart outlining appropriate uses and technical accuracy of assessments change across the early childhood age continuum is included. The

PRINCIPLES AND RECOMMENDATIONS FOR EARLY CHILDHOOD ASSESSMENTS



Submitted to

THE NATIONAL EDUCATION GOALS PANEL

by the Goal 1 Early Childhood Assessments Resource Group

Lorrie Shepard, Sharon Lynn Kagan, and Emily Wurtz, Editors

National Education Goals Panel

Governors

James B. Hunt, Jr., North Carolina (Chair, 1997–1998)
John Engler, Michigan
William Graves, Kansas
Paul E. Patton, Kentucky
Roy Romer, Colorado
Tommy G. Thompson, Wisconsin
Cecil Underwood, West Virginia
Christine Todd Whitman, New Jersey

Members of the Administration

Carol H. Rasco, Senior Advisor to the Secretary of Education
Richard W. Riley, Secretary of Education

Members of Congress

U.S. Senator Jeff Bingaman, New Mexico
U.S. Senator Jim Jeffords, Vermont
U.S. Representative William F. Goodling, Pennsylvania
U.S. Representative Dale E. Kildee, Michigan

State Legislators

Representative G. Spencer Coggs, Wisconsin
Representative Ronald Cowell, Pennsylvania
Representative Mary Lou Cowlishaw, Illinois
Representative Douglas R. Jones, Idaho

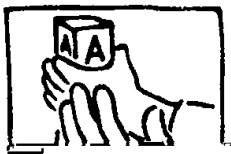
National Education Goals Panel Staff

Ken Nelson, Executive Director
Leslie A. Lawrence, Senior Education Associate
Cynthia M. Dixon, Program Assistant
John Masaitis, Executive Officer
Sherry Price, Secretary

Goal 1 Early Childhood Assessments Resource Group

Edward Chittenden, Educational Testing Service
Harriet Egertson, Nebraska State Department of Education
Eugene García, University of California, Berkeley
M. Elizabeth Graue, University of Wisconsin
Kenji Hakuta, Stanford University
Carollee Howes, University of California, Los Angeles
Annemarie Palincsar, University of Michigan
Tej Pandey, California State Department of Education
Catherine Snow, Harvard University
Maurice Sykes, District of Columbia Public Schools
Valora Washington, The Kellogg Foundation
Nicholas Zill, Westat, Inc.

PRINCIPLES AND RECOMMENDATIONS FOR EARLY CHILDHOOD ASSESSMENTS



Goal 1: Ready to Learn

preschool programs that help prepare children for school.

Every parent in the United States will be a child's first teacher and devote time

care needed to arrive at school with healthy minds and bodies, and to maintain the mental alertness necessary to be prepared to learn, and the number of low-birthweight babies will be significantly reduced through enhanced prenatal health systems.

FORM LEADERSHIP, STANDARDS, AND ASSESSMENTS

PART A—NATIONAL EDUCATION GOALS PANEL

SEC. 201. PURPOSE.

SEC. 207. EARLY CHILDHOOD ASSESSMENT.

(c) ADVICE.—The Groups shall advise and assist the Congress, the Secretary, the Goals Panel, and others regarding how to improve the assessment of young children and how such assessments can improve services to children.

(d) REPORT.—The Goals Panel shall provide reports on the work of the Groups to the appropriate committees of the Congress, the Secretary, and the public.

Introduction

Americans want and need good information on the well-being of young children. Parents want to know if their children will be ready for school. Teachers and school administrators want to know if their programs are effective and if they are providing children the right programs and services. Policymakers want to know which program policies and expenditures will help children and their families, and whether they are effective over time. Yet young children are notoriously difficult to assess accurately, and well-intended testing efforts in the past have done unintended harm. The principles and recommendations in this report were developed by advisors to the National Education Goals Panel to help [illegible] assessing young children appropriately and effectively.

The first National Education Goal set by President Bush and the nation's [illegible]

Childhood Assessments Resource Group to address this charge.

Assessment and the Unique Development of Young Children [illegible]

and the learning setting.

Because young children learn in ways and at rates different from older children and adults, we must tailor our assessments accordingly. Because young children come to know things through doing as well as through listening, and because they often represent their knowledge better by showing than by talking or writing, paper-and-pencil tests are not adequate. Because young children do not have the

opportunities, it is a mistake to interpret measures of past learning as evidence of what could be learned.

Recent Assessment Issues

These observations have proven effective for purposes of chronicling children's development, cataloging their accomplishments, and tailoring programs and activities within the classroom to meet young children's rapidly changing needs.

Recently, however, there has been an increase in formal assessments and testing,

using test results to sort children into or out of kindergarten and preschools. In

The Current Climate

Despite these difficulties, demands for assessments of student learning are increasing. Pressed by demands for greater accountability and enhanced educational performance, states are developing standards for school-aged children and are creating new criteria and approaches for assessing the achievement of challenging academic goals. In this context, calls to assess young children—from birth through the earliest grades in school—are also increasing. This document attempts to indicate how best to craft such assessments in light of young children's unique development, recent abuses of testing, and the legitimate demands from parents and the public for clear and useful information.

recommendations include warnings to protect against potential misuse. To explain the basis of these recommendations, there is a definition of each of four categories

General Principles

The following general principles should guide both policies and practices for the assessment of young children.

- **Assessment should bring about benefits for children.**
  Gathering accurate information from young children is difficult and potentially

other purposes. In the past, many of the abuses of testing with young children

assessment data. It is particularly difficult to assess children's cognitive or affective accurately before age 6. Because of problems with reliability and validity, some types of assessment should be postponed until children are older, while other types of assessment can be pursued, but only with necessary safeguards.

- **Assessments should be age-appropriate in both content and the method of** and development, including physical well-being and motor development;

- **Assessments should be linguistically appropriate, recognizing that to some extent all assessments are measures of language.**

Each child's first- and second-language development should be taken into account when determining appropriate assessment methods and in interpreting the meaning of assessment results.

- **Parents should be a valued source of assessment information, as well as an audience for assessment results.**
Because of the fallibility of direct measures of young children, assessments

For example, if data from a statewide assessment are going to be used for school accountability, then it is important that data be collected in a standardized way to ensure comparability of school results. If children in some schools are given practice ahead of time so that they will be familiar with the task formats, then children in all schools should be provided with the same practice; teachers should not give help during the assessment or restate the questions unless it is part of the standard administration to do so; and all of the assessments should be administered

assessments almost always occur in the context of activities and tasks that are already familiar, so practice or task familiarity is not at issue. In the classroom

learning opportunity and to figure out exactly how a child is thinking by seeing

teachers as part of the learning process lack the uniformity and the standardization that is necessary to ensure comparability, essential for accountability purposes.

Similarly, the technical standards for reliability and validity are much more stringent for high-stakes accountability assessment than for informal assessments

sufficiently accurate to ensure that important decisions about a child are not made as the result of measurement error. In addition, accountability assessments are usually "one-shot," stand-alone events. In contrast, caregivers and teachers are constantly collecting information over long periods of time and do not make high-stakes decisions. If they are wrong one day about what a child knows or is

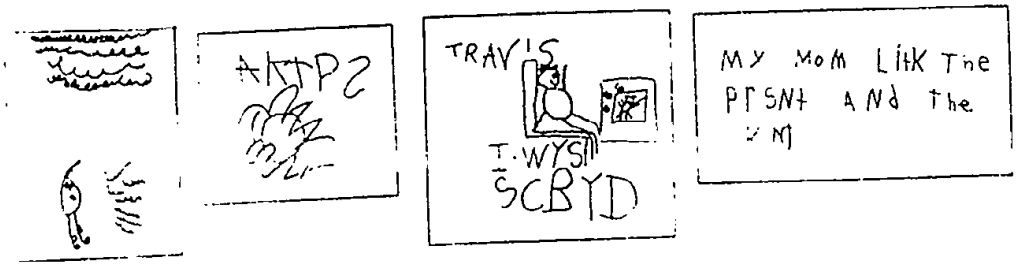
Serious misuses of testing with young children occur when assessments intended for one purpose are used inappropriately for other purposes. For example, the content of IQ measures intended to identify children for special education is not appropriate content to use in planning instruction. At the same time, assessments designed for instructional planning may not have sufficient validity and technical accuracy to support high-stakes decisions such as placing children in a special kindergarten designated for at-risk children.

An appropriate assessment system may include different assessments for different categories of purpose, such as:

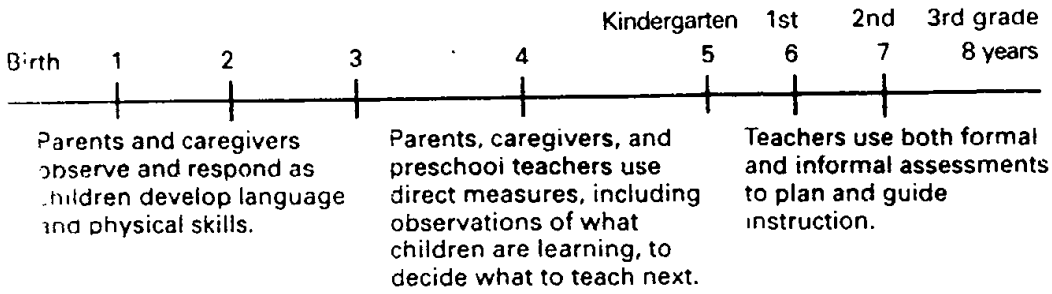
- assessments to support learning,
- assessments for identification of special needs,
- assessments for program evaluation and monitoring trends, and
- assessments for high-stakes accountability.

In the sections that follow, the requirements for each of these assessment

Photo: Martin Deutsch



(from the North Carolina Grades 1 and 2 Assessment)



Definition of purpose. Assessing and teaching are inseparable processes. When children are assessed as part of the teaching-learning process, then assessment

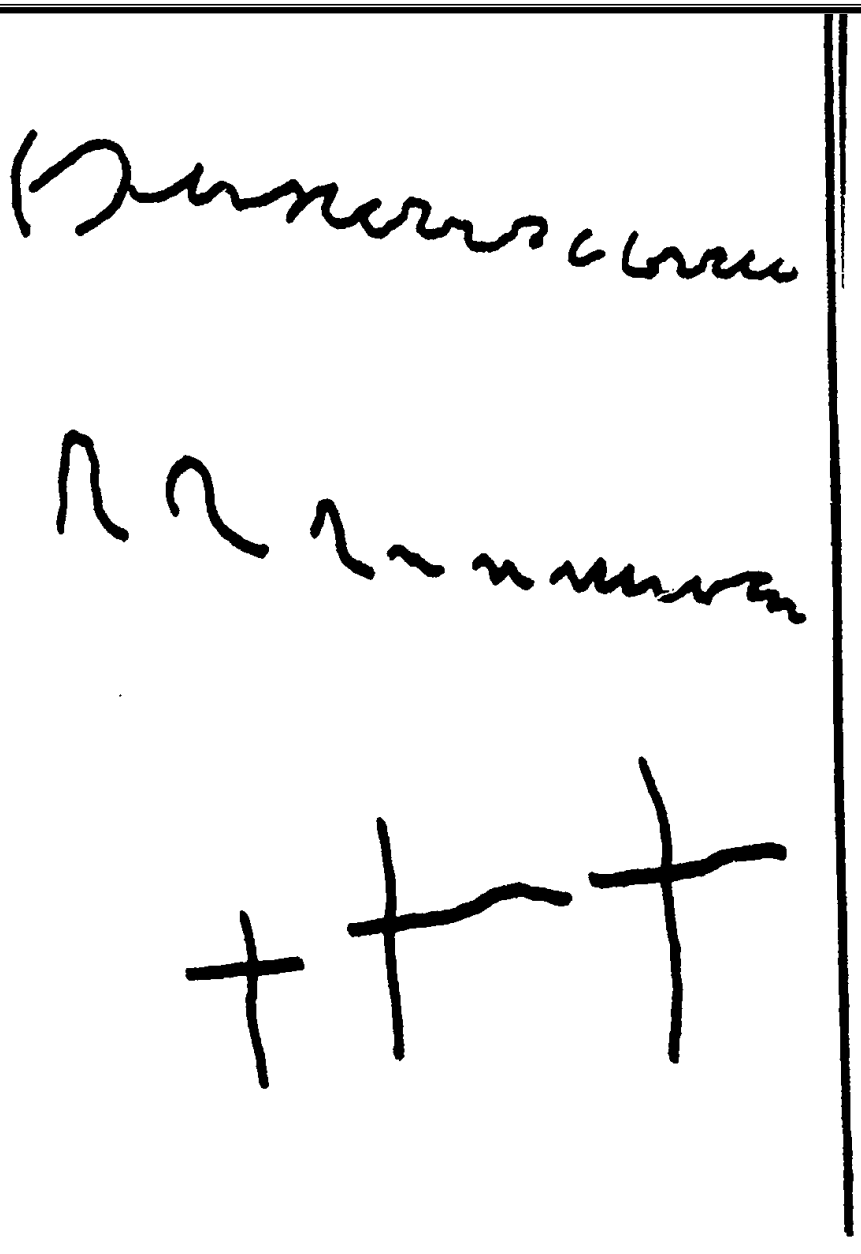
gauge what things children already know and understand, what things could be understood with more practice and experience, and what things are too difficult without further groundwork. This may include appropriate use of early learning readiness measures to be used in planning next steps in instruction. Teachers also use their assessments of children's learning to reflect on their own teaching practices, so that they can adjust and modify curricula, instructional activities, and classroom routines that are ineffective.

Audience. The primary audience for assessments used to support learning is *the teacher*, recognizing, of course, that parents are each child's first teachers. The primary caregiver is asking himself questions about what the child understands, what she does not understand, what she should be learning, and what is too soon

opportunities to learn that are closely congruent with

continuum. In more structured settings, classroom assessments are used by teachers

as part of instruction. Children benefit from seeing samples of their own work



teachers are now actively involving children in sharing their accomplishments with parents during conferences. Parents also want and need good information about how their child is doing. Although teachers collect much more information

Principals and primary-grade teachers may also work together to review instructional assessments to make sure that the school's programs are succeeding in helping young children meet developmental and academic expectations. Although external accountability testing should be postponed until third grade because of the difficulties in testing young children, grade-level teams of teachers and school administrators can use instructional assessments for purposes of internal, professional accountability to make sure that children who are struggling receive special help, to identify needs for further professional training, and to

knowing that such assessments are being used at the school level to monitor

Technical requirements. In order for assessments to support learning and

teachers must have enough knowledge about child development and cultural

assessments, conducted to improve learning, must also be tied to the preschool or primary curriculum and should have clear implications for what to do next.

The reliability and validity requirements for assessments used to support learning

for example in judging a child's reading level to help select a book from the library (this book is too easy), that decision is easily changed the next day when new assessment data are available. Because assessments used as part of learning do not have to meet strict standards for technical accuracy, they cannot be used for

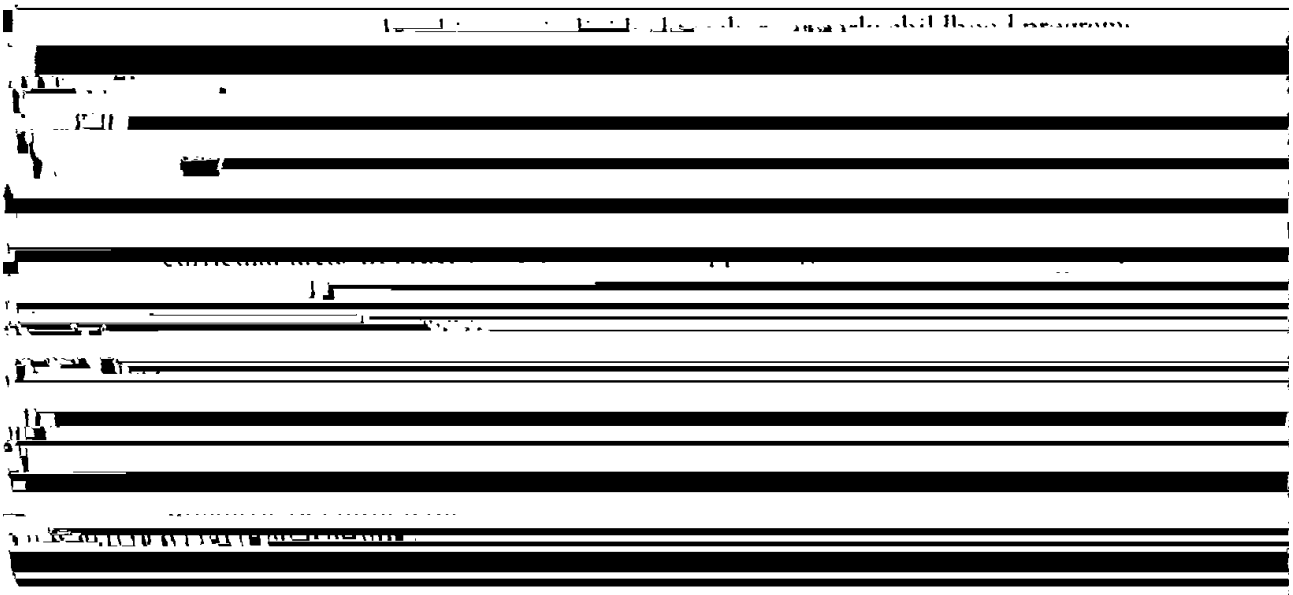
Age continuum. How old a child is within the early childhood age span of birth to 8 years old affects both the what and how of assessment. At all ages, attention should be paid to all five of the dimensions of early learning and development identified by the Goals Panel's Goal 1 Technical Planning Group: physical well-being and motor development; social and emotional development; approaches toward learning; language development; and cognition and general knowledge. Parents of toddlers and early caregivers address all five areas. Beginning in first

goals should continue to be part of classroom teaching and observation.

Methods of collecting assessment data include direct observation of children during natural activities; looking at drawings and samples of work; asking questions either orally or in writing; or asking informed adults about the child. The younger the child, the more appropriate it is to use observation. As age increases, especially

Recommendations for what policymakers can do

1. Policymakers should develop or identify assessment materials, to be used instructionally, that exemplify important and age-appropriate learning goals. At the earliest ages, caregivers need tools to assist in observing children. Lacking such assessment materials, preschool programs may misuse screening measures



thinking, to understand and analyze errors in thinking, and to build on each child's strengths.

Photo: Marietta Lynch

Sample of student work: the North Carolina Grades 1 and 2 Assessment

All children should be screened regularly for health needs, including as part of routine health care services.

Children entering Head Start and other preschool programs should be [illegible] including hearing and vision checks.

All children should be screened at school entry for vision and hearing immunizations.

Some mild disabilities [illegible]

suspected, parents and physicians should seek in-depth assessments.

children with potential disabilities are referred for in-depth assessment.

Definition of purpose. Assessments described in Purpose 1 are used by caregivers and teachers as part of supporting normal learning and development. Assessments used for Purpose 2 help to identify special problems and to determine the need for additional services beyond what regular caregivers can provide. The purpose of identification is to secure special services. Purpose 2 refers to identification of disabilities such as blindness, deafness, physical disabilities, speech and language impairment, serious emotional disturbance, mental retardation, and specific [illegible] immunization to ensure that appropriate health services are provided.

Because of the potential inaccuracy of nearly all sensory and cognitive measures and the cost of in-depth assessments, identification of special needs usually occurs in two stages. *Screening* is the first step in the identification process. It involves a [illegible] referred to a physician or child-study team for a more complete evaluation. For mental retardation and other cognitive disabilities, the second-stage in-depth assessment is referred to as a *developmental assessment*.

who must be involved in understanding and meeting their child's needs; and the preschool or primary-grade teacher who works with the child daily and who, most likely, made the referral seeking extra help.

Technical requirements. Except for extreme disabilities, accurate assessment of possible sensory or cognitive problems in young children is very difficult. The instruments used are fallible, and children themselves vary tremendously in their responses from one day to the next or in different contexts. In the field of special education, there is a constant tension between the need to identify children with disabilities to ensure early intervention and help, versus the possible harm of labeling children and possibly assigning them to ineffective treatments.

At step one in the identification process, the screening step, there are two [redacted] shortened versions of more in-depth assessments, and are therefore less reliable.

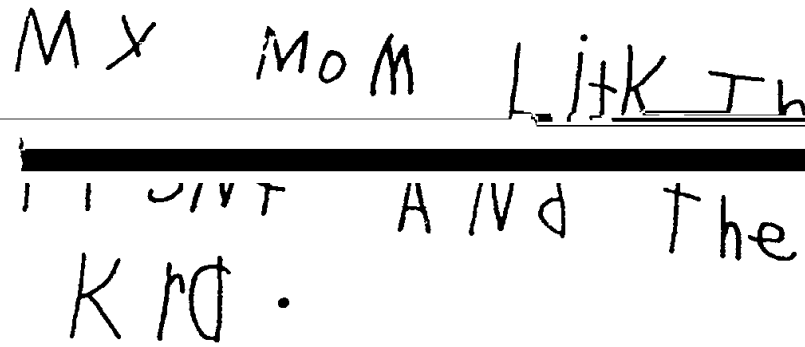
- Screening measures are only intended for the referral stage of identification. They are limited assessments, and typically are administered by school personnel who are not trained to make interpretations about disabilities.

- Screening measures should never be the sole measure used to identify children for special education. Because screening instruments have content like IQ tests, they should also not be used for [redacted]

For physical disabilities such as vision or hearing impairment, the second-stage in-depth assessment involves more sophisticated diagnostic equipment and the clinical skills of trained specialists. For potential cognitive and language disabilities,

[redacted]

competence in both home and school settings; and (4) for children with more



Sample of student work: the North Carolina Grades 1 and 2 Assessment

Age continuum. Special needs identification starts with the most severe—and

assessments are inaccurate at too early an age, but every child should have access to a regular health care provider, and children should be promptly referred if parents and physicians see that they are not reaching normal developmental

individual children with possible developmental delays should be referred for

undetected. For example, if a child has not received regular health checkups, a routine kindergarten screening may uncover a need for glasses.

Sample of student work: the North Carolina Grades 1 and 2 Assessment

Recommendations for what policymakers can do

1. States should ensure that all children have access to a regular health care provider to check for developmental milestones and to ensure that children

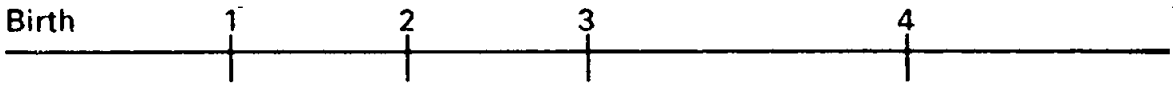
for in-depth assessments. Child Find is typically an organized effort by public health, social welfare, and educational agencies to identify all disabled children in need of services.

3. Mild forms of cognitive and language disabilities are particularly hard to identify. We know, however, that effective treatments for children with mild cognitive and language disabilities and most children at-risk for significant reading difficulty all involve the same kinds of high quality, intensive language and literacy interventions. Therefore, policymakers should consider increasing the availability and intensity of such services for broader populations of students who are educationally at-risk, including children in poverty and children thought to have special learning needs.

4. Given the potential for misuse of screening measures, states and districts that mandate screening tests should consider how they are being used and should evaluate whether identifications in their jurisdiction are more accurate with the use of formal tests than in states or districts where only parent and teacher referrals are used.

5. States that mandate administration of cognitive screening measures should expressly forbid the use of screening tests for other than referral purposes. Specifically, screening tests should not be used as readiness tests to exclude children from school; they should not be used to track children by ability in kindergarten and first grade; and they should not be used to plan instruction unless a valid relationship with local curricula has been established.

Appropriate Uses and Technical Accuracy of Assessments Change Across



Purpose 1: Assessing to promote children's learning and development

Purpose 3: Monitoring trends and evaluating programs and services

Because direct measures of

living and social conditions that affect learning and the adequacy of services.

Assessments, including direct and

programs, but such measures would not be accurate enough to make

Purpose 4: Assessing academic achievement to hold individual student

...Early Childhood Age Continuum (Birth to Age 8)

Kindergarten ... 8 years Beyond age 8

assessments to plan and guide instruction.

school entry for vision and hearing needs and checked for immunizations.

Some mild disabilities may only become apparent in the school context. Districts and states must by law have sound teacher and parent referral policies, so that children with potential disabilities are referred for in-depth assessment.

Beginning at age 5, it is possible to use direct measures, including measures of children's early learning, as part of a comprehensive early childhood assessment for monitoring trends. Matrix sampling should be used to ensure technical accuracy and to provide safeguards for individual children. Because of the cost of such an assessment, states or the nation should pick one ... early childhood, most likely kindergarten or first grade.

...e ...eachers, and schools accountable

about individual children and schools.

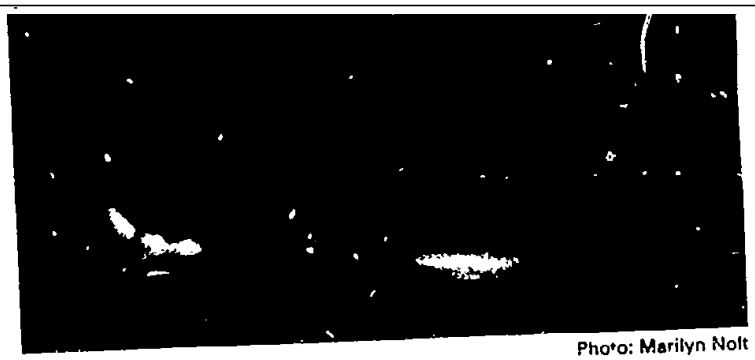
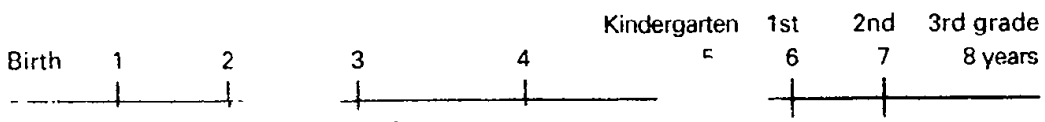


Photo: Marilyn Nolt

Mi mamal
and mi StrJohn

Purpose 3. Monitoring trends and evaluating programs and services



Because direct measures of children's language and cognitive functioning are difficult to aggregate accurately for ages from birth to 2, state reporting systems should focus on living and social conditions that affect learning and the adequacy of services.

Assessments, including direct and indirect measures of children's physical, social, emotional and cognitive development could be constructed and used to evaluate pre-kindergarten programs, but such measures would not be accurate enough to make high-stakes decisions about individual children.

Beginning at age 5, it is possible to use direct measures, including measures of children's early learning, as part of a comprehensive early childhood assessment for monitoring trends. Matrix sampling should be used to ensure technical accuracy and to provide safeguards for individual children. Because of the

Definition of purpose. For assessment Purpose 1 and Purpose 2, assessment data

degree of language development or familiarity with concepts of print. (For example, does the child come to school knowing how to hold a book and knowing

We have combined within Purpose 3 two closely related uses of aggregate data, *monitoring trends* and *program evaluation*. Large-scale assessment programs such as

document any changes in levels of student performance. Assessments designed to monitor trends could be used to monitor progress toward Goal 1 or to answer the question, "How is my state doing compared to the United States, another state, or

Purpose 3 hold programs "accountable" and hold states "accountable" for the adequacy of social conditions and services to young children. However, because the use of data to judge national or state programs entails consequences for the programs rather than for individuals, it is still relatively low-stakes for the

monitoring and program evaluation uses of data and the high-stakes accountability uses of assessments described in Purpose 4, which entail consequences for individuals.

improving programs. For example, national evaluations of Head Start provide evidence to Congress of the benefits of early educational interventions, which ensures continued funding as well as the establishment of related programs, such as Early Head Start and Even Start. In addition, more detailed evidence gathered as part of Head Start demonstrations and evaluations gives feedback to the system, and can be used for subsequent improvement of the overall Head Start program. For example, early evaluations documented and reinforced the importance of parent involvement in accomplishing and sustaining program goals. Similarly, the data from Goal 1 activities can be used to inform the public regarding the overall status of America's young children, as well as identifying where services are needed

large-scale assessment data must meet high standards of technical accuracy. For example, if policy changes are going to be made because reading scores have gone

Because of their visibility, state and national assessments also serve important symbolic functions. For example, when the NAEP results are reported, they are often accompanied by sample problems illustrating what students at each age

should know and be able to do. Because teachers and school administrators often make changes in curriculum and instructional strategies in an effort to improve performance on such external assessments, it is important that the NAEP for fourth and eighth graders include challenging open-ended problems, and not just

conditions of learning—the adequacy of [illegible] and the [illegible] support

Purpose 2, gathering data from sufficient numbers of children can ensure accuracy for purposes of evaluating programs. *Matrix sampling* is a statistical technique

of the total assessment, it is impossible to use the results to make decisions about individual children. This safeguard feature is especially important as a safeguard

range from birth to age 8. For children from birth to 2, the only direct measures that are sufficiently accurate to be feasible in a large-scale, every-child data collection effort are measures of physical characteristics such as birthweight. For children in this youngest age range, monitoring systems should focus on the conditions of learning by creating social indicators that track characteristics of families and the adequacy of [illegible] Important indicators in this earliest age range include

percentage of low-birthweight babies or the percentage of 2-year-olds getting

or preschool. It is also possible to assess learning of 3- and 4-year-olds directly.

motor learning for 3- and 4-year-olds. To avoid overtesting and protect against misuse, these assessments should use matrix sampling procedures. To ensure appropriate and accurate procedures, assessments should be administered to children individually by trained examiners under controlled conditions. Direct measures of learning would be costly to develop and administer, but the

Although direct measures of learning are possible in the context of

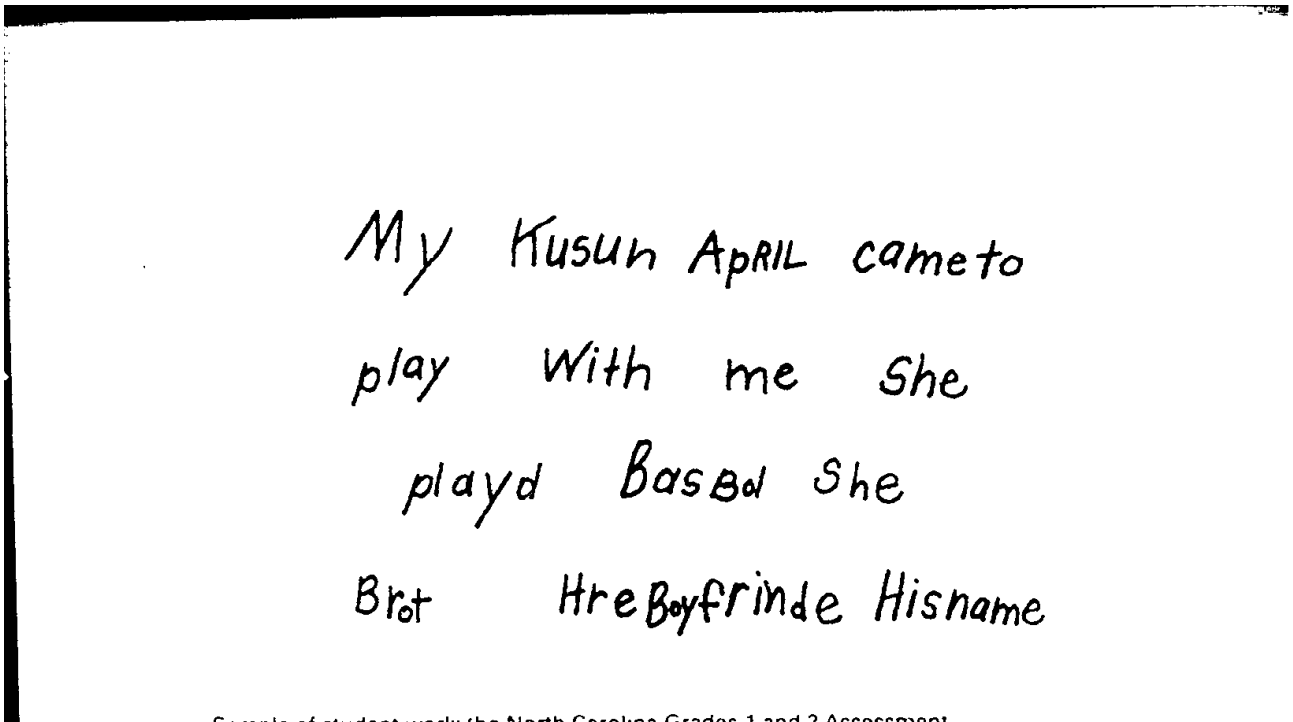
the program, a state monitoring system would require a household survey and individual assessments for a sample of children in their homes, at a cost that would outweigh potential benefits.

Beginning at age 5, however, it would be possible to administer direct measures of learning outcomes to children in school as part of a monitoring system. For

a national Early Childhood Assessment to provide comprehensive information about the status of the nation's children during their kindergarten years. The envisioned assessment would not only address the multiple dimensions of early learning and development, but would also counteract the fallible nature of each data source by collecting information from parents, teachers, and children themselves, through both direct measures and portfolios of classroom work. The five dimensions of early learning suggested by the Resource Group are being used by the National Center for Education Statistics as the framework for developing measures for the National Early Childhood Longitudinal Survey. Although these measures would not be available for widespread use, the insights gained from their development and field testing should be helpful to states trying to develop their own assessments.

Individual states could consider developing an early childhood assessment

effects of learning opportunities and services available in the years before school



Sample of student work: the North Carolina Grades 1 and 2 Assessment

blurring of preschool and school effects. However, a kindergarten-year assessment would have special sampling problems, because participation in kindergarten is voluntary in many states. At a minimum, accurate interpretation of trend data [illegible] of children in private kindergartens as well as in public

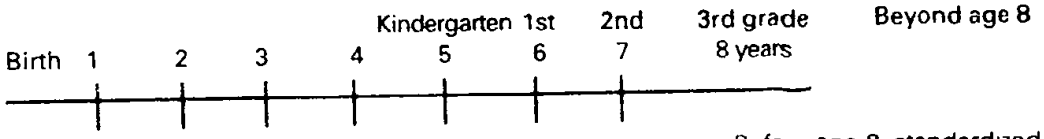
1. Before age 5, large-scale assessment systems designed to inform educational and social policy decisions about young children should focus on social indicators [illegible] the conditions of learning. Direct measures of learning outcomes

2. Beginning at age 5, it is possible to use direct measures, including measures of children's learning, as part of a comprehensive early childhood system to monitor trends. Matrix sampling procedures should be used to ensure technical accuracy and at the same time protect against the misuse of data to make decisions about individual children. Because such systems are costly to implement, states or the nation should pick one grade level for purposes of [illegible]

Photo: Michael Tony

We put books on the table and
nade a Maze for the guinea
Pigs. We put a guinea pig
way in the back of the house.
They went to find the grass
at the Other end of the
house. We were trying to find out
how many Seconds it would take
them to find the grass. It took
Rodney 2 minutes and 20 seconds

Purpose 4. Assessing academic achievement to hold individual students, teachers, and schools accountable



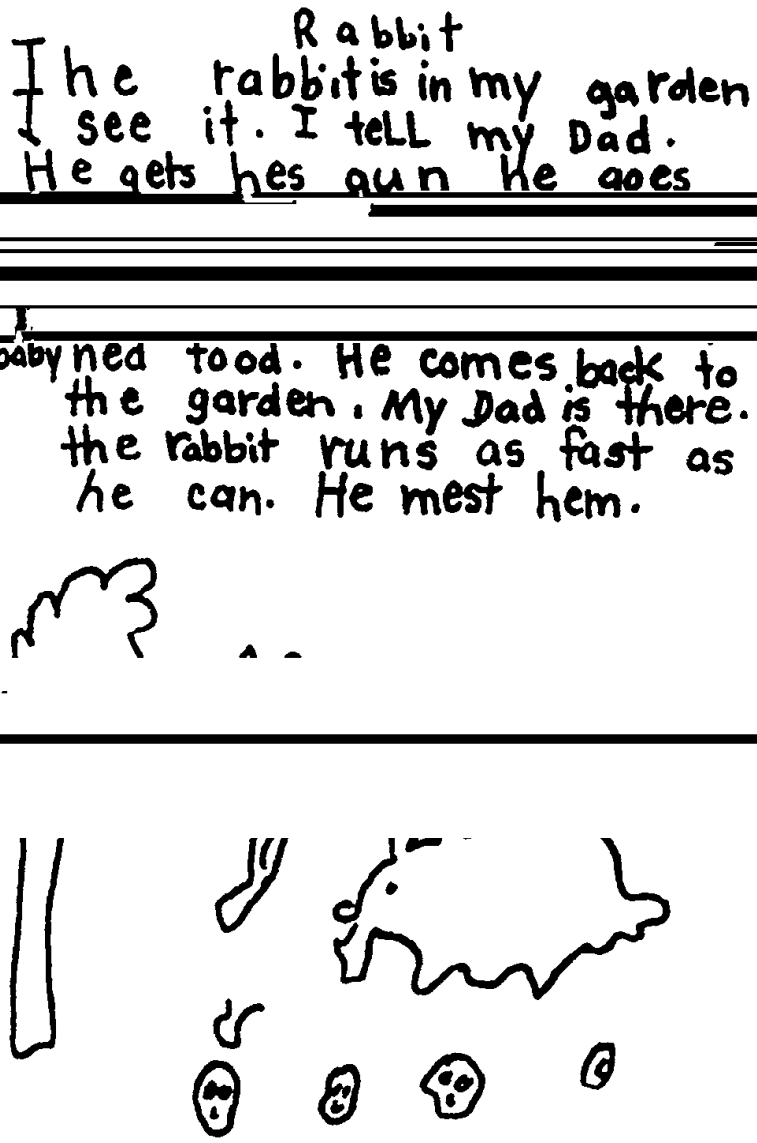
Before age 8, standardized achievement measures are not sufficiently accurate to be used for high-stakes decisions about individual children and schools. Therefore, high-stakes assessments intended for accountability purposes should be delayed until the end of third grade (or preferably fourth grade).

Definition of purpose. Purpose 4 refers to external examinations, mandated by an [illegible] usually the state or school district, and administered [illegible] accountable for desired learning outcomes. For policymakers, there is a close similarity between the use of assessment data for Purpose 3 and Purpose 4. Both [illegible]

results. Included in this category are external assessments administered nationally or by states and school districts. If results are reported for individual students, classrooms, or schools, then the assessment has much higher stakes than either [illegible]

however, that merely reporting [illegible] high stakes to assessment results with accompanying changes produced in instructional practices. Therefore, *the decision to report scores for individual students and schools* places assessments in this "accountability" category, whether or not the assessment is explicitly labeled as an accountability system.

Audience. Policymakers and the general public are, again, the primary audience for accountability data. An expressed intention of school-by-school reporting and reporting of individual student results is to give local constituencies, especially parents, the data they need to be informed about the quality of local schools and to lobby for program improvement.



Sample of student work: the North Carolina Grades 1 and 2 Assessment

Technical requirements. Accountability assessments may be similar in content to assessments used for monitoring trends. Both should be comprehensive measures of important learning goals. At higher grade levels, in fact, some states have school accountability systems that are also used to report state and district trends in achievement. Standards for reliability and validity are more difficult to meet for accountability purposes, however, because *standards for technical accuracy must be met at the lowest unit of reporting*. Thus, individual student scores must be sufficiently reliable, instead of just the state or district mean being reliable. Because

each individual score must be sufficiently reliable and valid, it is not possible to use the aggregation of scores to compensate for inaccuracies in individual measures. Individual-score reporting also precludes the use of matrix sampling to sample an assessment domain broadly. Instead, for fairness reasons, all students must take the same test.

The high-stakes nature of accountability assessments also contributes to their

standardized measures of children's physical, social, emotional, and cognitive development could be constructed and administered for purposes of program evaluation and monitoring trends-- because data aggregation would provide both

1. Before age 8, standardized achievement measures are not sufficiently accurate to be used for high-stakes decisions about individual children and schools. Therefore, high-stakes assessments intended for accountability purposes should

providing intensified special help if children are having difficulty, especially in learning to read.

Combining Assessment Purposes

There is a natural tendency for policymakers and educators to want to use assessment data for more than one purpose. The cost of developing new assessments would be better justified if the results could be used for multiple purposes, and if teachers and children go to the trouble of participating in an assessment, it would be desirable to get as much use from the data as possible. Many parents, teachers, and policymakers also want a combined system so that individual student results can be compared to standards set by the state or district. However, these desires for efficient use of assessment results must be weighed

Similarly, it seems reasonable to use the same assessments to serve Purposes 1, 3, and 4 on the grounds that all three involve measures of learning outcomes. However, reporting individual student and school-level data for accountability purposes (Purpose 4) requires a higher level of technical accuracy than the other two purposes, a level of accuracy that cannot be attained in large-scale programs for children younger than age 8. Therefore, the Resource Group has made quite

including accountability uses of assessment data.

Individual assessments, Purposes 1 and 2. In the past, screening measures intended as a first step in referral for special-needs identification have been misused for instructional purposes. For example, screening instruments designed to resemble short-form IQ tests have been used inappropriately to plan instruction or to hold children out of kindergarten. Although it would be possible, in theory, to develop assessments that could be used legitimately for both classroom assessment and screening for special needs (Purposes 1 and 2), extensive investment would be required to develop both curricularly relevant assessment content and empirical norms for evaluating disability.

To support teaching and learning (Purpose 1), assessment tasks should be as

to use the most generic tasks possible, so that all children from a wide variety of backgrounds will be equally familiar with the content of the assessment. Of course, this has not always worked even when seemingly familiar content was used; hence the problems of cultural bias.

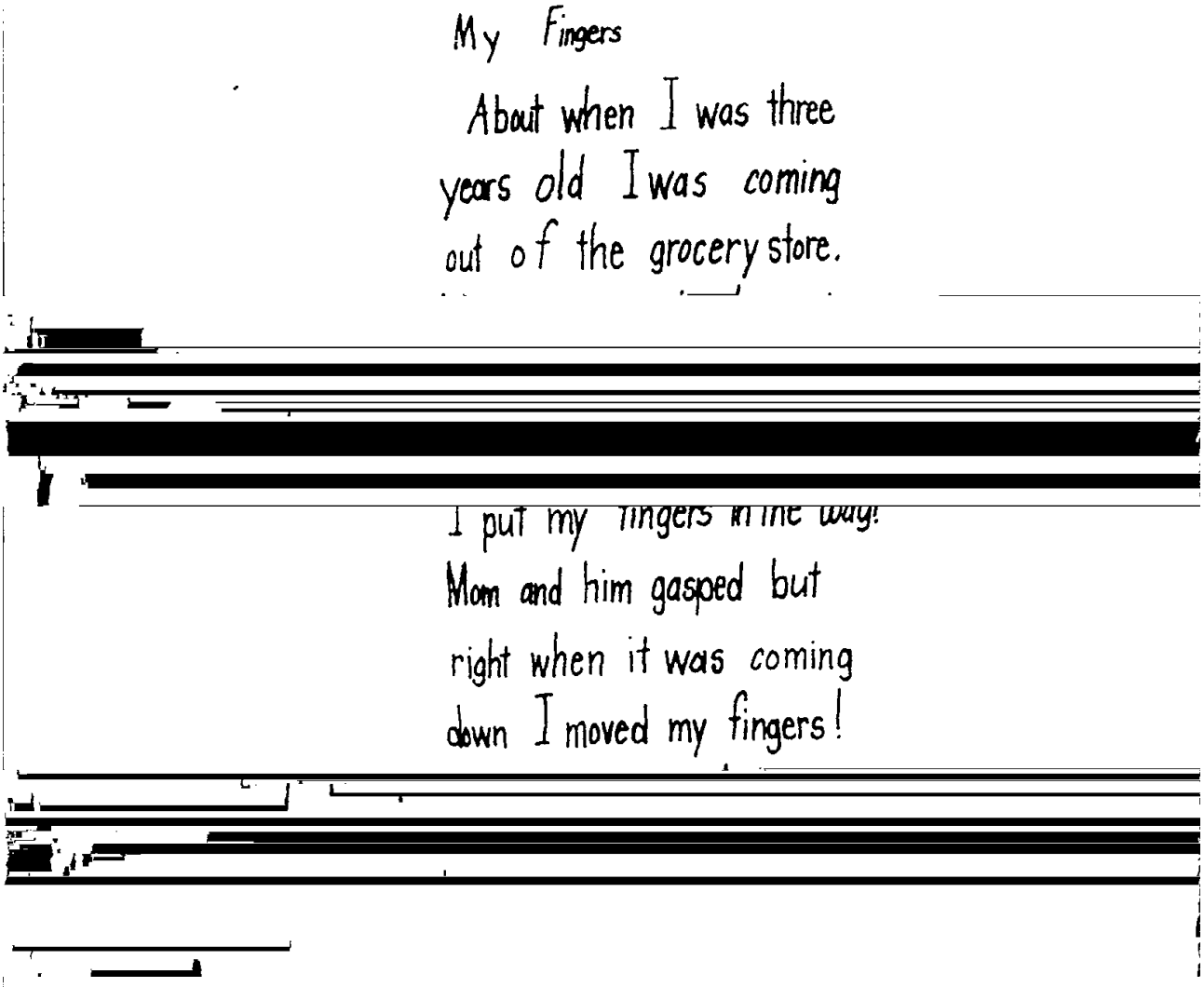
An alternative method of assessment for special-needs identification would

would need extensive training to use dynamic assessment with curriculum-aligned assessment tasks. We should also note that assessment materials intended for use in making special education placement decisions would require normative data and an empirical basis to support interpreting low performance as evidence of a disability, and would have to meet the more stringent reliability and validity standards for Purpose 2. In the meantime, the most appropriate policies are those that prevent the misuse of existing instruments.

curriculum. Use for accountability purposes also requires standardization of scoring across schools and rigorous external checks to make sure that the data being

States considering early childhood assessments to monitor trends (Purpose 3, a low-stakes type of program accountability) could, however, work to ensure that the content of assessments used for Purpose 1 is closely aligned with the content of the

for Purpose 1 and Purpose 3, but the two types of assessments could be developed in parallel so that they would be conceptually compatible and mutually supportive.



Sample of student work: the North Carolina Grades 1 and 2 Assessment

Conclusions

[redacted] child [redacted] is important both to support the learning of each individual child and to provide data—at the district, state, and national level—for improving services and educational programs. At the level of the individual child, teaching and assessment are closely linked. Finding out, on an ongoing basis, what a child knows and can do, helps parents and teachers decide how to pose new challenges and provide help with what the child has not yet mastered. Teachers also use a combination of observation and formal assessments to evaluate their own

preconditions of learning—such as the adequacy of health care, child care, and preschool services. Direct measures of children's early learning are also needed to make sure that educational programs are on track in helping students reach high standards by the end of third grade.

Assessing young children accurately is much more difficult than for older students and adults, because of the nature of early learning and because the language skills needed to participate in formal assessments are still developing. Inappropriate testing of young children has sometimes led to unfair and harmful decisions. Such testing abuses occur primarily for one of two reasons: either a test designed for one purpose is improperly used for another purpose, or testing procedures appropriate for older children are used inappropriately with younger children. In making its recommendations, the Resource Group has emphasized how technical requirements for assessments must be tailored to each assessment purpose, and we have tried to explain how the increasing reliability and validity of [redacted] for ages from birth to age 8 should guide decisions about what kinds

[redacted]

[redacted]

- **Assessing to promote children's learning and development.** The most important reason for assessing young children is to help them learn. To [redacted]

[redacted]

materials, to be used instructionally, that exemplify important and age- [redacted]

[redacted]

having a health or learning problem are referred for in-depth evaluation.

[redacted]

exclude children from school or to plan instruction. Often, the need for costly assessments could be eliminated if intensive language and literacy programs were more broadly available for all of the groups deemed educationally at-risk,

outside the classroom. Before age 9, assessment systems designed to gather data at the state or national level should focus on social indicators that describe the conditions of learning, e.g., the percentage of low-income children who attend quality preschool programs. Beginning at age 5 it is

- **Assessing academic achievement to hold individual students, teachers, and**

continuum of progress in Grades K, 1, and 2 that leads to expected standards

Congress charged the Goals Panel advisors to offer "clear guidelines regarding the nature, functions, and uses of early childhood assessments." In examining current trends in state and local policies, we found numerous efforts to guard against testing misuses of the past, as well as positive efforts to devise standards and assessments that would clearly document children's learning. We hope that these recommendations and principles will be useful to educators and parents, as well as

Glossary

Accountability: The concept of trying to hold appropriate parties accountable for their performance; in education these are usually administrators, teachers, and/or students. Beyond fiscal accountability, this concept currently means responsibility for student academic performance, usually by publicly reporting student achievement data (often test scores). Accountability mechanisms vary among states and local districts in the types of [illegible] but all are used and in the degree to which rewards, sanctions, or

assessment, but data on student attendance or homework completion [illegible] adult observations of student proficiency, or evaluations of projects, oral presentations, or other forms of problem-solving may also be assessments.

Child Find programs: Organized efforts by health, welfare, and education agencies to locate and identify children in need of special education services

Development: Growth or maturation that occurs primarily because of the emergence of underlying biological patterns or preconditions. The terms *development* and *learning* are distinguished by the presumption that one is caused by genetics and the other by experience. However, it is known that development can be profoundly affected by environmental conditions.

Developmental assessment: Measurement of a child's cognitive, language, knowledge, and psychomotor skills in order to evaluate development in comparison to children of the same chronological age.

Developmental continuum: A continuum that describes typical milestones in children's [illegible]

Formal assessment: A systematic and structured [illegible] student performance that both teachers and students recognize as an assessment event.

High-stakes assessment: Assessments that carry serious consequences for students or for [illegible]

results, but can provide useful insights about a child's learning.

Large-scale assessment: Standardized tests and other forms of assessment designed to be administered to large groups of individuals under prescribed conditions to provide information about performance on a standardized scale so that results for districts, states, or nations can be fairly compared

Learning: Acquiring of knowledge, skill, ways of thinking, attitudes, and values as a result of experience

Matrix sampling: A way to select a subset of all the students to be tested and subsets of various parts of a test so that each student takes only a portion of the total assessment, but valid [illegible] can be drawn about how all students would have performed on the entire test

Norms: Statistics or data that summarize the test performance of [illegible]

Readiness test: A test used to evaluate a student's preparedness for a specific academic program.

Reliability: The degree to which a test or assessment measures consistently across different instances of measurement—for example, whether results are consistent across raters, times of measurement, or sets of test items.

Screening: Selecting individuals on a preliminary test who are in need of more thorough evaluation.

Screening test: A test used as a first step in identifying children who may be in need of special services. If a potential problem is suggested by the results of a screening test, then a child should be referred for a more complete assessment and diagnosis.

Social indicator: A statistic (usually non-[illegible]

[illegible] special education is the specially designed instruction that public schools are required to offer either in a separate or regular classroom to meet the unique needs of a child with a disability.

Standardized test or assessment: Standardized [illegible]

[illegible] Education and Gall

Bibliography

Bredekamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs* (Rev. ed.). Washington, DC: National Association for the Education of Young Children.

Federal Interagency Forum on Child and Family Statistics. (1997). *America's children: Key national indicators of well-being*. Washington, DC: U.S. Government Printing Office.

Gredler, G.R. (1992). *School readiness: Assessment and educational issues*. Brandon, VT: Clinical Psychology Publishing Co.

Greenspan, S.I., & Meisels, S.J. (1996). Toward a new vision for the developmental assessment of infants and young children. In S.J. Meisels & E. Fenichel (Eds.), *New visions for the developmental assessment of infants and young children*. Washington, DC: ZERO TO THREE: The National Center for Infants, Toddlers, and Families.

Hills, T.W. (1992). Reaching potentials through appropriate assessments. In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Appropriate curriculum and assessment for young children* (pp. 43–63). Washington, DC: National Association for the Education of Young Children.

Kagan, S.L., Moore, E., & Bredekamp, S. (Eds.). (1995, June). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Goal 1 Technical Planning Group Report 95–03. Washington, DC: National Education Goals Panel.

Kagan, S.L., Rosenkoetter, S., & Cohen, N. (1997). *Considering child-based results for young children: Definitions, desirability, feasibility, and next steps*. Based on Issues Forums on Child-Based Results, sponsored by the W.W. Kellogg Foundation, the Carnegie Corporation of New York, and Quality 2000: Advancing Early Care and Education. New Haven, CT: Yale Bush Center in Child Development and Social Policy.

Langhorst, B.H. (1989, April). *A consumer's guide: Assessment in early childhood education*. Portland, OR: Northwest Regional Educational Laboratory.

Meisels, S.J. (1994). Designing meaningful measurements for early childhood. In B.L. Mallory & R.S. New (Eds.), *Diversity in early childhood education: A call for more inclusive*

testing. *Young Children* 42: 4–6, 68–73.

Meisels, S.J., with Atkins-Burnett, S. (1994). *Developmental screening in early childhood: A*

guide (4th ed.). Washington, DC: National Association for the Education of Young Children.

Meisels, S.J., & Fenichel, E. (Eds.). (1996). *New visions for the developmental assessment of infants and young children*. Washington, DC: ZERO TO THREE: National Center for Infants, Toddlers, and Families.

Meisels, S.J., Liaboe, J.R., Marsden, D.B., Dichtelmiller, M.L., & Dorfman, A.B. (1994). *The Work Sampling System*. Ann Arbor, MI: Rebus, Inc.

Meisels, S.J., Marsden, D.B., Wiske, M.S., & Henderson, L.W. (1997). *The Early Screening Inventory* (Rev. ed.) [ESI-R7R]. Ann Arbor, MI: Rebus, Inc.

Meisels, S.J., & Provence, S. (1989). *Screening and assessment: Guidelines for identifying young disabled and developmentally vulnerable children and their families*. Washington, DC: National Center for Clinical Infant Programs.

Michigan State Board of Education, Early Childhood Education & Parenting Office. (1992, April). *Appropriate assessment of young children*. Lansing: Michigan Department of Education.

Minnesota Department of Education. (1990). *Model learner outcomes for early childhood education*. St. Paul: Author.

National Association for the Education of Young Children and National Association of Early Childhood Specialists in State Departments of Education. (1991). Guidelines for appropriate curriculum content and assessment in programs serving children ages 3 through 8. *Young Children 46*(1): 21-38.

National Association for the Education of Young Children. (1988). NAEYC position

National Education Goals Panel. (1997, January). *Getting a good start in school*. Washington, DC: U.S. Government Printing Office.

National Education Goals Panel. (1997, October). *Special early childhood report 1997*. Washington, DC: U.S. Government Printing Office.

National Forum on Education Statistics. (1994). *A statistical agenda for early childhood care and education: Addendum to "A Guide to Improving the National Education Data System."* Adopted by the National Forum on Education Statistics, January 1994.

Neisworth, J.T. (1993). Assessment: DEC recommended practices. In *DEC recommended practices: Indicators of quality in programs for infants and young children with special needs and their families* (see EC 301 933).

Perrone, V. (1991). On standardized testing: A position paper of the Association for

development and learning. New York: Merrill, an imprint of Macmillan College Publishing Company.

Shepard, L.A. (1994). The challenges of assessing young children appropriately. *Phi Delta Kappan 76*(3): 206-213.

Shepard, L.A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools 34*(2): 85-97.

Shepard, L.A. (1991). The influence of standardized tests on the early childhood curriculum, teachers, and children. In B. Spodek & O.N. Saracho (Eds.), *Yearbook in early*

Goal 1 Advisors to the National Education Goals Panel

Technical Planning Group on Readiness for School

Leader: Sharon Lynn Kagan, Yale University

[redacted]

[redacted]

[redacted]

Annemarie Palincsar, University of Michigan
Tej Pandey, California State Department of Education
Catherine Snow, Harvard University
Maurice Sykes, District of Columbia Public Schools
Valora Washington, The Kellogg Foundation
Nicholas Zill, Westat, Inc.

Goal 1 Ready Schools Resource Group

Leaders: Asa Hilliard, Georgia State University
Sharon Lynn Kagan, Yale University

Barbara Bowman, Erikson Institute
Cynthia Brown, Council of Chief State School Officers

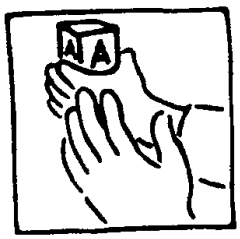
[redacted]

[redacted]

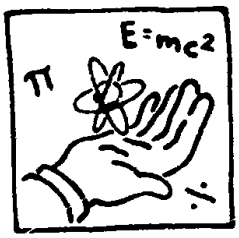
Lilian Katz, ERIC Clearinghouse for Elementary and Early Childhood Education
Michael Levine, Carnegie Corporation of New York
Evelyn Moore, National Black Child Development Institute
Tom Schultz, National Association of State Boards of Education
Barbara Sizemore, DePaul University
Robert Slavin, Johns Hopkins University

Typography and design by the U.S. Government Printing Office.
Editorial assistance provided by Scott Miller, Editorial Experts, Inc.

THE NATIONAL EDUCATION GOALS



READY TO LEARN

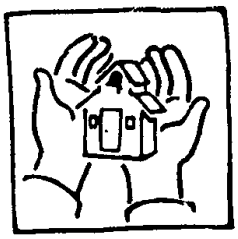


MATHEMATICS AND SCIENCE

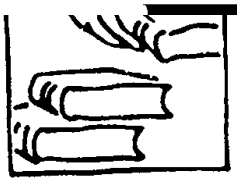
SCHOOL COMPLETION

ADULT LITERACY AND LIFELONG LEARNING

STUDENT ACHIEVEMENT AND CITIZENSHIP



SAFE, DISCIPLINED, AND ALCOHOL- AND DRUG-FREE SCHOOLS



PARENTAL

NATIONAL EDUCATION GOALS PANEL
1255 22nd Street, N.W., Suite 502
Washington, DC 20037
202-724-0015 • FAX 202-632-0957
http://www.negp.gov
E-mail: NEGP@goalline.org