DOCUMENT RESUME

ED 232 768 PS 013 730

TITLE	Instruments Developed in the Head Start Program Effects Measurement Project.
INSTITUTION	Mediax Associates, Inc., Westport, Conn.
SPONS AGENCY	Administration for Children, Youth, and Families (DHHS), Washington, D.C.
PUB DATE	May 83
CONTRACT	ACYF-105-77-1006
NOTE	32p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Early Childhood Education; Field Tests; *Interpersonal Competence; Program Descriptions; *Program Evaluation; *Test Construction
IDENTIFIERS	*Project Head Start

ABSTRACT

A test battery was prepared for use in assessing the effectiveness of Head Start and similar programs in fostering young children's development. The instruments were designed to measure program effects on several dimensions of the cognitive, social-emotional, and applied strategies domains. Specific competencies measured are presumed to define in operational terms Head Start's overall goal of developing participating children's social competence. Each of the instruments is theoretically grounded, reflecting explicit conceptualizations of the domain and dimensions of behavior measured. All reflect hypotheses concerning sequential or hierarchical processes of development. Further, the measures are designed to accommodate the cultural diversity of Head Start and similar populations, and to identify the different but equally valid paths along which individuals, racial/ethnic groups, and sex groups of children progress toward common goals. With reference to the test battery, this report describes the areas of child development measured, and lists principles and procedures used in developing the instruments. The largest portion of the report presents rationales for and characteristics of the measures. The concluding section describes pre- and post-field testing, preliminary analyses of the data, and revisions of the instruments. Finally, features of the planned final form of the test battery are indicated. (RH)

INSTRUMENTS DEVELOPED IN THE HEAD START PROGRAM EFFECTS MEASUREMENT PROJECT

Mediax Associates, Inc.
21 Charles Street
Westport, Connecticut 06880

May, 1983

This project was supported through funding by the United States Government, Administration for Children, Youth and Families (ACYF), under contract 105-77-1006. As of April 8, 1983, ACYF terminated its contract with Mediax "for the convenience of the Government." Thereafter, the project was conducted independently by Mediax with its own resources.

CONTENTS

I. INTRODUCTION

The Head Start Program Effects Measurement Project, begun in the fall of 1977, undertakes to prepare a battery of measures of program impact on the development of children between the ages of 3 and 7, on the effectiveness of the program in developing their "Social Competence"--i.e., children's success in everyday activities at home, at school and in the community. The measures are to be used by local administrators and teaching staffs in assessing the strengths and weaknesses of their own programs (Head Start and otner preschool programs and kindergartens) and in guiding such corrective steps as may be indicated.

The measures being developed in this project differ in several important ways from those previously used with early childhood populations. As is explained in greater detail subsequently, these measures are designed to assess the effects of the programs, not to evaluate individual children; they address the specific objectives of programs; they measure development over time, not status in terms of fixed norms; they are sensitive to different but equally valid paths along which children progress toward common goals; and they yield profiles across several areas of growth, thus providing comprehensive insight into developmental change.

The measures herein described are in tentative form, and in the process of evaluation and revision. The battery ultimately disseminated for use by programs will be constructed from items in the several instruments that survive critical analysis and appraisal.

Attention is given in this report to the areas of child development selected for measurement, guiding principles and procedures followed in preparing the measures, the rationales and characterisitics of the several measures, and the approach being used to refine the measures on the basis of the judgments of experts and extensive tests in the field.

II. AREAS OF CHILD DEVELOPMENT MEASURED

One of the early tasks of the project was to conceptualize the many facets of "child development" in terms that provide fruitful guidance to efforts at measurement. Important developmental characteristics of young children that warrant measurement were determined on three bases -- (1) a survey of early child-development scholars, conducted by J. McVicker Hunt, Senior Scientist of the project; (2) an analysis of relevant theoretical and research literature; and (3) mainly, the competencies important for Head Start to develop, as identified by Head Start parents and staffs and K-2 teachers of Head Start "graduates". In this latter connection, Mediax conducted a series of two-day Input Workshops in seven geographical regions of the country at which 375 participants listed and gave relative-importance ratings to more than 1,700 specific child characteristics that Head Start should seek to develop.

The important child characteristics identified through these procedures were organized into the following 4 broad domains and 21 subordinate dimensions of development.

I. Cognitive Development

1. Perception
2. Language
3. Reading
4. Math Concepts
5. Nature and Science
6. Social Organization

II. Social-Emotional Development

7. Sensitivity to Feelings of Others
8. Expression of Own Feelings
9. Self-Concept
10. Attitudes Toward Success in School
11. Independence
12. Sharing and Competing
13. Peer Relationships
14. Adult Relationships

III. Health and Physical Development

15. Health and Safety
16. Dental
17. Nutrition
18. Gross Motor
19. Fine Motor

IV. Applied Strategies

20. Task Competencies
21. Interpersonal Competencies

Additional characteristics were identified in the Aesthetic and Ethical domains of development. Mediax originally recommended that measures be prepared for all of these areas of child development, but available technology and budgetary limitations made this impractical. It was decided by the federal sponsor of the project, the Administration for Children, Youth and Families (ACYF), that measuring instruments be prepared for the several dimensions of attitudes, skills and knowledges outlined below.

I PRECURSORS TO INSTRUCTIONAL SUSCEPTABILITY

A. Social-Emotional

1. Interaction Attitudes (prosocial-antisocial)
2. Interaction Skills (sharing-competing, level of interaction)
3. School-Task Attitudes (attention-avoidance)

B. Applied Strategies

1. Task Attack Strategies (range and level)
2. Task Assistance Strategies
3. Organizational Competence (success in affecting others and in achieving goals)

II. COGNITIVE COMPETENCIES

1. Perception
2. Language
3. Reading
4. Mathematics
5. Nature and Science
6. Social Organization (subsequently replaced by Social Understanding)

This restructured set of competencies includes, to some extent, most of those originally recommended. The notable exception is the domain of Health and Physical Development. Some of the competencies in that domain are included in the cognitive measure of Nature/Science. It should also be noted that, as explained in Section VI, field tests of the Social Organization measure led to its omission from the battery, and to the inclusion of some of its items along with others in a new measure of Social Understanding.

All of the areas of development for which measures are being prepared are thought to be of critical importance to children's success in early schooling and to Head Start's overall goal of "Social Competence".

III. GUIDING PRINCIPLES FOR DEVELOPING MEASURES

The following general principles, previously summarized in the "Introduction", were established early in the project as guides to the development of measures.

1. The measures should be designed for program evaluation, i.e., the assessment of program effects on children's development; they should not be appropriate for the evaluation of individual children.

2. The measures should provide indices of change in children's development between "entrance" to and "exit" from the program (or within a designated program period).

3. The measures should be path-referenced, sensitive to the diverse paths along which children may develop toward common objectives.

4. Evaluative criteria should constitute "dynamic norms", reflecting the changing performance of children over time, rather than static status norms.

5. The measures should allow for culturally diverse manifestations of development in a given dimension, including multiple appropriate responses to the same stimuli in most instruments.

6. The measures should be formulated in terms that accommodate diversity among children. That is, to the extent possible, they should use illustrations with which children of different racial and social backgrounds can identify, language with which all children can be comfortable, instructions in English and Spanish, and scoring criteria that do not penalize children for responses in dialect, colloquial or other non-standard forms of expression.

7. The content of the measures should reflect both the objectives of Head Start as set forth in the program's Performance Standards, and the specific characteristics identified as important by parents and staffs and K-2 teachers in the Input Workshops conducted by Mediax.

8. Where appropriate, Spanish-language versions of the measures should be developed simultaneously with the English-language versions.

9. Multi-methods of assessment should be used to measure children's development in the several dimensions.

10. To the extent possible, the measures should use scales that are developmentally sequential or hierarchical.

11. The measures should be appropriate for children in the 3 to 7 age range.

12. The measures should be appropriate for administration by paraprofessional examiners after a brief period of training.

13. The measures developed initially should require approximately 20 minutes for each administration, and the overall battery should require between 2 and 2½ hours. Subsequently, the battery should be modified to require 45 minutes for 3-year-olds, and 60 minutes for 4-to-7 year olds.

14. The measures should adhere to applicable standards for measurement as stipulated in Standards for Educational and Psychological Tests and Manuals (1974) and as updated in the (Draft) Joint Technical Standards for Educational and Psychological Testing prepared by the joint committee of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (February 1983).

This set of guiding principles posed unusually demanding standards for test developers, and they were not fully satisfied in all of the instruments prepared. Even so, the impact of these guides resulted in measures that are more appropriate for their targeted use than any previously developed.

IV. PROCEDURES IN DEVELOPING MEASURES

There follows a summary description of the agencies involved and the procedures followed in developing the measures of this project.

Participating Agencies

Several consulting firms have participated for varying periods of time in developing the measures subsequently described.

The primary contractor is Mediax Associates, Inc., Herman P. Taub, Project Director. Mediax defined the taxonomy of children's competencies to be measured, and also developed the theoretical permises and general approaches used in the preparation of all measures.

After preliminary work by two other firms on measures in the Social-Emotional and Applied Strategies domains, Mediax assumed responsibility for completing the measures in these areas. To this end, the services of several consultants were engaged -- Barry J. Zimmerman, City University of New York, for measures of Sensitivity to the Feelings of Others; William L. Goodwin, University of Colorado, for measures of Sharing and Competing; and Martha B. Bronson, Brookline Early Education Project, for inventories of School-Related Social Skills and School Task Behaviors. Dr. Bronson also assumed overall responsibility for restructuring and coordinating measures in the Social-Emotional and Applied Strategies domains and, with Anthony Bryk, Harvard University Graduate School of Education, for field-test design and analysis.

Mediax maintains general oversight of the project, including the evaluation of all measures, editing and revisions, and the "packaging" and dissemination of the final battery.

ACYF contracted directly with three other agencies for the preparation of measures in particular domains, in cooperation with Mediax Associates. Two of them did the preliminary work in areas for which Mediax later assumed responsibility. The Urban Institute for Human Services, Jean E. Wofford, Project Director, developed the theoretical concept paper and tentative approaches to measurement in the Social-Emotional domain; and the Bank Street College of Education, Doris B. Wallace, Project Director, did the same for measurement in the Applied Strategies domain.

The third independent contractor, University of Arizona, John R. Bergan, Project Director, assumed responsibility for developing all measures in the cognitive domain. The University undertook directly to develop measures in the Perception and Math dimensions. It engaged the services of two sub-contractors for the other cognitive dimensions: University of California at Santa Cruz, Ronald W. Henderson, Project Director, for Reading and Nature/Science; and Indiana University, Sadie A. Grimmett, Project Director, for Language and Social Organization.

The draft measures prepared by the several contractors were refined by Mediax Associates, notably through the provision of a common format, illustrations modified by artists commissioned by the firm, adaptation of Spanish-language versions to dialects current in the United States, editorial corrections, and the purchase or construction of manipulatives and other stimulus materials.

A National panel of 16 members has exercised general oversight of the project since its early beginning. It includes practitioners in early childhood education, especially from Head Start, and child-development scholars with diverse areas of specialization. Together, they represent a broad range of expertise, experience and racial/ethnic populations. This National Panel has met once or twice a year to review developments in the project and to offer recommendations.

The general procedures by which the measures were developed are outlined below.

Concept Papers

A comprehensive paper was prepared conceptualizing the content and process of young children's development in each dimension. Based on exhaustive analysis of relevant theoretical and empirical literature, alternative models of hierarchical development were identified and/or hypothesized. The specific areas to be measured in the dimension were recommended, along with suggested procedures for such measurement.

Thus, each of the concept papers defined the rationale and general approach to the development of the measure for one dimension. The papers were submitted by their authors to ACYF and Mediax for comment and approval.

Item Formulation

Criteria were then defined for the formulation or selection of "items" (i.e., children's response tasks) to be included in each dimensional measure, such items being conceived as empirical indicators of developmental change in the content areas previously selected. The criteria called mainly for items that are compatible with the conceptualized model of developmental change and the guiding principles noted above.

On the basis of these criteria, tentative items for each dimensional measure were formulated anew or adapted from existing measures. These tentative items were then organized in draft manuals for use with children in testing periods of approximately 20 minutes.

Item Try-Outs

The draft measure for each of the six cognitive dimensions was tried out during the 1981-82 program year with a sample of several hundred Head Start children who were representative of the program's diverse population as regards sex, ethnicity and other background characteristics. The results were analyzed statistically for two purposes: (1) to identify items the instructions and/or content of which needed modification, or which should be eliminated entirely; and (2) to "scale" the items tentatively according to age.

In this latter connection, the draft items were organized into three age levels:

Level I - 3 to 5 years
Level II - 5 years to 6 years/6 months
Level III - 6 years/6 months to 8 years/6 months

Subsequently, largely because relatively few children over 6 years of age were included in the field test sample, Level III items were omitted from the several measures.

The several social-emotional measures were prepared very late in the project, only a few weeks before field tests, and they were tried out with a much smaller sample. Even these limited try-outs sufficed to identify a number of "bugs" calling for correction.

Experts' Evaluations

All of the measures have been subjected to continuous evaluation and revision since the item tryout versions (and their modifications) were received by Mediax during the summer and fall of 1982, and such appraisals will continue through 1983. Involved are critiques by project staff members, project consultants, and "outside" experts in the several areas of child development, together with rigorous tests in the field.

The nature and results of field tests are reported in Section VI. Attention is here called to evaluation of the measures by experts.

Seven scholars with established expertise in the areas of child development for which measures are being prepared were commissioned to appraise the draft instruments and suggest needed revisions. They and the measures they evaluated are listed below.

Perception: Charles Brainerd, Western Ontario University
Math: Merle Wittrock, University of California at Los Angeles
Nature/Science: Ronald Good, Florida State University
Reading: Roger W. Shuy, Georgetown University
Language: Richard Duran, Educational Testing Service
Social Organization: William Damon, Clark University
Inventory of School - Related Social Skills and Inventory of School Task Behaviors: Craig T. Ramey, University of North Carolina.

Each evaluator was asked:

1. To rate each item of the assigned measure as satisfactory (S) or unsatisfactory (U) on each of the following criteria:
    a. Valid indicator of development along the intended path?
    b. Appropriate level of difficulty for age group?
    c. Free of bias toward racial, ethnic, sex groups?
    d. Measures something different from other items (i.e., not redundant)?
    e. Appropriate for administration by paraprofessionals?
    f. Appropriate for scoring by paraprofessionals?
2. To explain briefly the reason for each unsatisfactory (U) rating.
3. To suggest revised or alternative or additional items for the dimension.

4. To express judgments regarding the following general questions:

   a. Are the content areas "covered" by this measure appropriate and adequate for assessing program effects on the development of children aged 5 to 7? Specifically: Are some of the content areas inappropriate for this age group? Are essential and appropriate content areas omitted? Explain. Offer suggestions for needed corrections.
   b. In general, is the form in which the items are cast appropriate for the content involved and for children aged 5 to 7? What, if any, alternative format do you recommend?
   c. Does the measure seem to reflect a valid conceptual framework of children's development in the dimension? Explain.
   d. What other evaluative judgments and/or suggestions do you offer for improving this measure?

The evaluators were generally positive in their critiques of the draft measures. However, two of the measures--Nature/Science and Reading--were adjudged conceptually inadequate; and revisions were made to improve the measures. Specific criticisms were offered for all of the measures, together with suggested corrections. These reports by evaluators were appraised by project staff and passed on to test developers for appropriate action. The revised measures will be resubmitted to the evaluators with supporting rationale and data. However, none of these expert evaluators bears any responsibility for the content and form of the measures eventually prepared.

V. RATIONALES AND CHARACTERISTICS OF THE MEASURES

Separate measures are being prepared for each of the dimensions of the cognitive domain outlined in Section II (except that Social Understanding is substituted for Social Organization). One of the cognitive measures (Nature and Science) now incorporates items assessing development in health, safety and nutrition. Two instruments are being prepared to assess development in the several dimensions listed for the Social-Emotional and Applied Strategies domains. Both are measures that tap several dimensions of child development in each of these domains.

A. General Characteristics

Common to all of these measures, in somewhat varying degrees, are the general characteristics noted in the preceding discussions of guiding principles and procedures. They also share several other general characteristics.

With few exceptions, the manual for each measure includes (1) an "Introduction", which states its purpose and rationale; (2) an outline of the subtests that constitute the measure, with the numbers of the items that relate to each subtest and the age levels of children for whom they are intended; and (3) the items to be administered, together with instructions (in English and Spanish), lists of materials, and scoring criteria. English and Spanish instructions are arranged in parallel columns on the same page to allow for bilingual administration where necessitated by the child's language proficiency.

In order to assess accurately the development of children from the wide range of cultural backgrounds served by Head Start, the test items of all measures are designed to maximize the likelihood that children will understand what they are to do, and to encourage them to show what they know. Instructions are given in simple language; unscored "practice items" introduce many subtests; many items posit game-like situations; and illustrations depict objects and events generally common in the environments of Head Start children.

Two types of materials are used with most of these measures. Predominant are pictures of objects, scenes in nature and society, events, people, etc., drawn by artists commissioned by this project. In some cases they are enclosed in separate "picture binders"; in others, they are interspersed among items of the manual. Manipulatives constitute the other type of materials. They are objects of various kinds, to be handled by children or by examiners in the view of children -- e.g., blocks, geometric forms, toy trains and cars, colored strips of paper, rocks, coins, paper clips, puppets, plates and play food, and many more.

Pictures of children are used extensively in some measures. They are artists' sketches depicting youngsters of different racial groups.

Whereyer possible, examiners record the child's actual responses, thus providing a basis for the analysis of error patterns. In the case of items where this is not possible, children's responses are simply scored as right or wrong. On the observational measures, examiners record the occurrence of defined behaviors, making possible analysis of the frequencies and proportions of different categories of behavior. The scores for the measure are then recorded in vertical columns, by items, on the front side of a score sheet. On the back side of this sheet, the examiner checks several groups of statements to indicate significant behaviors of the child during the testing session -- (1) problems (e.g., loud noises) that may have affected the child's performance; (2) selected behaviors of the child (e.g., "attentive", "uncooperative", "overly talkative", "very interested", etc.); and (3) the examiner's perception of the appropriateness of the "preferred language" (English or Spanish) selected for use in administering the instrument to Hispanic children. Both the item scores and the behavioral checks are designed for optical scanning.

The manual for each measure is "packaged" in a hard-cover, loose-leaf binder in which items are grouped by age levels. The binder, specially designed for this project, can be made to stand A-shaped between the child and the examiner, with pages bearing pictures facing the child and pages bearing related instructions facing the examiner.

In addition to manuals for the several measures, there is also a Data Collectors Manual, prepared by Mediax, that provides detailed instructions for administering the instruments. Addressed to examiners, it includes sections on "Introduction", "Overview of Project Organization", "Description of Data Collection Tasks", "Maintaining Relations with Teachers, Other Head Start Staff and with Parents", "Cost Control Procedures", "Coping with Special Situations", and "Questions and Answers".

Video tapes have also been prepared to provide instruction and practice in administering and scoring each measure, and in interpreting results.

These are general characteristics of the whole group of measures. There follow for the measure or measures in each dimension of child development: (1) a brief summary of rationale, consisting largely of quotations from the related concept paper and "Introduction"; (2) a list of subtests; and (3) illustrative types of response tasks children are called upon to perform.

B. Cognitive Measures

PERCEPTION

Rationale: The concept paper that guided the development of the Perception measure defines "perception" as "that subtest of cognitive processes involved in extracting information from physical stimuli which serves to facilitate the construction of higher order concepts." It is conceived as "a process existing on a continuum with both sensation and cognition, rather than as a separate category of behaving."

The model here used "includes four levels of perceptual processing, in four degrees of alteration, which represent sequential transformations made on a stimulus by the processor. Level 1 encompasses the detection of information; Level 2 the representation of that detected information, involving extraction of relevant features; Level 3 the storage of the essential features extracted; and Level 4, extrapolation beyond the information provided in the stimulus itself."

"There are two basic ways in which the model relates to development. The first (A) relates to the proficiency with which the individual can accomplish different processes given a consistent stimulus. The second (B) involves the range of possible stimuli to which a given process can be applied. It seems reasonable to predict that children will differ on both (A) and (B) as they develop and increase in perceptual proficiency."

Thus, as noted in the "Introduction", perception is much more than simply reacting to physical stimuli; it involves deriving meanings, and it is a developmental process. Its importance lies in the fact that "one cannot expect successful completion of a cognitive task unless the task-relevant information is processed, a perceptual act. Cognition, therefore, presupposes perception, and the latter serves as the basis for accomplishment of the former."

The content component of the Perception measure is reflected in two forms of items: "(1) the perception of temporal/auditory information, and (2) relations among units (e.g., a series of sticks arranged on a size dimension). All of the items represented in this test have been systematically selected to tap a structure which represents the kinds of content categories and the relations between them."

The set of perceptual skills measured by this instrument "are precisely the kinds of performances relevant to instructional priorities--those to which educational experiences are directed. Therefore, they have educational relevance in the larger picture (i.e., being prerequisites for other academic behaviors), and in the smaller picture, which are the educational goals of Head Start."

Subtests: The subtests of the Perception measure are listed below.

Judging Shape and Form
Judging Size and Length
Working With Spatial Relations
Working With Perspective Relations
Building Visual Patterns
Seriation

Types of Response Tasks: For purposes of illustration, some of the types of tasks children are called upon to perform are listed below.

Match pictures of geometric forms on cards.
Construct geometric form with stips to match form on card.
Identify picture that shows how an object looks from two perspectives: the position of the child and the position of a doll.
Match cards by shape.
Observe card with geometric form for 3-4 seconds, then (with the card face down) identify that form on another card.
Rotate a triangular disc to match changing position of triangle attached to face of a clock.
Observe card with bar of given length for 3-4 seconds, then (with card face down) identify bar of same length on card with four bars of different lengths.
Construct red and whict block patterns to match model on card.

MATHEMATICS

The concept paper on which the Math measure is based examines, among other questions, competing theoretical issues concerning the developmental structure of early mathematical knowledge. Notable among them are process-content issues, process-competence issues, and issues related to developmental change. For reasons there fully explicated, the model adopted for this project does not separate content from process, but relates the two. By tieing process to content, this approach is capable of representing "cases in which processes may be applied across different task contents"; and it is also able "to identify the limits of generality of processes that do apply to more than one content category".

Further, the position is here taken "that there is an advantage to considering both competence and process in the assessment of mathematical knowledge. Information about process can provide an indication of how competence is achieved."

Still further, as regards developmental change: (1) it is here assumed "that it can be useful to conceptualize developmental sequences in terms of the processes underlying mathematical task performance; (2) it seems advisable to include task representation as a variable in the construction of hypothesized hierarchical sequences", since "the way in which children represent mathematical tasks may affect hierarchical ordering"; and (3) although the study of errors may be useful for individual diagnosis, analysis of "intellectual processes including performance errors will be limited to processes that reflect developmental progress that can be assessed in program evaluation."

The "Introduction" to this instrument states that "the measures in the math dimension are organized in three broad areas: working with numbers, working with shapes, and working with measurement units. These areas in turn are divided into subtests, each reflecting a separate set of skills in the dimension. The content reflected in the subtests is designed to articulate directly the Head Start goals in mathematics." These measures are also "designed to overcome some of the shortcomings apparent in conventional achievement tests insofar as they assess mathematical competencies (e.g., conservation) shown by developmental research to be fundamental to the mastery of mathematics skills."

The purpose of this measure "is not just to determine the extent to which children know more at the end of instruction than they did at the beginning. Rather it is to ascertain qualitative changes in children's cognitive skills . . . The subtests in the math dimension are designed to make this possible. Items assessing developmental skill variations are included in the measures to make them sensitive to developmental change. For instance, counting tasks include counting forward, counting backward, and counting by multiples (e.g., by two's)."

Since children may solve mathematics problems in different ways, the measures are also "designed to be sensitive to diversity in development and to reveal alternative paths to development when they exist."

Subtests: The subtests of the Math measure are listed below.

Numerical Recognition
Math Signs
Conservation of Number
Recognizing Set Size
Numeration
Addition
Subtraction
Multiplication
Division
Recognizing Shapes
Money
Time
Ordination

Types of Response Tasks: Illustrative types of items in the Math measure are the following:

Recognize numbers of blocks, math signs, etc.
Count objects.
Add, subtract, multiply and divide--with objects and verbally.
Recognize circle, square, rectangle, triangle.
Recognize same or different number of blocks in two groups.
Recognize coins of different value.
Recognize comparative value of different groups of coins.
Tell time from pictures of a clock.

NATURE AND SCIENCE

Rationale: The purpose of the Nature and Science measure is to assess those aspects of children's knowledge of objects, events and relations that contribute to a growing understanding of science and the processes that science uses to discover, describe and explain the natural world. The measure originally developed and pilot-tested in the fall of 1982 placed too heavy a reliance on verbal responses and multiple-choice items, using drawings. Following the advice of the expert's evaluation, the instrument was re-conceptualized to focus on the processes of

science, using tasks that actually involve the child in active observing, manipulating and discovering with a variety of objects and situations. Two major sources provided the main ideas for translating this purpose into items: the writings of Piaget, and the specific experience of the Process Instrument originally developed by the American Association for the Advancement of Science in the early 1960's.

Thus, the Nature and Science measure is also based on a particular orientation to the child's role as learner. According to the original concept paper on measurement in this dimension, "the most important principle emanating from Piaget's work, and the most robust factor reflected in our conceptual framework for the Nature/Science dimension is the view that the young child is an autonomous, active, self-discovering learner, involved in the first-hand manipulation of physical phenomena." This means that the measure de-emphasizes factual scientific knowledge. Although some of that is included, we recognize that the ability to name something is only a superficial kind of knowledge, whereas knowing "how" to do something represents a more fundamental competency. Therefore, the bulk of the Nature and Science items actually engage the child in operations that will indicate competency in carrying out "scientific" processes.

Subtests and items. The Nature and Science measure is not divided into specific subtests, and many of the items assess more than one process. The processes involved are observing, describing, classifying (grouping), explaining, predicting, and measuring. The content to which these processes are applied are living and inanimate objects, energy and force relations, biological processes and functions, and seasonal relationships. In addition, children's knowledge of health, safety and nutrition are assessed, using techniques that also require the processes of observing, describing, classifying and explaining.

For many items a range of responses is possible. The scoring system reflects this range, rather than being simply a right-wrong procedure. This also means that most of the same items can be administered to children of varying ages, with developmental and learning differences being reflected in different scores obtained. Therefore, the experimental version of the Nature and Science measure administerd in the spring of 1983 does not have two separate levels as the other measures do. It should also be noted that scientific processes other than those listed above, such as ordering and using spatial relations, overlap constructs measure by the Perception instrument; hence, they are not included in the Nature/Science Measure.

Types of Response Tasks: There follow descriptions of three illustrative types of responses children are called upon to perform by the Nature/Science measure.

In one item, the child is presented with 9 squares of fabric that can be sorted into 3 groups, either by material (wool, nylon, cotton) or by color (blue, white, print), and is asked to sort the fabrics into 3 groups. (Either classification is acceptable.) The child is then asked to sort the fabrics by a different criterion. Then a new square of fabric is presented for the child to place in the correct group.

In another item, the child is shown pictures of rectangles of varying lengths, but close enough in size to prohibit accurate comparisions. The child is asked to use a white card with colored markings to measure the rectangles and determine which is longer, shorter, the same length as the green mark, etc.

In still another item, a car is placed on an inclined plane, so it will "roll down this hill". The plane is tilted further to "make the hill steeper"; and the child is asked whether the car will roll "faster than before, just the same, or slower". After responding, he/she is asked: "How would you check to see if you were right?"

READING

Rationale: The concept paper on the development of competence in reading and pre-reading includes substantial analysis of "the two theoretical traditions dominating reading theory and curriculum organization . . . often referred to as bottom-up and top-down". It concludes that "an ideal model is one that provides a coherent description of test-driven or bottom-up processes and reader-driven or top-down processes in reading". The proposed model is said to provide "a resolution to the apparent conflict in the two views of reading performance". How this accommodation is reflected in the Reading measures is explained in the "Introduction" to the manual.

"On the one hand, reading was viewed as a text-driven, or bottom-up process which is controlled by textual input. Learning to read involves translating graphics into speech with the focus on decoding the written symbols into speech sounds. Comprehension then occurs as oral language processes take over. During early reading acquisition, proponents of this position place an emphasis on students' concepts of units of language and their ability to manipulate those units. In adherence with this position, subtests of the reading dimension measure the students' ability to identify and manipulate language units.

"In the other position, top-down, the reader becomes a much more active participant. Meaning is gained through a process of hypothesis formation, data sampling, and confirmation. Readers use their knowledge of the world and language to gain an understanding of the text. This view places a greater emphasis on the purposes and processes of print within the context of the students' environment. The reading subtests adjust to this position by including knowledge of the language of instruction, understanding the purposes of print, and the use of semantic and syntactic knowledge.

"Thus the subtests within the reading dimension attend to both theoretical positions. The working assumption was that reading involves an integration of readers' knowledge and goals within the intended message on the printed page. The subtests sample a sequential attainment of decoding skills, along with concepts related to the top-down theoretical position."

The concept paper organizes "the structure of reading knowledge into four broad categories called reading production, comprehension, utility, and writing production". The subtests of the reading measure tap reading-related behaviors in each of these areas, with varying emphasis corresponding to emphases in the Head Start goals and curriculum.

Subtests: The subtests of Reading measure are noted below.

Capital and Lower Case Correspondence
Knowledge of Print Process
Word Reading
Naming Letters

Orthographic Structure Knowledge
Rhyming Concepts
Auditory Segmentation
Cloze
Writing Production
Word Segmentation

Types of Response Tasks: Illustrative types of tasks children are called upon to perform in the Reading measure are listed below.

Name letters.
Read words from a list.
Recognize different syllables of a spoken word.
Recognize pictures the names of which rhyme.
Tell what word is left if part of it (e.g., "cow" in "cowboy") is taken away.
Recognize part of own name missing as pronounced by examiner.
Identify (from picture and text) what people "look at when they read."
Recognize errors in spelling own name with letters on table, etc.; tell how to correct.
Supply missing word in sentence read by examiner.
Write on a blank sheet of paper (e.g., letters, numbers, sentences, stories or just scribble).

5. LANGUAGE

Rationale: The concept paper in the Language dimension reviews competing theories of language acquisition in young children, and opts for "a functionalist view of language-- a focus on how the child brings language to bear to meet the demands of the situation in which language is used."

"The key to this approach is the notion that grammatical structure cannot be understood outside the context in which language is used. The functionalist approach holds that grammar is a secondary or derived system, related to the constraints of the communication task".

This point of view is especially important for the assessment of development in Head Start children. Here, even more than in other cognitive dimensions, assessment must cope with cultural diversity. "Language is learned within a child's culture, and children coming from different cultures will use language in ways that reflect their different cultures".

As regards assessment, "the following assumption about the goal has been made: we wish to know the level at which the individual child is capable of using language in a given situation." It is important, therefore, "to devise situations in which the child needs to use language, and then to score the level of what the child does". This focus "precludes the traditional assessment of isolated linguistic forms". Moreover, "the functionalist approach to language assessment mandates an emphasis on the child's spontaneous production (as opposed to comprehension or imitation of language)"; because "production of language appropriate for context clearly implies the ability to imitate or comprehend that language".

Reflecting this point of view, the "Introduction" to the Language manual specifies three kinds of language competencies the measures are designed to tap. The first major area is that of semantics, or "What Words Mean". Emphasis is on "the use of those relational words which play so important a part in the child's overall cognitive development. These reflect expression of spatial, temporal, causal, conditional, class inclusion, and hierarchical relations."

The second competency category is syntactics, or "How Words Work Together". "Verb tenses and other inflectional word endings as well as sentence complexity are stressed."

The third major area of competency is pragmatics, or "Using Words to Communicate." This component includes two subcomponents--(1) "conventional situations in which knowledge of the rules which guide conversation are assessed"; and (2) "Telling Things to Others", which "taps the child's skill in story telling and handling hierarchical and sequential elements in stories", and also giving directions.

Subtests: The subtests of the Language measure are listed below.

Show Me
Same and Different
School Time
After School
What Would You Say
Giving Directions
If and Unless
How Stories are Put Together
Telling About Pictures
Before And After
Explanations
Comparing (English only)
Changing Words (English only)
Cambiando Las Palabras
Encontrando La Palabra Correcta

Types of Response Tasks: Illustrative of some of the types of children's response tasks posed by items in the Language dimension measure are the following:

Presented with a doll and car, child is directed: "Show me: The doll pushes the car."

Shown pictures of boxes containing geometric forms, child is instructed: "Point to the box where the pictures are the same;" also where the pictures are different.

Child is shown three boxes containing different numbers of cupcakes. Examiner points to box with 2 cakes, saying that it "has some cupcakes"; points to box with 5 cakes, saying that it "has even _____ cupcakes"; and points to box with 9 cakes, saying that "it has the very _____ cupcakes," Child supplies missing words.

Child is told: "You are walking home from school with your friend. When you get to your house, you and your friend walk inside and see your mother." Child is then asked: "What is the first thing that you should say to her?" Similarly: "If your mother does not know your friend, what should you say to her?"

Presented with a puppet, the child is instructed: "Tell Sandy how to play this game. Remember, he can't see the game, so you have to tell him about it."

Child is instructed to use toy telephone to "call your friend and ask him/her if he/she can come and play with you."

Examiner reads "stories" to child. E.g., "The frog is sitting on a log in the stream. Then he jumps into the water." Child is instructed to select pictures and arrange them in correct sequence to "tell the story with pictures."

Examiner tells a "story" and asks child what happens next. E.g., "If it is sunny the children will go to the zoo, unless it is cold outside. Today is a cold day. What will the children do?"

With manipulative objects at hand, the child is asked to perform certain tasks. E.g.: "Put a penny in the cup after you put the button in."

Child is instructed to use pictures and manipulable accessories to depict a sentence the examiner reads. E.g.: "The boy wearing the hat waves to a friend carrying the bag."

UNDERSTANDING OF SOCIAL RELATIONS

The instrument designed to measure children's Understanding of Social Relations seeks to elicit responses that provide insight into children's knowledge of generally-accepted conventions guiding relations with others, sensitivity to the feelings of others, and patterns of sharing and cooperating. It is organized in three parts.

Part I, Social Roles and Rules, consists of 6 items and taps a child's knowledge of social roles and rules and taking turns. It involves role-playing and the use of dolls and other objects.

Part II, Interpersonal Perception of Affect, includes 4 items that "call upon the child to respond to brief stories by selecting a social representation of four possible emotions: happiness, sadness, fear and anger." One panel of pictures shows the faces of four children, each expressing one of these emotions. The child is told a brief "story" about Johnny or Nancy, and asked to point to the face showing how he/she would feel in the situation described.

Part III, The Pictorial Scale of Sharing, is an 8-item measure of children's prosocial behaviors in the area of sharing and helping. Sharing is defined as "the giving up or dividing of material possessions, human relationships, time or skill, or the communicating of ideas, information or feelings to someone else". In each item, the child is presented with a panel of pictures showing what "some children" might do in a defined situation. The child is then asked what he or she would do in that situation, and the choice is recorded.

C. Social-Emotional and Applied Strategies Measures

The Social-Emotional and Applied Strategies measures prepared by this project are designed to asses the early development of children's school-related attitudes and overt behaviors that are not tapped by the cognitive measures. Thus it is that the several dimensions of these two domains are characterized in Section II as "Precursors to Instructional Susceptability". They seek insight into the nature and quality of children's social interactions and approaches to cognitive tasks that are hypothesized as critical for effective performance in school.

Predominant among traditional assessment in this general area is the use of attitude scales or inventories in which children's verbal responses to selected stimuli are interpreted as evidence of hypothesized abstract "traits". Relevant research findings suggest, however, that young children's attitudes tend to be mercurial, and the validity of their self-reports highly suspect. Moreover, the conception that some unitary sets of attutudes and behaviors are here involved (e.g., "attitude toward school") is theoretically questionable.

In the light of these and related considerations, this project uses two complementary approaches to measuring children's development in the Social-Emotional and Applied Strategies domains, without any assumptions about whether they reflect some unitary traits. First, observational records are made of children's "school task" behaviors as they respond to cognitive tests. Second, observational records are made of children's "school-related social skills" as they react to structured situations of social interaction.

SCHOOL TASK BEHAVIORS

The Bronson Inventory School Task Behaviors uses structured observational categories and trained observers to assess the behaviors of individual children in structured task or test situations. As explained in its "Introduction":

"Coping effectively with a structured task or test situation requires that a child be able to respond appropriately to a (possibly unfamiliar) adult, to a (possibly) novel situation or setting, and to a variety of different tasks which vary in interest, familiarity, and difficulty. The child must be able to listen attentively to instructions and directions, to resist distraction and discouragement, and to respond with effort and persistence to the demands of each task. In order to manage tasks successfully the child must be able to understand the requirements of the task, check or scan and notice the relevent features of the task, organize task relevant materials when necessary, use an organized systematic plan of attack in complex tasks, and correct errors or try again when difficulty arises."

"Competence in structured tasks or tests requires both a repertoire of appropriate strategies and the motivation or willingness to try the task. This instrument provides categories that reflect these two aspects of performance. It also includes several categories designed to record the child's evaluation of his or her own ability and performance within each task. The self-evaluation component is included in the Inventory because it may be related to the self-concept and thus to the child's willingness to try and to persist in task or test situations."

Four major categories of behavior are observed and recorded: (1) RESPONSE TO TASK, (2) TASK AVOIDANCE BEHAVIORS, (3) TASK ATTACK STRATEGIES, and (4) OUTCOME. The components of these categories are listed on the following page. Precise definitions and illustrations are provided the observer for each sub-category of behavior to be observed, together with detailed procedural instructions.

Child's Name: _____

Class: _____

The Bronson Inventory of School Task Behaviors

Observer: _____ Code:Z__ Date: ___/___/83

CATEGORIES	Perception :	ITEMS											
RESPONSE TO THE TASK	Attends to Instructions	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Answers Too Soon	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Tries Task on Request	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Tries Task with Encouragement	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Requests Help	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Requests Clarification	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Requests Evaluation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Evaluates Self -Positive	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	-Negative	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
TASK AVOIDANCE BEHAVIORS	No Response/Ignores (Passive)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Resists/Refuses -Verbal	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	-Physical	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Becomes Distracted	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Irrelevant/Off Task Comment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Requests to Stop/Leave	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Cries	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Uses Materials Inappropriately	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Moves Excessively in Seat	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Leaves Seat	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Other (Describe Below)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
TASK ATTACK STRATEGIES	Verbalizes Rules/Requirements	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Organizes/Groups Materials	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Uses Systematic Approach	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Checks/Scans Carefully	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Notices Features of Task/Materials	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Corrects Error	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Tries Again/Starts Over	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
OUTCOME	Completes Successfully	☐			☐			☐			☐		
	Completes Not Successfully	☐			☐			☐			☐		
	Starts but Does Not Complete	☐			☐			☐			☐		
	Does Not Start	☐			☐			☐			☐		

COMMENTS:

This instrument is administered while children respond to test items selected from cognitive measures. It does not matter whether the child responds correctly or incorrectly; the observer is concerned only with how he or she behaves while responding.

SOCIAL-SKILLS

The Bronson Inventory of School-Related Social Skills is used to obtain insight into nature and quality of children's behaviors in their relation with other children. It is administered as randomly-selected pairs of children interact in two structured situations.

In the first situation, "Building Together," pairs of children are given 10 small red squares and 10 small blue squares of DUPLO, together with 10 red and 10 blue rectangles of DUPLO. They are instructed: "Build something together that you would like to build. Use the red and blue blocks to build something together." This situation lasts five minutes.

In the second situation, "The Farm," each child is given a part of the manipulatives in the Fisher-Price "Family Play Farm" set, with the silo removed. One child is given the barn, with the fence, the feeding trough and the horse cart inside. The other child is given the other toys in a bucket--including a baby, cradle, stroller, playpen and 2 small dinosaurs (about the same size as farm animals). They are instructed: "You can play with them together or by yourselves." This situation lasts ten minutes.

The observational instrument defines school-related social skills as "the ability to become involved in organized social interaction with others, the ability to use positive social strategies to influence others or solve social problems, and the ability to act effectively and successfully to influence others and solve social problems." Thus, three major categories of behavior are observed and recorded: (1) INVOLVEMENT categories, (2) SOCIAL ORGANIZING STRATEGIES categories, and (3) SOCIAL ACCOMMODATING STRATEGIES categories. The components of these categories are listed on the record form on the next page. Precise definitions and illustration are provided for each sub-category of behavior to be observed, together with detailed procedural instructions.

The psychometric properties of the cognitive, social-emotional, and applied-strategies measures described in this section are being tested through empirical evaluations based on pretest and posttests in the field.

CHILD'S NAME: _____

SEX: M / F BIRTH DATE: __/__/7_

CLASSROOM: _____

BRONSON INVENTORY OF SCHOOL-RELATED SOCIAL SKILLS

SITUATION: _____ OBSERVER: _____ CODE:Z__

OBSERVED AT _____ A.M./P.M. ON ___/___/83

OTHER CHILD'S NAME: _____

CATEGORIES		MINUTES:	1	2	3	4	5	6	7	8	9	10	11	12
INVOLVEMENT	SOCIAL	ORGANIZED	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		NOT ORGANIZED	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	NON-SOCIAL	INVOLVEMENT WITH MATERIALS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		WATCHING	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		NO INVOLVEMENT	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		OTHER (DESCRIBE BELOW)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

CATEGORIES	SUCCESS IN INFLUENCING OTHERS:		SUCCESS 1	2	3	4	NO SUCCESS 5	6	7	8	NOT APPLIC. 9	10	11	12
SOCIAL ORGANIZING STRATEGIES	SUGGESTS/DIRECTS ACTIVITY		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	ASSIGNS ROLES OR RESOURCES		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	STATES RULES		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
SOCIAL ACCOMODATING STRATEGIES	HELPS	SPONTANEOUS - SUGGESTS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		- AGREES									☐	☐	☐	☐
		- REFUSES									☐	☐	☐	☐
	SHARES	SPONTANEOUS - SUGGESTS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		ALLOWS - AGREES									☐	☐	☐	☐
		- REFUSES									☐	☐	☐	☐
	TAKES TURNS	SPONTANEOUS - SUGGESTS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		ALLOWS - AGREES									☐	☐	☐	☐
		- REFUSES									☐	☐	☐	☐
	TRADES	SPONTANEOUS - SUGGESTS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	BARGAINS/BRIBES	- POSITIVE	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	(THREATENS)	- NEGATIVE	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	ASSERTS RIGHTS		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	COMPETITIVE COMMENT										☐	☐	☐	☐
											☐	☐	☐	☐
	RESISTS/IGNORES	- CHILD									☐	☐	☐	☐
		- ADULT									☐	☐	☐	☐
	USES PHYSICAL FORCE/TAKES/GRABS		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
			☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	SHOWS HOSTILITY	- VERBAL									☐	☐	☐	☐
		- PHYSICAL									☐	☐	☐	☐
	ASKS INFORMATION	- CHILD	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		- ADULT	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	ASKS HELP	- CHILD	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
		- ADULT	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

COMMENTS: _____

VI. EMPIRICAL EVALUATION

Attention is called in Section IV to the process of continuous evaluation and revision of the measures being prepared in this project, with emphasis on appraisals by selected experts and project staff. The instruments are also being subjected to rigorous evaluation on the basis of empirical data assembled through pretests and posttests in the field.

Pretests In The Field

Pretests of the cognitive measures were conducted during the fall of 1982 and winter of 1983 with a representative sample of approximately 3,000 Head Start and K-2 children in 19 sites, located in urban, suburban and rural communities in 17 states around the country. The social-emotional measures were tested in 2 of these sites.

The pretest sample included 2,370 children in Head Start classes and 379 children in school grades K-1. Ninety-five per cent of the Head Start children were in classes where 5 or more children were tested, thus permitting classroom-level analyses. They were selected to satisfy two basic criteria: (1) all children in a class except those moderately or severely handicapped or whose parents did not give permission; and (2) all children in a class who meet particular cell design requirements (e.g., membership in a particular age-ethnic-language group). In addition, 59 children were used for practice and certification of data collectors.

For purposes of these field tests, site managers were selected and trained; and they, in turn, selected and trained paraprofessional data-collectors, who administered the tests.

Operations in the field were guided by a Data Collectors Manual, previously described, and a Site Manager's Manual, both prepared by Mediax. The manual for Site Managers includes sections on recruiting and hiring data collectors, training, data-collection operations (including monitoring and transmittal of data), and cost control procedures.

Score sheets for children tested in the field were mailed daily to Mediax, where they were entered into a computer, recorded on diskettes, and sent air-freight to the University of Arizona for analysis.

The research design for appraisal of the draft measures on the basis of field tests was prepared by the University of Arizona project director, John Bergan, and Mediax consultant, Anthony Bryk, Associate Professor of Evaluation, Measurement and Statistics at the Harvard University Graduate School of Education, and Project Director of the Huron Institute.

Preliminary Analyses and Revisions

Several types of analysis were made of pretest data.

1. The construct validity of the whole battery of draft measures, initially appraised by expert evaluators, was further examined through content analysis of the extent to which all of the measures combined include items

that correspond to the specific knowledges and skills and attitudes identified as desirable objectives of Head Start by participants in the Input Workshops conducted early in the project. The analysis revealed a very substantial correspondence. Although the test items do not reflect all of the desirable competencies identified by the Input Workshops and include some items designed to assess competencies not so identified, the high degree of correspondence between the two warrants the judgment that the whole group of instruments address in very large measure the developmental competencies that Head Start parents and staff and K-2 teachers think the program should foster.

2. Estimates of concurrent validity are being made through analysis of children's pretest performance on the draft measures and two independent but related instruments: Preschool Inventory and Metropolitan Readiness Test (Kindergarten). These are widely used measures of outcomes related to school readiness. Posttests on these measures with the same children are being conducted this spring. The data collected will be analyzed to determine (a) whether there is strong correlation between children's performance on the project measures and the two independent measures, and (b) whether the newly-developed measures are sensitive to children's growth.

3. Analysis was made of the reliability of scoring for the several instruments by having the Director of Field Operations and the Site Manager score pupils' responses as they observed data collectors administer and score the tests. The percentage of agreement among these three scores (or pair of scores) was computed. Such parallel scores were obtained for more than 100 administrations; and analysis revealed high percentages of agreements, generally in the 90's. Although the percentages tended to vary somewhat among the data-collectors and among the several instruments, the generally high level of agreement among different scorers was adjudged acceptable evidence of reliability. It is anticipated that the more rigorous monitoring of posttest administrations will result in even higher levels of inter-scorer agreement.

4. The internal consistency of each measure--the degree of correlation among its items--is being analyzed, using Kuder-Richardson internal consistency estimates and parallel forms reliability estimates. If the measure addresses a true construct, its several items are expected to correlated highly with one another--to "hang together" as a unit. The results of this analysis will provide further evidence indicative of the reliability of the several measures.

5. The extent to which each test item discriminates among children of different ages was determined by comparing the percentage of correct responses for children at 3-month age intervals. Items which do not differentiate in the expected direction among such age groups were modified or eliminated.

6. The extent to which each item of each measure discriminates among children with varying ability was determined through use of latent trait techniques (BICAL). Children's ability levels were estimated on the basis of their performance on each measure as a whole. Test items that do not discriminate among children of different ability levels were modified or eliminated from the measure.

7. The relative difficulty of the items of each measure was analyzed by its several "task strands" (e.g., Letter-Word concepts, Visual Memory, Naming Isolated Letters, Orthographic Stucture Knowledge, etc. in the Reading measure). For this purpose, the median number of children's correct responses on each item was plotted on an item-difficulty scale ranging from +4.9, through .0, to -4.7, on which the median difficulty levels for all children 3, 4, 5 or 6 years old are designated. These distributions were inspected for each task strand to determine whether (a) the items represented a substantial range of difficulty, without gaps; (b) some items were obviously "too easy" or "too difficult"; and/or (c) some items tended to cluster at a given difficulty level, revealing little or no differentiation in this regard. Desirable, of course, is the continuous distribution of the items of each task strand across the median ability levels of the several age groups. Some items were modified or eliminated and some items were added in an effort to approximate such distributions.

8. Latent trait techniques were also used to identify test items that reflect bias toward sex, racial/ethnic or language groups. Controlling for ability levels, differences in the percentage of correct responses to an item by subgroups of children of the same age were interpreted as evidence of bias. Items reflecting bias were modified or eliminated.

9. Following the collection of pretest data, a Field Test Personnel Survey was conducted "to collect feedback information from individuals actually involved in the use of the measures during field testing. The survey was designed in such a way as to preclude responses if an individual had not been trained and had not administered the (test for the) dimension in question. In other words, the responses should only reflect the results of opinion informed by experience".

The survey was conducted by mail, and the arbitrary cut-off date was set as January 31, 1983. Responses were received by that time from 11 (58%) of the 19 site managers, and from 44 (55%) of the "current" data collectors as of mid-December, 1982.

The survey instrument was a 35-item check-list on which respondents rated each cognitive measure on a number of specific criteria relating to its (a) Item Wording (English), (b) Art Work, Graphics, (c) Manipulative Materials, (d) Administerability, and (e) Spanish Text. (There were also some survey items concerning General Issues, e.g., packaging of the measures, training tapes, etc.). Ratings were made on a 4-point scale, ranging from unacceptable quality (1) to high quality (4).

Results of the survey showed that the site managers and data collectors who administered the measures in the field perceived real differences among them on all of the five groups of criteria. The Reading measure received the highest ratings in three areas: Artwork and Graphics, Manipulatives and Administerability; but it was rated lower than any of the others on the quality of the Spanish Text. The Math instrument received the lowest average rating for Wording, Art Work/Graphics, quality of Manipulatives, and Administerability. No measure was rated wholly unacceptable on any of the five groups of criteria. Project staff re-

examined each measure in the light of these ratings by field personnel, and made revisions designed to correct perceived weaknesses as regards the specific characteristics adjudged inadequate.

10. A substudy of the instructional sensitivity of the battery of measures was begun during the pretest period and will continue into the posttest period. Approximately 300 children in 30 classrooms of 3 sites were administered the measures at four time intervals of about 6 weeks. Analyses of children's performance at these successive points will provide external evidence on whether the battery is sensitive to continuing instruction over time, an important consideration for judging the validity of the measures for program assessment. It will also provide important psychometric evidence of the change score characteristics of the new measures.

11. Several investigations are being conducted to shed light on the measures' sensitivity to program-related variables. In one substudy, classrooms at four sites were observed using the CDA Checklist, a measure of "overall classroom quality" that taps the dimensions on which Head Start personnel are trained and assessed in the Child Development Associate credentialing program. Analyses are being made to determine whether the new instruments relates to classroom process in expected ways. A second substudy involves administering a survey to teachers in order to determine the extent to which their classroom instruction emphasizes areas that are assessed by the measures. Analysis of their responses will provide a validation of the content of the measures in terms of current Head Start practices in participating classes

On the basis of these analyses, many revisions were made in the measures used in pretests in the field—minor in some cases, major in others; and, as previously noted, the Social Organization measure was withdrawn. The revised versions of the measures are being administered in posttests currently under way.

Posttests in the Field

Posttests of the several measures are being administered to a selected sample in a number of field sites. The following cognitive measures are being administered at all sites:

Perception
Math
Nature/Science (including health, safety and nutrition)
Understanding Social Relations
Reading
Language

The following applied-strategies and social-emotional measures are being administered at a subset of the sites.

Bronson Inventory of School Task Behaviors (used in observing children's behaviors while responding to one of the cognitive measures).

Bronson Inventory of School-Related Social Skills (used in observing children's behaviors while responding to two situational measures of social interaction).

Sites that administered the Preschool Inventory and Metropolitan Readiness Test during the pretest period will administer these instruments to the same children during the posttest period; and the substudy of instructional sensitivity begun during pretests will be concluded during the posttest period.

In addition to these outcome measures, several related instruments are being administered for the purpose of assembling data that will help assess the validity of the measures, and also provide insights into factors influencing children's performance on them. These instruments include:

1. Teacher Rating Scale -- an instrument that consists of 30 terse descriptions of child behavior in the social-emotional and applied-strategies domains. Teachers are asked to rate the degree to which each behavior is "like" that of a named child. Their ratings are analyzed as a partial check on the validity of related outcome measures.

2. Classroom Staff Questionnaire -- a 12-item inquiry form designed to obtain information on the training and experience of teachers of the Head Start classes used in field tests.

3. Content Validation Survey -- a 77-item check-list of items in the several cognitive and social-emotional measures, each of which teachers rate on two bases: (A) relative emphasis in their classrooms, and (B) relative emphasis desirable in the project measures. Ratings of test items are made on a 4-point scale:

   Level 1. Not emphasized at all
   Level 2. Slight emphasis
   Level 3. Important emphasis
   Level 4. Most emphasized

4. Family Questionnaire -- a 21-item check-list on which the parents or guardians of children provide (and mail back to Mediax) SES and related information about their children's backgrounds.

5. Family Background Data Report -- a 1-page inquiry form on which data gatherers assemble similiar SES information from Head Start records.

6. Mobility and Retention Report -- a 2-page inquiry form on which data gatherers assemble from Head Start records information on the extent of and reasons for mobility of the children in the field-test sample.

Preparation of the Final Battery

The data assembled from posttests of the cognitive and social-emotional measures and the several related instruments will make possible many additional analyses and assessments of the instruments. On the basis of those analyses and assessments, indicated revisions will be made of all measures.

The final battery of measures prepared for use by Head Start and other early child-development programs will reflect these revisions. It will include items that field tests have demonstrated to be valid, reliable, sensitive to instructional programs and that paraprofessionals can administer and score effectively.

It is anticipated that this final battery will be administered in three sessions, requiring a total of approximately one hour with each child. There will be start-stop rules that vary with children's ages and previous test experiences, thus permitting each administration to be tailored to the individual child.

The instruments tentatively planned for administration in each of the three sessions are listed below.

Session 1. Perception, Reading and School-Task Behaviors
Children (a) respond to the cognitive measures of Perception and Reading, and (b) their behaviors are observed and recorded as they respond to the Perception measure.

Session 2. Language, Understanding Social Relations, and Social Skills
Children (a) respond to the cognitive measures of Language and Social Understanding, and (b) participate in two social-interaction situations as their behaviors are observed and recorded.

Session 3. Mathematics and Nature/Science (including Health, Nutrition and Safety)
Children respond to these two cognitive measures.

On the basis of children's performance on the above outcome measures, their development in the following dimensions will be assessed.

I. PRECURSORS TO INSTRUCTIONAL SUSCEPTIBILITY

A. Social-Emotional
1. Interaction Attitudes (prosocial-antisocial)
2. Interaction Skills (sharing-competing, level of interaction)
3. School-Task Attitudes (attention-avoidance)

B. Applied Strategies

1. Task Attack Strategies (range and level)
2. Task Assistance Strategies
3. Organizational Competence (success in affecting others and in achieving goals)

II. COGNITIVE COMPETENCIES

1. Perception
2. Language
3. Reading
4. Mathematics
5. Nature and Science
6. Social Understanding

The measures will be accompanied by a manual of detailed instruction for administering and scoring each part of the battery. There will also be a video tape of children taking each part of the battery, illustrating the range of behaviors to be observed and scored. This will facilitate the self-instruction of teachers on how to use the measures.

The scoring sheet for each part of the battery, along with instruction in the manual, will facilitate immediate interpretations by the teacher, and also provide for more detailed central analysis interpretation.

The use of these measures with preschool and kindergarten children will yield data and experiences on the basis of which further developmental efforts will be undertaken -- to improve the psychometric properties of the instruments; to facilitate administration, scoring and interpretation of the measures through further application of advanced technology; and to extend beyond kindergarten the age levels of children for whom the battery may appropriately be used.

VII. SUMMARY

The Head Start Program Effects Measurement Project is preparing a battery of instruments for use in assessing the effectiveness of Head Start and similar programs in fostering young children's development in several dimensions of the cognitive, social-emotional and applied strategies domains. The instruments are designed to measure program effects, not to evaluate individual children.

The specific competencies measured were identified early in the project, mainly on the basis of relative-importance ratings by Head Start parents and staff and kindergarten-through-second-grade public school teachers assembled in a series of regional workshops in different parts of the country. They are presumed to define in operational terms Head Start's overall goal of "Social Competence."

Thus, these measures, unlike any previously available, are addressed directly to the child-development objectives Head Start seeks to achieve. Moreover, they are designed to accommodate the cultural diversity of Head Start and similar populations, and especially to identify the different but equally valid paths along which individuals and racial/ethnic and sex groups of children progress toward common goals.

The instruments are theoretically grounded. They reflect explicit conceptualizations of what constitute the several domains and dimensions of behavior measured, together with hypotheses concerning the sequential or hierarchical processes of development involved.

A common set of procedures guided the preparation of all measures. Preliminary drafts of the instruments were tried out with samples of Head Start and public school children, revised, submitted for evaluation by experts in the several dimensions of child development, and revised again. The draft measures were then pretested in the field with approximately 3,000 Head Start children in 19 sites around the country, and revised again. They are currently being posttested in the field, and will be revised still further on the basis of analyses of findings. All measures are designed for administration and scoring by paraprofessionals. A central scoring and interpretation service will also be provided by Mediax.

The final battery of measures will be administered in three sessions requiring a total of about one hour of testing time with each child. A manual accompanying the measures will provide users with detailed instructions for administering, and scoring each part of the battery. There will also be a video tape of children responding to the measures, useful as an aid in self-instruction of program staff on how to use and interpret the instruments.

The Head Start Program Effects Measurement Project was begun about six years ago by Mediax Associates, under contract with the Administration for Children, Youth and Families, U.S. Department of Health and Human Services. Several other agencies and many professional consultants have participated in its development, and a National Panel of scholars and Head Start practitioners have monitored the project throughout. The project is now being carried to completion by Mediax Associates, without further funding by the Federal Government.

It is anticipated that the final battery of measures will be available for use by Head Start, other preschool programs and kindergartens in the fall of 1983. Further developmental efforts are projected to improve the psychometric characteristics and facilitate the use of the measures, and also to make them appropriate for children beyond the kindergarten level.