R E P O R T R E S U M E S

ED 020 790 PS 000 991

A STUDY OF A MEASUREMENT RESOURCE IN CHILD RESEARCH, PROJECT HEAD START.

BY- JOHNSON, ORVAL G. BOMMARITO, JAMES
SOUTHERN ILLINOIS UNIV., EDWARDSVILLE

REPORT NUMBER OEO-2458
EDRS PRICE MF-$0.75 HC-$6.16 152P.

DESCRIPTORS- *MEASUREMENT INSTRUMENTS, MEASUREMENT TECHNIQUES, RATING SCALES, TESTING, EDUCATIONAL RESEARCHERS, INFANCY, *EARLY CHILDHOOD, LITERATURE REVIEWS, *SURVEYS, PERSONALITY TESTS, ACADEMIC APTITUDE, ATTITUDE TESTS, PERCEPTION TESTS, SOCIAL RELATIONS, *TESTS, *TEST RELIABILITY, APTITUDE TESTS, ACHIEVEMENT TESTS, HEAD START,

MEASURES OF CHILD BEHAVIOR AND CHARACTERISTICS, NOT YET PUBLISHED AS SEPARATE ENTITIES, WERE COLLECTED THROUGH A PAGE-BY-PAGE SEARCH OF ISSUES OF 46 JOURNALS (LISTED IN APPENDIX A) PUBLISHED DURING THE PERIOD OF JANUARY 1956 TO DECEMBER 1965 AND 50 RELEVANT BOOKS. CORRESPONDENCE WITH RESEARCHERS AND AUTHORS OF MEASURES YIELDED ADDITIONAL MEASUREMENT RESOURCES. AS PRESENTED IN THE REPORT, THE MEASURES WERE GROUPED INTO SIX KINDS, (1) DEVELOPMENT, ACADEMIC APTITUDE, AND ACHIEVEMENT, (2) PERSONALITY, (3) ATTITUDES, (4) SOCIAL INTERACTION AND SKILLS, (5) PERCEPTUAL SKILLS, AND (6) MISCELLANEOUS. THE LISTING FOR EACH TEST INCLUDED ITS NAME, THE AUTHOR, THE AGE OF THE POPULATION FOR WHOM IT WAS DESIGNED, THE GENERAL AREA OF INTEREST, THE TYPE OF MEASURE, AND THE SOURCE FROM WHICH A COPY OF THE MEASURE MIGHT BE OBTAINED. A DESCRIPTION OF THE MEASURE (OFTEN QUOTING ITS AUTHOR) INCLUDED SAMPLE ITEMS AND AN OUTLINE OF THE ADMINISTRATIVE AND SCORING PROCEDURES. WHEN AVAILABLE, RELIABILITY AND VALIDITY DATA WERE BRIEFLY SUMMARIZED. A BIBLIOGRAPHICAL REFERENCE WAS PROVIDED FOR EACH MEASURE. (MS)

A Study of a Measurement Resource in Child Research

Project Head Start

Office of Economic Opportunity

Contract No. OEO-2452

Orval G. Johnson, Ph.D.
Project Director

James W. Bommarito, Ed.D.
Project Co-Director

Southern Illinois University
Edwardsville Campus
Edwardsville, Illinois 62025
Contractor

INTRODUCTION

The fields of psychology, child development, education, and related fields have contributed their share to the information explosion within recent years. Workers in these fields have found increasingly difficult the task of keeping up with the literature, particularly if their interests are not narrowly defined. Researchers find the review of the literature involving more and more references from a wider variety and number of journals. As the fund of information increases, the information retrieval problem becomes more complicated, requiring more and more central sources to which the scholar can go to save time and effort.

A fast-growing number of researchers are planning and carrying out research with young children. Many of them are conducting experimental and evaluative studies in connection with government- and foundation-sponsored programs. Much child research is sponsored or carried out without any outside sponsorship in universities, state institutions, private schools, and public schools. Attention has been focused most recently on the culturally deprived child.

The rigorousness of the research depends upon the quality of the overall design and upon the extent to which the terminology of the study can be made operational. Many of the variables with which researchers deal in studies of young children are operationalized in the form of scores on tests, rating scales, frames of reference for observation, structured interviews, etc.

Unfortunately, since these measurement devices are so widely scattered in various sources, there is a vast duplication of effort from study to study in ferreting out of the literature the data-gathering instruments or techniques which are pertinent to each investigation. There is at present

no satisfactory basic source to which researchers can go for the specialized information they need on measures suitable for children. The Mental Measurement Yearbooks (O. K. Buros) cover widely varied types of tests for all ages. Other disadvantages of the Yearbooks for child research purposes are that they are limited to published tests and the series is revised infrequently. Tests in Print (O. K. Buros) while including more tests for preschool age children, has the same drawbacks as the Yearbooks. Both are excellent beginning points in an intensive search for specialized data-gathering techniques, however. A comprehensive up-to-date compendium of instruments and techniques usable with young children would meet a growing need, by simplifying the search for them, which presently consumes so much useful time and energy of researchers. In addition, this compilation would have heuristic value, stimulating the planning of evaluation and measurement phases of research.

This total project was initiated under Grant Number 1 R11 MH02347-01 from the National Institute of Mental Health. This report covers that portion of the work done under the current contract with the Office of Economic Opportunity.

PROCEDURE

The task of finding as many unpublished measures of child behavior and characteristics in the literature as possible gave rise to two important questions. What kind of broad screening procedure could be set up to find the largest number of measures, within certain limits of staff and time? Secondly, within the screening procedure mentioned above, how might we be assured of not missing useful measures?

It was our experience that the usual reference sources were not appropriate to the task which we have set for ourselves. Psychological

Abstracts, The Education Index, The Annual Review of Psychology, The Review of Educational Research, and similar sources, while useful in providing leads, were nevertheless not appropriate as basic documents from which to work.

It was decided that we would search the journals in the areas of psychology, psychiatry, and education, which appeared likely to yield measures of child behavior and characteristics. It seemed obvious, and our experience confirmed this, that a page by page perusal of the materials would be necessary to extract the information we needed. Article titles, while frequently helpful, simply did not give us enough clues that there might be a useful measure described in the article. We found neither the table of contents nor the yearly index to any of the journals useful for our purposes. Every page of every journal which we searched was scanned for references to measures, whether in that article or elsewhere.

Every volume of every journal over the period January 1, 1956 to December 31, 1965 was searched, whenever the journal encompassed that span. If it was a new journal with its beginning somewhere in this ten-year period, the issues from the beginning to December 31, 1965 were scanned. Forty-six journals were scanned. The list of journals is included as Appendix A to this report. In addition to the journals, approximately fifty books were used because they were known to have descriptions of measure in them. Additionally, measures came to our attention from various sources, including friends who were conducting research using measures which they had constructed, and who kindly offered them for inclusion in this compilation.

When, using the above-described search procedure, we found a measure mentioned, we looked for a reference to it in either Buros' Mental Measurement Yearbooks, or in his Tests in Print. If we did not find it in either

of these sources, we then looked in the article or in the bibliography (if any) for a copy of and a description of the measure. Whenever possible, which was in practically all cases, we had a copy of the measure as used by the author. Frequently the measure itself was published in the original article where we first saw a reference to it. This was obviously the most convenient situation. If the measure itself was not available in the original article or in the usual reference sources, we wrote a letter to the author. This is a story in itself, both in the magnitude of the task and in the nature of the response. The response to our request for information was excellent. With very few exceptions, colleagues were generous in giving their time to writing answers to specific questions we had, and in general, for seeing that we got all the information we needed to write a meaningful description of a measure. A few letters probably never reached the addressee, a few people simply did not respond, even after a follow-up letter; and a few authors indicated that they wished to copyright and distribute their measures themselves. We tried, in these cases, to point out that they could be their own agents whether or not the measure appeared in our compilation, but we were not always successful.

When we had all the necessary information on a measure, we were ready to write our description. After trying out a couple of formats for the descriptions, we finally settled on a nine-point format, set up as follows:

The first item of information which we list is the name of the test. In every case where the author had given his measure a name, we used the one which he had given. If the author had not named his instrument, we searched the literature with regard to the measure to find a word or phrase which aptly described it. In a very few cases, it was necessary to concoct a name. Some measures, on the other hand, were known by several names.

which resulted in duplicate files until, in the process of working with the measures, it became clear that the two and sometimes more names referred to only one measure.

In general, there was little difficulty with the problem of authorship. Occasionally, a researcher developed the instrument and described it in a study authored by him and several colleagues. Then, later on, he authored singly a study in which he used the measure. Sometimes, it was not clear whether he alone was the author or whether the measure had been authored by the larger group.

Another class of information which we included in the description of every measure is the age of the population for whom the measure is intended. Occasionally, authors specified the age level for which their measure appeared appropriate. In most cases, however, it was necessary to use as the appropriate age for the measure the age of the population who were the subjects in the studies conducted using the measure. Thus, it could very well be that many of the measures are applicable to ages outside the limits indicated in the age range for the test as given in the test description.

Many of the studies did not specify in any percise terms what was the variable measured by the instrument. In a large number of cases, therefore, the statement of the variable being measured is the judgment of the person who did the test description, which would be one of the authors of this report.

The type of measure was frequently a difficult matter to determine, in large part because there appears to be no well-accepted over-all classification of measure that could be used. Since we were concerned with other types of measures in addition to tests, we could not use some of the classifications used for tests alone. Whenever possible, we used any description which was given by the author. We tended to fall back on some

commonly used measurement terminology, frequently using phrases to qualify the classification of a measure when we felt the classification did not quite fit.

The sixth point which is covered in the test descriptions is the source from which the measure may be obtained. This was both one of the simplest and one of the most difficult pieces of information to get about a measure. If a measure was contained in one of the items in the bibliography, there was obviously no difficulty in directing the potential user to the proper source. On the other hand, if the measure had never been published intact in any readily available source, it was necessary to contact the author and attempt to get his permission to make the measure available in some way. Most authors were notably cooperative, many of them offering to provide the measure free to anyone who wanted it. Frequently with more complicated measures, especially if they involved any equipment, the author specified a price for the measure.

A large part of the description is made up of the section entitled, "Description of Measure." This section frequently expanded on the "Type of Measure" section and in every case where it was feasible, which was almost all cases, samples of the items were quoted. This section usually also contained an outline of the administration and scoring procedures.

The eighth point in the description format is that relating to the reliability and validity of the measure. Data on reliability and validity would be crucial to the potential user of a measure, and they were summarized briefly in this section. Whenever information on reliability and validity was available from several studies, the data were included in the test description in this section.

The final section was the bibliography, which was made up on the basis of the information that we had coming from the original source of the measure

and from reports by the author that the measure had been used by another researcher whose reference he provided for us. The bibliography is not intended to be a complete bibliography for the use of any measure.

MEASURES OF DEVELOPMENT, ACADEMIC APTITUDE, AND ACHIEVEMENT

TEST DATA WORKSHEET

Name of Test ABC Inventory

Author Normand Adair and George Bleach

Age 4 years 9 months through 4 years 11 months

Area (i.e. language development) School readiness

Type of measure Individually administered paper and pencil test

Source from which the test may be obtained Educational Studies and Development
1357 Forest Park Road
Muskegon, Michigan 49441

Description of Measure

The ABC Inventory consists of a series of tasks comprising four sections which are orally administered to the child. The responses required of the child consist of naming tests, common knowledge, simple drawings, elementary counting skill, and so forth. According to the authors, the inventory is of sufficient simplicity to be administered by unskilled help, such as teacher's aids, and each requires only eight or nine minutes to complete. In scoring the results, raw scores may be converted to "readiness ages" by means of a conversion table.

"The principal purpose of the ABC Inventory is to identify children who are immature for a standard school program. Operationally, the inventory serves best when used in pre-school screening and has been designed with this function in mind. Aims in developing inventory were to (1) devise a screening technique that was reliable and valid; (2) construct a format that was easily managed by inexperienced examiners; (3) outline administration, scoring and interpretation procedures that were direct and uncomplicated; (4) maintain economy by minimizing equipment needs and time consuming procedures; and (5) be suitable to children in the pre-school age range."

Evidence of reliability, validity, and standardization

"Item analysis, weighting and refinement continued through 1962. The standardization group throughout consisted of boys and girls whose ages ranged from 4 years 9 months through 4 years 11 months. The average age was 4 years 10 months. No effort was made to separate the scores of boys and girls or to make allowances for socio-economic differences. The fifteen schools involved in the study included K-12 systems in areas with 400 total enrollment to moderate sized systems enrolling over 5,500. Because the number in the standardization study is large (N=166) and the age range small, biases in selection are believed to be diluted.

"Reliability was established by matching comparable groups and assuming group equivalency. Scores for children of the same age who enrolled in the same school districts two years apart were compared with the following results:

	1962 group	1964 group
No. in Both Groups	166	314
Means for Both Groups	65.51	66.71
Standard Deviations	22.66	21.78
Standard Error of Means	1.76	1.23
Mean Differences	1.20	
Standard Error	2.08	
Critical Ratio	.58	

According to Table of t, no significant difference between means.

"Validity was determined by comparing 'pass-fail' features between children in the upper and lower half of the score distribution. Eighty-three in the standardization group obtained scores 63 and above while 83 scored below 68. Forty-three children failed their first year of school. Of those failing, 37 or 86% were identified accurately. Seventy-seven or 63% passing, scored above 68 (tetrachoric correlation .70).

"A ready age scale was constructed by combining all test scores over a three-year period (N=619) into a frequency distribution and examining its resemblance to a normal bell-shaped curve. Features of divergence from symmetry were studied for significance. The test for skewness and kurtosis was not significant. Deviating scores were measured from the mean and on a basis of their percentile rank were given an age index. This index, called a 'readiness age,' approximates the mental age feature described for other tests and carries similar implications. This treatment was tested by comparing the ABC Inventory ready age with the Stanford-Binet mental age. In a small sample study (N=14), the product moment correlation between ready age and mental age equaled .78. Investigation of this relationship is being extended and subsequent statistical measures may not yield so high a correlation."

Bibliography

Adair, N., and Blesch, G. The ABC Inventory. Muskegon, Michigan, Educational Studies and Development, 1965.

TEST DATA WORKSHEET

Name of Test — Adjective-Noun Paired Associates Test

Author — Allan Paivio

Age — Fourth and fifth grades

Area (i.e. language development) — Verbal learning

Type of measure — Paired-associates word lists

Source from which the test may be obtained — See Paivio for examples

Description of Measure

"The learning material consisted of lists of 16 adjective-noun paired associates. The lists were constructed from words occurring 50 or more times per million according to Thorndike and Lorge (1944). Thirty-two adjectives and equal numbers of relatively concrete and relatively abstract nouns were selected such that each adjective was a relevant modifier of either a concrete or a more abstract noun. Since only a few high-frequency nouns could be found which are abstract in the sense of having no objective referents and yet are appropriate for children the concrete and abstract nouns chosen differ mainly in generality:... the concrete nouns refer to specific things or events while the abstract nouns are more general in their referents (e.g., meat versus food). The items (listed in the order: adjective-concrete noun, abstract noun) are: Wooden-Box, Meal; Soft-Chair, Spot; Bright-Star, Idea; Yellow-Corn, Grain; Clean-Dress, Clothes; Gentle-Boy, Child; Pretty-Rose, Flower; Beautiful-Queen, Lady; Proud-Indian, People; Heavy-Lead, Load; Loud-Shot, Noise; Sharp-Pencil, Stick; Green-Ball, Toy; Deep-Valley, Place; Equal-Half, Share; Fancy-Shoes, Things; Open-Window, Space; Hungry-Bear, Animal; Early-Morning, Time; Brave-Hunter, Fellow; Hard-Iron, Metal; Sweet-Orange, Fruit; Stone-Castle, Building; Fresh-Meat, Food; Paper-Dollar, Money; Cold-Wind, Air; Soft-Music, Sound; Tall-Oak, Tree; Happy-Christmas, Holiday; Wise-Man, Person; Busy-Bee, Bug; Large-Army, Crowd.

Four lists, each consisting of 16 paired associates, were constructed from the above items, the assignment of items to different lists being random. Within each list, eight of the adjectives were paired with concrete nouns and eight with abstract nouns. Since each list could be presented with the pairs in either the A-N or N-A order, eight different lists were used in the study."

The author contends the procedure has potential relevance "...to theories of learning as well as to the practical problems of language...(Paivio, 1956, p. 370). As the theoretical basis for his findings, the author concludes: "The theory of nouns as conceptual pegs, which incorporates imagery as a mediator of recall, is relevant even in learning situations involving associatively unrelated pairs. Thus, with randomly paired nouns varying in abstractness, superior learning would be predicted with concrete nouns as stimuli and abstract nouns as responses rather than the reverse--a prediction that seems not to follow from the associative probability hypothesis or existing data on the effects of meaningfulness. Research is currently underway to test this prediction" (Paivio, 1963, p. 378).

Evidence of reliability, validity, and standardization

It was hypothesized that word pairs in a noun-adjective order (N-A) would be learned more easily than an adjective-noun order (A-N); that pairs with concrete nouns would be learned more easily than compared with abstract nouns; and that the interaction of word-order and noun abstractness was also significant. In the first two instances the hypotheses were verified at a very high level of significance ($p < .001$). In the last case, the hypothesis was also supported as statistically significant level ($p < .05$), "...indicating as predicted that the greater ease of learning the N-A as compared to the A-N order was superior with concrete as opposed to abstract nouns" (Paivio, 1963, p. 374).

The subjects and procedures for the investigation were as follows: "The Ss were 136 children (62 boys and 74 girls) from the fourth and fifth grades of an urban elementary school. Ss subjects were run in nine groups, arranged in size from 13 to 18 members, formed (with one exception) by dividing a classroom of children into two groups which met simultaneously with different Ss in separate classrooms... Each group learned two paired-associate lists, one with pairs in the A-N order, and the other in the N-A order."

Bibliography

Lambert, W. E., and Paivio, A. The influence of noun-adjective order on learning. Canadian Journal of Psychology, 1956, 10, 9-12.

Paivio, A. Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370.

Thorndike, E. L., & Lorge, I. The teacher's workbook of 30,000 words. New York: Bureau of Publications, Teachers College, 1944.

TEST DATA WORKSHEET

Name of Test	Bible Knowledge Test
Author	Parshall, H. W.
Age	Has been used with children with mental ages from 3-6 to 14-3
Area (i.e. language development)	Knowledge of the Bible
Type of measure	Objective test (yes-no type)
Source from which the test may be obtained	See Parshall (1960) in bibliography

Description of Measure

The object of the test is "to provide a simple scale of questions available to Chaplains as a tool in planning their Sunday School program" (Parshall, 1966).

Twenty-five items were prepared from the Bible, nine from the Old Testament and sixteen from the New Testament. They are arranged in the order in which they appear in the Bible rather than in any order of difficulty.

"The test was administered orally on an individual basis..."

Examples of the twenty-five items on the measure are the following:

"1. (Yes) God made the world.
2. (Yes) Moses was the name of the baby found in a basket in a river.
3. (No) Jesus was born in the city of Jerusalem.
4. (No) The Good Samaritan was one of Jesus' disciples.
5. (Yes) Following his death, Jesus was placed in a tomb in the side of a hill."

Evidence of reliability, validity, and standardization

"In order to test the reliability of the answers given by the remaining 48 subjects, of the original 153, the Bible Knowledge Test was readministered one to three months later. The mean M.A. for these 48 subjects was 8 years, 10 months. The reliability coefficient was .92, which indicates satisfactory reliability. Scores on this test correlated .61 with the M.A. of mental defectives and -.12 with the C.A. of the same group.

"This Bible Knowledge Test can be used as an aid in establishing a religious program of training for institutionalized mental defectives. Also, it can be administered to newly admitted patients as an aid to placement in the religious program of the institution. By administering this test and asking a few pertinent questions, the religious worker should be able to obtain a reliable estimate of the patient's knowledge of the Bible. It can be administered quickly and easily, has high 'face validity,' and has high reliability. It is presented to be of use to institutional chaplains and religious workers in developing their training programs for mental defectives" (Parshall, 1960, p. 961).

Bibliography

Parshall, H. W. A Bible knowledge test for institutionalized mental defectives. American Journal of Mental Deficiencies, 1960, 64, 960-962.
Parshall, H. W. Personal Correspondence, November, 1966.

TEST DATA WORKSHEET

Name of Test	Draw-A-Scene Test
Author	Lowenfeld, Viktor
Age	2-17
Area (i.e. language development)	Growth (intellectual, emotional, social, perceptual, physical, aesthetic, creative)
Type of measure	Free drawing
Source from which the test may be obtained	See (2) in bibliography

Description of measure

The author provides a framework that can be used to evaluate the spontaneous drawings of children to get measures of growth in the areas mentioned above in Item 4. He provides several evaluation charts at different age levels. The first evaluation chart is formulated to provide an evaluation method for growth of children aged two to four. This is called the scribbling stage. For each area of growth, there are items which must be checked. For example under the emotional growth area, the first question is, "Does the child enjoy his scribbling?" The scorer or evaluator is then to check either "none", "some", or "much". In the book in which the evaluation charts are given, the author elaborates on the question and gives some suggestions and rather loose criteria that the scorer can use in making his judgment. The next evaluation chart is set up to measure growth at the "preschematic" stage, between the ages of four and seven. Again, there is a limited number of items within each category, the numbers ranging from three (in intellectual growth) to six (in emotional growth). The scoring criteria are very loose, which would lead one to expect the reliability of the measure to be relatively low.

Growth during the "schematic" stage (ages 7-9) and the "gang" stage (ages 9-11) is assessed with the use of evaluation charts which are separate but similar to those for other levels mentioned above.

A generalized evaluation chart is given which, according to Lowenfeld, could be used at any age level.

It is likely that with careful training of scorers a respectable reliability level might be achieved for this evaluation method.

The measure has the advantage of being usable with drawings which have already been made.

Evidence of reliability, validity, and standardization

None

Bibliography

1. Lowenfeld, Viktor. Your Child and His Art. The Macmillan Company, New York, New York, 1955.
2. Lowenfeld, Viktor. Creative and Mental Growth. The Macmillan Company, New York, New York, 1952.
3. West, J. H. Correlates of the Draw-A-Screne. Journal of Clinical Psychology, Volume 16, No. 1, January, 1960, p. 44.

TEST DATA WORKSHEET

Name of Test Hammond-Skipper Pre-School Achievement Rating Scale

Author Hammond, Sarah L., and Skipper, Dora

Age 3-5

Area (i.e. language development) Cognitive and social development

Type of measure Rating scale

Source from which the test may be obtained Skipper, Dora
The Florida State University
Department of Elementary Education
Tallahassee, Florida 32306

Description of Measure

"The writers have developed an achievement rating check list which has been used by teachers with three-, four- and five-year olds. This Achievement Rating check list is not a test. No attempt was made to establish norms for different age groups. It consists of a list of items for assessing achievement which were selected from the literature and have been reviewed by specialists as being appropriate for the early childhood years. The check list is intended to be used in helping the teacher gain a picture of the child's status to be used as a basis for planning and guidance and as a means of assessing a child's progress" (Skipper, 1966). The measure consists of 71 items on which the rater evaluates each child. Examples of items are: "Eats with spoon, puts on coat alone, pedals tricycle, follows simple directions, looks at page from top to bottom, counts rationally 1-5, speaks in sentences (uses sentences of at least five words), and listens and identifies sounds of different instruments. Each of the 71 items is checked in one of four columns headed "Always," "Sometimes," "Never," and "Does Not Apply."

"Directions for scoring: Note items in which the pupil has achieved, the items which need continued practice, and those which need to be introduced. Plan experiences in terms of analysis of achievement items. Ignore those items which do not apply" (Skipper, 1966).

Evidence of reliability, validity, and standardization

None

Bibliography

Skipper, Dora. Personal Communication. May 6, 1966.

TEST DATA WORKSHEET

Name of Test	Meaning of Opposition
Author	Kreezer, G. and Dallenbach, K.M.
Age	5 to 7-1/2
Area (i.e. language development	Concept of opposition
Type of measure	Test, individually administrated
Source from which the test may be obtained	See Kreezer and Dallenbach in bibliography

Description of Measure:
This test is made up of a 25-item word list as follows: White, short, fat, wet, glad, hard, in, slow, weak, open, heavy, up, smooth, little, sleep, shut, strong, quick, out, soft, sad, dry, thin, long, black. The S is given some examples of opposited, gets feedback and limited instruction about opposition, and is then presented with the stimulus words one at a time. The scoring system did not require that antonyms be given. "All that we required was evidence that the Ss, as demonstrated by their replies to the 25 test words, understood the nature of opposition. The answers 'not smooth,' or 'un-smooth,' to the word smooth, for example, were as significant for our purposes as the antonym 'rough.'" (Kreezer and Dallenbach, 1929, p. 433). Thus, the test is designed and scored to measure understanding of the concept or relation of opposition, rather than as an opposites vocabulary test.

Evidence of reliability, validity, and standardization:
Norms are provided for five age groups, from 5-0 to 7-6, at half-year intervals, with N=20 for each group. In general, the number of correct responses shows progression with age. Robinowitz (1956, p. 29) also found significant relationships between age of child and ability to learn the relation of opposition.

Bibliography:

Kreezer, G. and Dallenbach, K.M. Learning the relation of oppositions. Amer. J. Psychol., 1929, 41, 432-441.

Robinowitz, R. Learning the relation of opposition as related to scores on the Wechsler intelligence scale for children. J. genet. Psychol., 1965, 88, 25-30

TEST DATA WORKSHEET

Name of Test: Number Facility Scale--Lesser, et al.

Author: Lesser, Fifer, and Clark

Age: 6 years, 2 months through 7 years, 5 months

Area (i.e. language development): Intelligence

Type of measure: Number facility

Source from which the test may be obtained: U.S. Office of Education (See Reference)

Description of Measure:
The authors' purpose ". . . is to measure numerical concepts prior to formal training in mechanics of computation or in the operational symbols." The following account describes the manner in which they attempted to achieve their aim as well as the nature of the test:

"The resultant numerical test consists simply of two pictures: one for the Enumeration, Addition and Subtraction subtests, and one for the Multiplication and Division subtests. The former is a street scene showing houses, trees, autos, children, and other objects in quantity. The latter picture is of a fruit and vegetable stand with some children in the scene. The test items require the subject to enumerate, add, subtract, multiply and divide the persons or objects in the scene. On the enumeration subtests the subject is permitted to count with his finger actually touching the picture if necessary, as this test was introduced into the scale in order to determine the subjects' ability to count. On all other subtests the subjects were required to compute without touching the pictures.

It was recognized that there was a risk of introducing unwanted verbal variance in presenting the tests in the manner. However, the names of objects to be manipulated, the phrasing of the questions and the directions given were all put in the simplest possible language. Various versions of the specific methods applicable to each of the subtests were tested during item tryouts and the wording producing the least confusion was adopted. Furthermore, the examiner was prepared to give the entire test in the native language of the child, if necessary" (Lesser et al., 1964, pp. 77-78).

"The directions for giving these tests appear in Appendix A (Lesser et al., 1964, pp. A-11, A-15), and the specific enumerations and answers required are indicated in the test blank included as Appendix B" (Lesser et al., 1964, p. B-6).

As in the previous scales, the authors' definition of the skill involved in the test and its significance for intellectual ability may be cited: "This is defined as skill in enumeration, and in memory and use of the fundamental combinations in addition, subtraction, multiplication, and division. It is of great importance in arithmetic in elementary schools and mathematics in secondary schools" (Lesser et al., 1964, p. 68).

Evidence of reliability, validity, and useability:
The reliability coefficient for the entire sample was .96. "All ethnic groups show consistently high reliabilities for Number (.94-.96). . ." (Lesser et al., 1964, p. 96).

The range of intercorrelation for this scale with other scales was .28 (Lower-Class) to .58 (Puerto Ricans and Negro) on the Verbal Scale; .52 (Middle-Class) to .74 (Puerto Rican) on the Reasoning Scale, and .40 (Jewish) to .54 (Negroes) on the Space Test (Lesser et al., 1964, pp. 86-101).

Bibliography of test usage:

1. Lesser, Gerald S., Fifer, Gordon, and Clark, Donald H. Mental Abilities of Children in Different Social and Cultural Groups. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, Comparative Research Project No. 1635, 1964.
2. Davis, F.D., Lesser, G.S., and French, Elizabeth G., et al. "Identification and Classroom Behavior of Gifted Elementary-School Children." Cooperative Research Monographs, 1960, No. 2, 19032.
3. In "The Gifted Student," OE-35016, Monograph No. 2, U.S. Office of Education, Department of Health, Education, and Welfare.

TEST DATA WORKSHEET

Name of Test	Probst Test
Author	Probst, Cathryn A.
Age	Preschool and primary school
Area (i.e. language development	General information
Type of measure	Individually administrated, question-answer
Source from which the test may be obtained	See Probst, 1931, and Rosen (undated)

Description of Measure:
The items on this measure are similar to the information items on the Weschler Scales. There are two forms of the test, each consisting of 66 items. These items are divided into eleven areas or topics and six items measuring knowledge about the topic. The topics or areas measured are as follows: local points of interest; time and number; current topics; natural phenomena; literature and music; animals, birds, and insects; plants & flowers; occupations & industries; household arts; simple mechanics; and games & amusements. Topics 1 and 3, Local Points of Interest and Current Topics, respectivel obviously need to be modified for both the time and locale where the test is to be used For example, Rosen's adaptation of the items in Topic 1 for use in Milwaukee contains the following items: Tell me the name of a Milwaukee newspaper. What is the name of a lake in Milwaukee? What is the Mississippi? What is Wisconsin Avenue? What is the North Shore? What is WOKY? Similarly, in Topic 3, Current Topics, some modifications must be made according to the time. Rosen has modified the Probst Test in this way for his Milwaukee version: 1. Who is Freddy? 2. Who is Khrushchev? 3. Who was the first president? 4. Who is Mickey Mantle? 5. Who is Beaver? 6. Who is Dwight Eisenhower? Examples of other items in the test are the following: Topic V (Literature and Music), "What colors are the keys on the piano? What was Cinderella's coach or carriage made of?" Topic III (Occupations & Industries) shows the following items: "1. Who makes money by cutting hair? 2. What is a shoe made of?" The items are scored right or wrong, and the two forms of the test may be combined to increase reliability. This test, like any other information measuring test, may be considered in large part a measure of opportunity to learn. The test is administrated individually, in either one or two sessions. Probst administrated it in two sessions and reported a mean administration time of 17 minutes for Form A and 19 minutes for Form B. Templin (1958) administrated it in a single session and found the mean time to be approximately 30 minutes.

Evidence of reliability, validity, and standardization:
Probst (1931) reports an uncorrected equivalent forms reliability of .94 and a corrected reliability of .97. She found that corrected reliabilities for the eleven topics or areas range from .45 to .89 with a median of .76. The equivalent forms data reported above were based on 100 kindergarten children. "A correlation of .81 was obtained between the equivalent forms" (Rosen, undated). Rosen has developed an

"abridged information test," a similar version of the Probst's Test, made up of a total of 44 items. Templin (1958) reports an equivalents form corrected reliability of .94.

Bibliography:

Probst, Cathryn A. A general information test for kindergarten children. Child Develpm., 1931, 2, 81-95.

Rosen, S. Information Range and Interest among Kindergarten Children. Marquette University, undated. (Unpublished paper).

Templin, Mildred C. General information of kindergarten children: a comparison with the Probst study after 26 years. Child Develpm., 1958, 29, 1, 87-96.

TEST DATA WORKSHEET

Name of Test Objective Language Scale

Author Williams, W. G.

Age 5-12

Area (i.e. language development) Language development

Type of measure Observation of behavior and individual test items

Source from which the test may be obtained Definition of the language scale items and other explanatory data may be obtained by writing the author at Purdue University, Speech and Hearing Clinic

Description of Measure

This measure is described as "a language scale listing items of language development at successive age levels" (Williams, 1960, p. 30) and is a "modification of one developed earlier by M. J. Mecham."

"Most of the items were taken from Doll's Vineland Social Maturity Scale, Gesell's Developmental Schedules, Terman and Merrill's revision of the Binet Scale, Poole's Maturation Scale of Articulation, and the works of McCarthy."

All of the items are listed in the Williams article, and each item is designated by category: Reading, Writing, Speaking, and Listening.

Evidence of reliability, validity, and standardization

"1. Scores on the language scale and Metropolitan results were tested by means of rank-order correlation. A significant relationship was found, except in three grades.

2. The language scale scores were analyzed by simple analysis of variance, F. A significant difference was found in successive grade levels.

3. The scores of speech-delayed children were correlated with the evaluations of speech clinicians. A significant relationship was found between these two measures.

4. The language scale was tested for reliability by use of the split-half technique. A significant correlation of reliability was found.

5. The scores obtained by teachers and clinician, all using the language scale, were tested by t-ration and rho correlation. A significance of difference was found with t; a significant relationship was found with rho for most grades.

6. The teachers' subjective evaluation and use of the language scale were correlated by using rho and the sign test. No significant relationship was found with most grades" (Williams, 1960, pp. 30-33).

Bibliography

Williams, W. G. The adequacy and usefulness of an objective language scale when administered to elementary school children. Journal of Educational Research, 1960, 54, 30-33.

TEST DATA WORKSHEET

Name of Test: Reasoning Scale--Lesser et al.

Author: Lesser, Fifer, and Clark

Age: 6 years, 2 months through 7 years, 5 months

Area (i.e. language development): Intelligence

Type of measure: Scales on Verbal Ability, Reasoning ability, Numerical ability, and Space ability.

Source from which the test may be obtained: U.S. Office of Education (See Reference)

Description of Measure:
As indicated previously, the present scales are modifications and adaptations of the Hunter Aptitude Test. "The Reasoning subtests used in this study are more similar in content and formats to those of the earlier study than any of the other subtests. . . The major changes in the items within the subtests consisted of revising the drawings for the Picture Analogies and Picture Arrangement subtests Drawings were made with a view to minimizing the picture-book quality previously present. A further modification was the addition of more abstract drawings; that is, a number of geometrical line drawings were utilized to measure the ability to recognize various relationships outside the usual context of the picture book or story book. All items included in the Reasoning subtests are such that no naming of objects is required. If the subject can perceive the relationship involved, the correct answer can be indicated without naming any object or, indeed, without knowing the names of any of the objects in the drawings. Verbal skills are thus minimized.

(1) Picture Analogies: Each of the Picture Analogy items consists of two large cards. The first card contains three drawings, two of which are paired The paired objects are similar and are related to each other in some way; e.g. small, large, young-old, narrow-wide, upright-turned. The second card consists of drawings of four objects similar to the third object on the first card One of the four objects on the second card bears the same relationship to the third object on the first card as do the paired objects. The subject is required to select the analogous pair.

(2) Picture Arrangement: This subtest consists of a series of picture stories and of abstract object sequences. These consist of from three to five drawings, each on a separate card. Each item is presented in a predetermined mixed order and the subject is asked to arrange the cards so the pictures make a good story or are in a good order.

(3) Jump Peg: This subtest consists of a specially-designed board with large holes in it (similar to a Chinese checkers board) and several large wooden pegs. The

subject is presented with a set of increasingly more difficult arrangements in the board and instructed to jump and eliminate pegs until only one peg remains and that one in the center hole of the board" (Lesser, et al., 1964, pp. 74-75).

The authors caution ". . . care be taken in each case to be certain that the subjects understand the procedure before scored items are presented" (Lesser et al, 1964, p. 77). To this end, they provide detailed directions for administering and scoring the subtests (Lesser et al., 1964, pp. A-5, A-10).

As in the Scale for Verbal Ability, the authors' definition of Reasoning and its significance for intellectual ability might be noted: "This involves the ability to formulate concepts, to weave together ideas and concepts, and to draw conclusions and inferences from them. It is, almost by definition, the central element of aptitude for intellectual activities and therefore is of primary importance in all academic fields and most vocations" (Lesser et al., 1964, p. 68).

Evidence of reliability, validity, and usability:
Reliabilities (corrected for attenuation) are as follows: The r for the entire sample (N=320) was .92. The range was from .79-.94. Jewish (.84) and Negro children (.85) show somewhat lower reliabilities on the Reasoning scale than they do on other scales. The reliabilities for these scales are lowest for both Middle-Class Negro and Jewish children (r=.79). Reliabilities for lower-class Negro children (r=.94) and lower-class Jewish children (r=.88) are distinctly superior to their middle class counterparts. Finally, reliability for boys (r=.92) surpasses that for girls (r=.84).

"Reasoning (Median r of .62) correlates most strongly with the other scales . . ." (Lesser et al., 1964, p. 89).

Bibliography of test usage:

1. Lesser, Gerald S., Fifer, Gordon, and Clark, Donald H. Mental Abilities of Children in Different Social and Cultural Groups. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, Comparative Research Project No. 1635, 1964.

2. Davis, F.D., Lesser, G.S., and French, Elizabeth G., et al "Identification and Classroom Behavior of Gifted Elementary-School Children." Cooperative Research Monographs, 1960 No. 2, 19032.

3. In "The Gifted Student," OE-35016, Monograph No. 2, U.S. Office of Education, Department of Health, Education, and Welfare.

TEST DATA WORKSHEET

Name of Test — Riggs-Rain Classification System

Author — Margaret M. Riggs and Margaret Z. Rain

Age — Almost any age from birth to maturity

Area (i.e. language development) — Diagnosis of type of mental retardation

Type of measure — Check scale

Source from which the test may be obtained — See References.

Description of Measure:
The Riggs-Rain System is a formal method of diagnosing mental retardation according to the degree of information available in the records. "Six major categories are to be used: I Familial, II Organic, III Unexplained, IV Mixed, V Monoloig, VI Non-Classifiable. Each category has several subdivisions which represent levels of certainty in terms of the amount of information available" (Riggs and Rain, 1952, p. 75).

"This classifications system involves the interaction of three kinds of information:

a) Positive evidence indicating familial or organic pathology.
b) Less positive evidence, sufficient to raise a slight but non decisive suspicion of familial or organic pathology.
c) Specific non-diagnostic information concerning family, pregnancy, birth, development, and neurological state, the presence of which suggests that if there had been pathology it would have been remembered and reported by the informant."

Specifically, the Classification System enables one to classify the etiology of a case as familial or organic according to "definite positive evidence" or "a slight suspicion" for each category. Examples of each category within each level of evidence follow:

"I Criteria for Classifying Cases as Familial

A. Definite evidence:

1. At least one parent or sibling (not half sib, not twin, not putative parent) is diagnosed as 'backward,' 'dull,' 'retarded,' 'mentally deficient,' of borderline intelligence or less, on the basis of an intelligence test, or by an individual or institution known to base diagnoses only on intelligence tests; such deficiency not accompanied by a possible organic cause or symptomatology indicative of such a cause, as these are defined in this paper."

"II Criteria for Classifying Cases as Organic

A. Definite evidence:

1. Diagnosis by a neurologist or endocrinologist on the basis of objective examination that a recognized organic syndrome is responsible for the

child's retardation, e.g. cerebral palsy, megalencephaly, Froelichs, hydrocephalus. If equally competent authorities in the same field have given conflicting opinions disregard the evidence entirely.

2. One or more of the following are in the case history:
   a) Severe falls or injuries to the mother during pregnancy.
   b) German measles (rubella) in first 3 months of pregnancy.
   c) Mother had severe anaemia during pregnancy."

Two helpful aids in the use of this System have also been provided: a complete guide to indicate levels of certainty in the diagnosis and a data sheet for recording all the necessary information that will enable one to make a diagnosis.

According to Riggs and Cassell (1952), the system offers several advantages.

"1. It is one of the advantages of the system that rules for classification remain fixed while the specific criteria for Familial and Organic Etiology may be and should be expanded or changed to fit new facts...."

2. ". . . inexperienced people might easily train each other in the use of this system."

3. "The reliability of this classification system depends so much upon the accuracy of the user, and so little on professional judgment, that a skilled clerk or registrar might do all the abstracts in case histories and most classifying on a routine basis."

4. With added practice, a speed of abstract in information from cases increases and the number of errors decreases. For example, in the original study with the use of the system, ". . . the amount of time spent classifying a case decreased from about 45 minutes to about 25, but the average number of errors for the first 25 and 100 cases was 5.2, and for the last 25 it was still 3.5."

"From the standpoint of research, this system holds certain unique advantages. We will be able to state how 'pure' a certain sample is with respect to etiology, or specify in advance the certainty level we are prepared to accept. For example, we might wish to survey the etiology of all children showing artistic skill; while low certainty of classification could not be avoided in some cases, we can at least assign a tentative label to each. For other problems we might want to compare the extent to which Not-Classifiable or Mixed cases resemble the 'purer' examples of a certain etiology. At times we may want the most sharply contrasted groups we can secure and would limit our sample to cases at the highest certainty level. All too often contradictory results are obtained for the same problem because the experimenters are not using the same type of case. We believe that the reliability and objectivity of the system we have just presented will help those who use it to exchange ideas concerning the mentally retarded with a minimum of confusion" (Rigge and Cassell, 1952).

<u>Evidence of reliability, validity, and useability:</u>
Rigge and Cassell (1952) seem to describe a form of content validity in support of the classification system. "In specifying the causes and symptoms which we are willing to accept as evidence of organic damage, we have included only those: (a) which we actually have found a series of about 300 cases, and (b) which are

also well recognized as being significant for the etiology and diagnosis of damage to the central nervous system."

After a practice and teaching session for rating cases, four judges classified the same 100 cases. "All but twenty-five of the final cases and practice cases were under age 20. All were taken in alphabetical order from the clinical files or had been encountered in recent departmental work. They formed a representative sample of the young resident population of the Training School." Several indices, then, furnish evidence for reliability.

1. "Analysis by Chi-square show that the four judges did not differ significantly in the number of cases placed in each subdivision. . .

2. Two or more judges agreed on the same classification for all 100 cases. At least three of the four agree in 85 cases. Agreement was unanimous for 61 cases. . . The over-all difference between chance frequency and obtained frequency is significant far beyond the .001 level of confidence, even allowing for . . . low theoretical frequencies."

3. The authors accepted agreement "of at <u>least</u> three or four judges as the criteria for reliability . . ." in the classification of any given case. On this basis, agreement was found in 85% of the 100 cases. "This means that we may expect between 76 and 91 percent agreement of this sort at or beyond the .05 level of confidence, and between 74 and 93 percent at or beyond the .01 level of confidence . . ." (Riggs and Cassell, 1962).

<u>Bibliography</u>:

Riggs, Margaret M., and Rain, Margaret E. A classification system for the mentally retarded. Part I: description. <u>Train. Sch. Bull.</u>, 1952, 49, 75-84.

Riggs, Margaret M., and Cassell, Margaret E. A Classification system for the mentally retarded. Part II reliability. Train. Sch. Bull., 1952, 49, 151-168.

TEST DATA WORKSHEET

Name of test	Rutgers Drawing Test
Author	Anna Spiesman Starr
Age	Four to seven years
Area (i.e. language development)	Intelligence
Type of measure	Paper and pencil performance test
Source from which the test may be obtained	See Starr (1952) in bibliography

Description of Measure

The Rutgers Drawing Test consists of 14 geometric figures". . . drawn to scale, and arranged in order of difficulty with ample space beneath each figure for the child's copy.... The following figures make up the series: cross, circle, square, square with extended sides, oblique cross, Indian swastika, triangle, baseball, block cross, straight line star, diamond, double rectangle, interlaced star, Maltese cross" (Starr, 1952, p. 47).

As part of the procedure, the child is expected to copy a figure directly beneath the geometric form on a single sheet of paper. Details on administering the test and the materials necessary are provided.

Responses are scored as 2 for success, 1 as partial credit, or 0 for failure.

As an aid to scoring, complete and extensive scoring samples are provided for all 14 figures (Starr, 1952, pp. 56-64). By means of a table for the derivation of scoring norms, raw scores may then be converted to an "equivolent drawing age in months."

Finally, "the Rutgers Drawing Test is given ordinarily as an individual test although it may be used with small groups of children already familiar with each other." No time limit is necessary but Starr has concluded that ". . . experience shows that about five minutes are required for the average child" (Starr, 1952, p. 47). Further, she has emphasized that the test should be limited to children between the four to seven year levels. "Failure above that level may be highly significant" (Starr, 1952, p. 50).

"Although there has been no standardization of the test on children of seven years and over, our practice is to give a drawing age of 84 months for a score of 24 and to add two months 'drawing age' for each additional score point. Thus, a maximum score of 28 is a drawing age gives a drawing age of 92 months" (Starr, 1952, p. 51).

Evidence of reliability, validity, and standardization

"The Rutgers Drawing Test has had a long period of incubation. In 1931, the first results were reported as tentative norms of one hundred preschool children copying simple geometric figures (10). In 1937, Stevens submitted a continuation study of 555 children (age range 40 thru 6.9 years) copying a series of fourteen figures (11). Since then further studies have been carried out with a total of some 3,000 protocols available. From time to time, various smaller studies have been made exploring how well mentally deficient adults and children score on drawing tests. Another such peripheral study was made of a group of Negro nursery school children, resulting in no indication of racial differences. More recently, a rescoring was done of all the normal cases in order to extend the spread of scoring values in the upper age range beyond the entrance to school" (Starr, 1952, p. 45).

"A further study was made of 277 normal children within the age range of 3 to 6, who had had individual examinations at the Rutgers Clinic. A comparison of their MA scores on Binet and the MA of the Drawing Test showed a product-moment correlation of .72 P. E., .019.... An additional study was made at Rutgers of 57 cases of mentally deficient children under 12 years of age, with no I.Q. over 60. The correlation between their Binet and Drawing scores was .61, P. E., .05. This further supports the validity of the findings with the 4 to 6 year group" (Starr, 1952, pp. 49-50).

To the best of the reviewer's knowledge, no data on reliability for the scale are available.

Bibliography

Starr, Anna S. Drawing test for pre-school children. Unpublished study. Preliminary report read before Association of Consulting Psychologists. New York, 1931.

Starr, Anna S. The Rutgers drawing test. The Training School Bulletin, 1952, 49, 45-64.

Stevens, Margaret C. A drawing test for the pre-school child. Unpublished Master's thesis. Rutgers University, 1937.

TEST DATA WORKSHEET

Name of Test	Scales of Mental Ability of Children in Different Social and Cultural Groups
Author	Lesser, Fifer, and Clark
Age	6 years 2 months, through 7 years 5 months
Area (i.e. language development)	Intelligence
Type of Measure	Scales on Verbal Ability, Reasoning Ability, Numerical Ability, and Space Ability
Source from Which the Test May Be Obtained	U. S. Office of Education (see ref.)

Description of Measure:
Possibly important features of these scales include the sample, the original stimulus for the scales, unique features in the standardization, the actual sub-tests involved in the scales, and the purpose for the development of the scales.

Sample selection: "Subjects were 320 first grade children, age 6 years 2 months, through 7years 5 months, 80 from each of four cultural groups (Jewish, Negro, Chinese, and Puerto Rican). Each cultural group was subdivided into 40 children of middle class and 40 of lower class status. The 40 children in each social class category within each cultural group were further subdivided into 20 girls and 20 boys. Thus, a 4 x 2 x 2 analysis of variance design was used (4 Cultural Groups x 2 Social Classes x 2 Sex Groups), each of the 16 cells containing 20 S's. All S's were drawn from public schools in New York City and its surrounding environs" (Lesser et al., 1964, p. 37).

Original basis of scales: The present scales were based on modifications of the Hunter Aptitude Scales for Gifted Children (Lesser, et al., 1964, p. 61) which were utilized with children ranking in age from 4 years to 5 1/2 years (Davis et al., 1960).

Unique features and standardizations: Two rather unique features were included in the standardization of the scales. First, the test ". . . could be administered in English or in the child's primary language or any combination of the two...." (Lesser et al., 1964, p. 65). This unique method of administration was made possible by two interrelated factors:

1. ". . . psychological testers who spoke one of the primary languages of the cultural groups (Spanish, Yiddish, several Chinese dialects) administered the tests so that instructions and test questions could be given in English, the primary language of the child's cultural group, or (more often) in the most effective combination of the two languages for the particular child" (Lesser et al., 1964, p. 26).
2. "Only words were used with suitable equivalents in Spanish, Chinese and Yiddish, indicating the same verbal concepts at approximately the same difficulty level for children of ages six to seven and one half years" (Lesser et al., 1964, p. 71). Secondly, extreme care was taken by the investigators to control

numerous variables affecting test performance. These variables related to the background and experience of the examinee, his personality and motivation, immediate situation, and the test demands (Lesser, et al., 1964, pp. 22-36). Not all of these 18 variables related to these four broad areas were adequately controlled, though most of them were. In any case, the authors described explicitly the rationale for the control, the method for this control, and the adequacy and/or inadequacy of their controls achieved.

Subtests of the scales: "The final form of the scales used in this study included the following subtests and numbers of items...." (Lesser et al., 1964, p. 6).

Scale	Subtests	Number of Items	
Verbal	Picture Vocabulary	30	
	Word Vocabulary	30	
			60
Reasoning	Picture Analogies	18	
	Picture Arrangement	16	
	Jump Peg	12	
			46
Numerical	Enumeration	6	
	Addition	10	
	Subtraction	10	
	Multiplication	10	
	Division	10	
			46
Space	Object Completion	16	
	Estimating Path	12	
	Jigsaw Puzzles	16	
	Perspective	10	
			54

Purpose of scales: The specific purpose of the study was ". . . to examine the patterns among various mental abilities in first-grade children from different social class and cultural backgrounds. Patterns among four mental ability areas (space conceptualization, verbal ability, number facility, and reasoning) are studied within four cultural groups in New York City (Puerto Rican, Negro, Chinese, and Jewish), with each culture group divided into middle class and lower class groups" (Lesser et al., 1964, p. 1). The general purpose of the study was ". . . to extend the general demonstration that differential mental abilities exist in young children by examining the variations in the patterns of these diverse abilities associated with variations in social class and cultural conditions. While it has been established that groups factor beyond a general ability factor exist in young children, little has yet been discovered about the differences in patterns of intellectual expression, the relative strengths related to different social class and cultural influences" (Lesser et al., 1964, pp. 1-2).

Evidence of Reliability, Validity, and Useability:
Reliability data is extensive and should be consulted for full details (Lesser et al., 1964, pp. 86-101). "Reliabilities (corrected for attenuation) and inter

correlations are presented for the entire group . . ., for each ethnic group. . ., each social class group. . ., and each sex group. . . In addition, each ethnic group is divided into its middle and lower social-class components and the reliabilities and intercorrelation described....

The reliabilities based upon the entire sample . . . are satisfactory except that the reliability of the Space test (.85) is somewhat lower than desired. Reliabilities do not vary greatly among ethnic groups.... However, the Jewish children (.87) show a slightly lower reliability on the Verbal test than the other groups, and both Jewish (.84) and Negro children (.85) show somewhat lower reliabilities on the Reasoning scale. All ethnic groups show consistently higher reliabilities for Number (.94-.96) and moderate reliabilities for Space (.80-.85).

Intercorrelations indicate ". . . a rather disappointingly high set of interrelationships." They depended in large part on homogeneity of the sample. Thus, ". . . when each social-class group (N=160) contains all ehtnic groups...." the median reliabilities for attenuation (r=.91) and intercorrelations (r=.53) were both high. In contrast, ". . . when the samples were reduced to more homogeneous groups by dividing each ethnic group into its separate social-class components. . . the reliabilities were reduced only negligibly (from .91-.89) but the intercorrelations dropped considerably (from 153 to .40). These data of test degree to which the current sampling of children from extremely diverse groups have effected the intercorrelations among the scales" (Lesser, et al., 1964, p. 88). Since the test consisted of power scales exclusively the statistics for attenuation probably were not spurious. "Under no circumstances was the child's test performance inhibited due to a lack of time" (Lesser, et al., 1964, p. 34). The authors suggest that the plan of reliabilities and intercorrelations on the scales gives some support to their validity. First, ". . . strong intercorrelations in the present study are accompanied by hgih reliabilities for the scales." Secondly, they proposed that if the predicted differences in patterns of mental abilities among these social class and ethnic groups emerged ".... despite the limited differentiation among abilities, such findings may then testify even more convincingly to the power of the social c class and cultural influences than if greater differentiation of abilities existed." Significant differences in patterns of mental abilities were found among the four ethnic groups (F ratio of 17.32, p .001). But these patterns of differences did not hold with respect to social class groups or the interaction of social class and ethnicity (Lesser et al., 1964, pp. 122-130).

In addition to the differences and patterns of mental ability, the analyses of variance indicated that highly statistically significant differences for level of performance existed with respect to differences between (1) social class groups, (2) differences among four ethnic groups, and (3) significant interactions between social-class ethnicity (Lesser et al., 1964, pp. 102-104).

Bibliography of Test Usage:

Lesser, Gerald S., Fifer, Gordon, and Clark, Donald H. Mental Abilities of Children in Different Social and Cultural Groups. Washington, D. C.: Office of Education, U. S. Department of Health, Education, and Welfare, Comparative Research Project No. 1635, 1964.

Davis, F. D., Lesser, G. S., and French, Elizabeth G., et al. "Identification and Classroom Behavior of Gifted Elementary School Children. Cooperative Research Monographs, 1960, No. 2, 19032.

In "The Gifted Student", OE-35016, Monograph No. 2, U.S. Office of Education, Department of Health, Education, and Welfare.

TEST DATA WORKSHEET

Name of Test	Sequentially Scaled Achievement Test
Author	Richard Cox, and Glenn T. Graham
Age	Kindergarten, first, and second grade
Area (i.e. language development)	Arithmecic
Type of measure	Sequential Achievement Test
Source from which the test may be obtained	Learning Research and Development Center, University of Pittsburgh

Description of Measure:
The development of the test was based on the following premise: "Items for achievement testing . . . should be designed to indicate whether or not the required behaviors had been mastered . . ." This premise, the authors claim, applies to ". . . performance criteria for successful completion of some specific year of work which are identical for all students" (Cox and Graham, 1966, p. 1). Under these conditions, a specific score would indicate whether a student has mastered specified behaviors. In extending the Coefficient of Reproducibility to achievement testing, moreover, the authors reasoned as follows: "If the behavior to be tested could be arranged in a sequential order, and the tests were scalable, a student who obtained a score of 5 would have answered items 1, 2, 3, 4, 5, 6, and 7, and did not answer any items beyond 7. Knowing the behaviors these items represent, the score on the test indicates to the teacher, guidance counselor, or researcher those behaviors the student has mastered and those behaviors he has yet to master. The present study is and attempt to develop such a test" (Cox and Graham, 1961, pp. 2-3).

The normative group for a validation of the test was not specified in terms of the size. The test was, however, administrated to the "kindergarten, first, and second grade." (Cox and Graham, 1966, p. 5).

Evidence of reliability, validity, and useability:
The authors use the Scalogram Analysis by Guttmann in the development of the achievement test. "The analysis yields a "Coefficient of Reproducibility" which indicates how well an individual's response pattern can be reproduced by knowing his total score. The value of .90 was arbitrarily established as an acceptable lower limit" (Cox and Graham, 1966, p. 1). The authors contend that the Coefficient of Scalability when used with the reproducibility coefficient ". . . further contributes evidence of scalability." The Coefficient of Scalability ". . . determines the degree to which the individual's performances can be reproduced from knowledge of the marginal totals. The coefficient prevents one from spuriously attributing high scalability from a sample of many extreme items and/or individuals. A coefficient of .792" (Cox and Graham, 1961, p. 5). No statistics on reliability coefficients are given. "Since this test is still on an experimental there is no data available concerning reliability . . ." (Cox, 1966).

Bibliography of test usage:

Cox, Richard C., and Graham, Glenn T. "The Development Of A Sequentially Scaled Achievement Test." Presentation at the American Educational Research Association, Annual Convention. Chicago, Illinois. February, 1966.

Cox, Richard. Personal Correspondence, March 3, 1966.

TEST DATA WORKSHEET

Name of Test: Space Conceptualization Scale--Lesser et al.

Author: Lesser, Fifer, and Clark

Age: 6 years, 2 months through 7 years, 5 months

Area (i.e. language development): Intelligence

Type of measure: Space conceptualization

Source from which the test may be obtained: U.S: Office of Education (See Reference)

Description of Measure:
"The directions for giving these tests appear in the Appendix A [Lesser et al., 1964, pp A-16, A-22] and the specific enumerations and answers required are indicated in the test blank included as Appendix B "[Lesser et al., 1964, p. B-6]. The Space Scale includes the following subtests:

(1) "Estimating Path: This subtest consists of twelve drawings of airplanes. Five drawings show a single plane entering a cloud and seven show two planes on intersecting routes. The drawings indicate the direction of flight of the planes. In the first set, the subject is asked to select from four choices the point the plane will come out of the cloud if it continues on the same flight path. In the latter set, the subject selects from four options the point the flight paths of the two planes will cross.

The Estimating Path subtest was retained from the earlier battery although it tends to have a somewhat limited three-dimensional quality. As it was reasonably easy for the very gifted children, it is seemed suitable for measuring normal children. Furthermore, being a completely novel test, it is unlikely that any current children's game would bias it.

(2) Jigsaw Puzzles: Each item in this subtest consists of two identical squares, triangles or circles made of a colorful heavy plastic. One of each shape is cut into three or more pieces. The subject is required to fit the pieces together to match the uncut shape which is left in the view but out of the reach of the subject.

The Jigsaw Puzzles test was retained because it was the most successful subtest in the Space Battery utilized with gifted children. For the purposes of this study, the only change made was to utilize the easier items from the previous study It is likely that this test yields a valid and reliable spatial-visualization factor, and is probably one of the better subtests. This proved to be the case in the study of gifted children, and the estimates of difficulty levels suggest that an appropriate set of items has been included in the current battery.

(3) Object Completion: This subtest consists of sixteen incomplete pictures of familiar objects. The subject is requested to identify each object. The list of objects appears in Appendix B.

The Object Completion subtest is less suitable for measuring spatial conceptualization than is desirable because it requires verbal identification as well as visual

perception. The attempt was made to minimize the verbal component by selecting extremely simple and familiar objects (e.g., dog, spoon, and car) In addition, this subtest was scored so that even if a child could not name the object precisely, his response was marked correct if it indicated a suitable identification of the object.

(4) Perspective: This subtest consists of three large drawings, two of street scenes and one of a children's playground. Each scene contains three or four persons. The subject is asked to identify from four or five options the part of the scene a particular person would be able to see from that person's position in the scene

The Perspective subtest provides good spatial measure if one can assume a minimum reasoning ability. To reduce the contribution of individual differences in reasoning ability, the items were designed to present perspective in its simplest form, directions were simplified as much as possible, and sample items were used" (Lesser et al., 1964, pp. 78-80).

Again, the authors' definition of Space Conceptualization and its significance and terms of intellectual ability should be noted: "This refers to a cluster of skills related to judging spatial relationships and sizes of objects, and in visualizing their movements in space. It is involved in geometry, trigonometry, mechanics, and drafting, in elementary school activities such as practical arts and drawings, and in occupation such as mechanics, engineering, and architecture" (Lesser et al., 1964, pp. 68-69).

Evidence of reliability, validity, and useability:
The reliability coefficient (corrected for attenuation) for the entire sample was .85. The reliabilities (corrected for attenuation) range from .63 to 86. The reliability coefficients was lowest for Middle-Class Negro Children (r=.63) They were also relatively low for Middle and Lower-Class Jewish Children (r=.74). The reliability coefficient for lower-class Negroes on this scale was appreciably higher (r=.81) than that for Middle-Class Negro Children (r=.63). "All ethnic groups showed consistently high reliabilities for Number (.94-.96) and moderate reliabilities for Space (.80-.85)." The data suggests, moreover, that the Space Test, along with the Verbal Scale, correlates least strongly with other scales (median r of .54). "The somewhat intercorrelations of the Space Scale. . . . are perhaps attributable to its lower reliability" (Lesser et al., 1964, p. 89).

Bibliography of test usage:

1. Lesser, Gerald S., Fifer, Gordon, and Clark, Donald H. Mental Abilities of Children in Different Social and Cultural Groups. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, Comparative Research Project No. 1635, 1964.
2. Davis, F.D., Lesser, G.S., and French Elizabeth G., et al., "Identification and Classroom Behavior of Gifted Elementary-School Children." Cooperative Research Monographs, 1960, No. 2, 19032.
3. In "The Gifted Student," OE-35016, Monograph No. 2, U.S. Office of Education, Department of Health, Education, and Welfare.

TEST DATA WORKSHEET

Name of Test	Test of Context Clues
Author	Dunn, L. M. C.
Age	Elementary school
Area (i.e. language development)	Ability to use context clues in reading
Type of measure	Test-group
Source from which the test may be obtained	See (1) in bibliography

Description of Measure

This is the type of short measure of the ability to use context in reading which is found in many of the standardized, published reading diagnostic scales.

Evidence of reliability, validity, and standardization

Dunn (1) used this measure to compare retarded and normal children of similar mental ages. He found the Mean Raw score for the retarded group to be 6.85, whereas that of the normal group was 12.53. He found this difference significant at the 1% level. "Measures of reliability and validity were beyond the scope of present investigation."

Bibliography

Dunn, L. M. C. A comparative study of mentally retarded and mentally normal boys of the same mental age on some aspects of the reading process. Unpublished docteral dissertation, University of Illinois, 1953.

TEST DATA WORKSHEET

Name of Test	Verbal Scale—Lesser et al
Author	Lesser, Fifer, and Clark
Age	6 years, 2 months, through 7 years, 5 months
Area (i.e. language development)	Intelligence
Type of measure	Scales on Verbal Ability, Reasoning Ability, Numerical Ability, and Space Ability
Source from which the test may be obtained	U.S. Office of Education (See Reference)

Description of Measure:
"The Verbal Scale consist of 30 Picture Vocabulary items and 30 Word Vocabulary items. The latter subtest was administrated in two parallel forms of 15 words each. The two forms were administrated at separate testing sessions. This technique was employed with the Word Vocabulary partly because of the length of time required to administer this test entirely to obtain a more rigorous of the reliability of the test" (Lesser et al., 1964, p. 69).

In scoring the items on this scale, ". . . the examiners entered on the answer blank the entire response given by each subject whenever it was anything other than an obviously correct one." Items were either scored right (+1) or wrong (0).

"Three test influences were statistically controlled in the analysis of covariance design of the study: effort and persistence. . . persuasibility or responsiveness to the examiner. . . and age of the subject...."

1. "A simple, rote cancellation task was used to provide a rough estimate of effort and persistence."

2. ". . . To measure each child's responsiveness to the examiner, the 'experimenter persuasibility' test was administrated.... The attempt was made "to access the child's general desire to seek the approval of the examiner in order to statistically control this variable in determining the impact of social class and cultural conditions upon mental abilities."

3. "Interest the child to participate in the testing was measured by a structure interview in the combination with the rank and procedure. Since this approach failed to discriminate any interest level among the children, it was abandoned. The authors provide detailed directions on the nature of the task, the scoring procedures, and the administration of the checks" (Lesser, et al., 1964, pp. 81-82).

As a final comment, the authors' definition of verbal ability as significance for intellectual ability should be noted: "This skill is defined as memory for verbal labels in which reasoning elements, such as those required by verbal analogies, are reduced to a minimum. Verbal ability has long been regarded as the best single predictor of success in academic courses, especially in the language and social science fields. It is involved to a marked degree in the work of all professions and most of the semi-professional areas" (Lesser et al., 1964, p. 68).

Evidence of reliability, validity, and useability:
Reliabilities (corrected from attenuation) range from .78-.94. For the entire sample, r was .93. Reliabilities were lowest for Middle-Class Jewish Children (r-.78) and highest for Middle-Class Children considered as one sample (r=.94). The reference should be consulted for details on these reliabilities as they relate to each ethnic group, each social-class group, each sex group, and each ethnic group subdivided into social-economic class components (Lesser et al., 1964, pp. 86-101).

The Verbal Scale (median r of .54) correlated least strongly with the other scales. "Since verbal ability is considered to be the most important confounding factor in testing mental abilities, these data suggest some degree of success in reducing the contributions of verbal differences to scores on other scales (at least relative to the contributions of the other scales to each other)." Further, the data suggests that, in this scale as well as among the other scales, ". . . differentiation among the various abilities is greatest for Chinese children (median r of .46) and least for the Negro children (median r of .61), despite nearly equivalent reliabilities" (Lesser et al., 1964, p. 89).

In developing the picture vocabulary, two principal procedures were followed:
1. Care was taken that the items selected would be those ". . . represented in the environments of all urban children, in the population . . ." to be sampled.
2. "In the pictorialization of the items, the artists were instructed to make their drawings of people neutral in tone and to include a minimum of the frills often associated with the pictures that appear in popular children's books" (Lesser et al., 1964, pp. 69-70). Further, the precautions with which the investigators developed the 30 item Word Vocabulary test in the try-out period strongly suggests that at least content validity can be claimed for it. First, in the selection of the items, care was taken so that they were representative ones, as indicated before. Secondly the word had to have an equivalent in another language and be similar in difficulty within this new context as determined by linquistic experts. Thirdly, a check was made to see that the original English word was retained in the process of being translated from English to another language and/or dialect and back to English. All experts were consulted for criteria to use in retaining words as in the Chinese group. Lastly, the five lists of 30 words included ". . . those displaying the best difficulty levels and their tryouts with the four ethnic groups.... These 30 words were then divided into two matched lists based upon the composite rankings of the words made by the Psychologists serving as test administrators and by the test constructors" (Lesser et al., 1964, pp. 69-74).

Bibliography of test usage:

Lesser, G.S., Fifer, Gordon, and Clark, D.H. Mental Abilities of Children in Different Social and Cultural Groups. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, Comparative Research Project No. 1635, 1964.

Davis, F.D., Lesser, G.S., and French K.G., et al. "Identification and Classroom Behavior of Gifted Elementary-School Children." Cooperative Research Monographs, 1960, No. 2, 19032.

In "The Gifted Student," OE-35016, Monograph No. 2, U.S. Office of Education, Dept. of Health, Education, and Welfare.

TEST DATA WORKSHEET

Name of Test	Wang Mental Ability Test
Author	James D. Wang
Age	Fourth, Fifth, and Sixth Grades
Area (i.e. language development)	Mental Ability
Type of measure	20 multiple-choice questions
Source from which the test may be obtained	(See reference)

Description of Measure:
"The Mental Ability Test which consists of only 20 questions was constructed in the same general manner as certain of the well-known, standardized multiple-choice types. The scoring was made thoroughly objective. The most outstanding feature of this test is that it can be taken by the children in the fourth, fifth, and sixth grades, within a total time of only six minutes" (Wang, 1941, p. 5). In a pre-trial, the test was administrated to 41 boys and 40 girls in grades 4-6" (Wang, 1941, p. 6).

Some sample items of the test are as follows (Wang, 1941, p. 117):

"1. Which one of the five words below means the opposite of south?
1west, 2 sunset, 3 north, 4 equator, 5 left.....( )
2. A toe is to a foot the same as a finger is to what?
1 head, 2 arm, 3 glove, 4 hand, 5 elbow.....( )"

Evidence of reliability, validity, and useability:
The test was correlated with the Haggmon-Nelson Tests of Mental Ability, Form A, for grades 3-8. The correlation between the two tests were .80. "A partial correlation was run between the two tests, eliminating or holding constant the effects of chronological age. This coefficient was .81" (Wang, 1941, p. 6).

Bibliography of test usage

Wang, James D. "A study of certain factors associated with children's play interests." Unpublished doctoral dissertation, George Peabody College for Teachers, 1941.

MEASURES OF PERSONALITY

TEST DATA WORKSHEET

Name of Test	The Anxiety Scale
Author	Jerry D. Alpern
Age	Pre-school
Area (i.e. language development)	Personality-anxiety
Type of measure	Modified interview
Source from which the test may be obtained	See (1) below

Description of Measure

The Anxiety Scale is a modified interview consisting of 79 items which the subject responded to by placing a steel ball in one of two boxes, and contiguous to and represented by a "happy" face of a child of the same sex as the subject, the other contiguous to and represented by a "sad" face of the same-sex child. "The Anxiety Scale interview was composed of three types of items. The first type of item was designed to elicit 'happy' responses; e.g., 'Which one looks like you when it's your birthday?' The second type of item was designed to elicit 'sad' responses; e.g., 'Which one looks like you when you've been spanked?' The third type of item, the ambiguous items, were designed to elicit either a 'happy' or 'sad' response depending upon the experience of the individual subjects; e.g., 'Which is you when you're going somewhere you've never been before?' It is from this last group of items (ambiguous items) that the anxiety scores were derived."

The first ten items of the measure included six "happy" and four "sad" ambiguous items. The other 69 items consisted of alternating ambiguous and unambiguous items, beginning with ambiguous item 11.

In administering the measure, the examiner calls attention to the "happy box" and points out that there is a picture of a happy boy (or girl) by a period. He does the same for the "sad box" and then instructs the child to put the marble in whatever box he wants to after he hears the question.

Four scores were derived from the Anxiety Scale: (1) the total "sad response" score, which is the number of times the child makes a sad response to the ambiguous interview items. (2) The speed score, which is the child's main latency on the ambiguous items, latency is defined as the time elapse between the presentation of the item or question and the response. (3) Latency Score. A basal response time was determined for each child by getting the average of the latency time on the 34 unambiguous items. The latency score, then, is the difference between the mean latency time on the 35 ambiguous items and the basal time mentioned above. (4) Variance score, which is simply the variance of the latency times on the ambiguous items.

While reliabilities were not high and validity was not demonstrated in the form of significant correlations between any of the four anxiety scale scores and either of the two criteria, teachers ranking of anxiety and a motor task, the test procedure used and the pool of items for the anxiety measure may be of interest to other researchers. Low reliability should be expected in view of the fact that the age range of the subjects was from 3-3 to 5-1.

Evidence of reliability, validity, and standardization

Test-retest reliabilities on the four anxiety scores, for the younger group (chronological ages 3-3 to 3-11) range from -.14 to .46. For the older group, with chronological ages from 3-11 to 5-1, the reliabilities ranged from .57 to .89.

None of the anxiety scale scores correlated significantly with teachers' rankings of anxiety. There was also nosignificant relationship between performance on a motor task and any of the anxiety scale scores.

Bibliography

Alpern, Jerry D. "The Relationship of an Objective Measure of Anxiety for Pre-School Aged Children to Two Criterion Measures." Unpublished Master Thesis. State University of Iowa. February, 1959.

TEST DATA WORKSHEET

Name of Test	Child Conflict Scale (Emmerich)
Author	Walter Emmerich
Age	Preschool children between the ages of three years, seven months, to five years, one month
Area (i.e: language development)	Personality (emotional conflict)
Type of measure	Rating scale
Source from which the test may be obtained	(Emmerich, 1959)

Description of Measure

The Child Conflict Scale represents "a measure of a child's conflict about his parental expectations derived from the interview protocols." Emmerich has further described the scale as follows: "This measure is composed of various aspects of behavior believed to signify that the child is avoiding an appropriate response to the item. (An 'appropriate' response was defined as one which can be scored with respect to content on the nurturance-control scale.) The conflict measure used was a composite one, as it was not feasible to evaluate separately the reliability and validity of each indicator of conflict and then select the best of these. Consequently, this measure includes some behaviors which may not be actual manifestations of attempts to avoid responding to the item. Also, it was not feasible to determine empirically which signs of avoidance behavior were of greater strength. Rather, it was assumed a priori that certain types of avoidance are stronger than others."

In scoring the child's responses on the scale the following procedure was follows:

"A five point scale of conflict was developed and applied to each of the child's responses to the six mother-child and six father-child items." In addition, conflict categories were used. The five point scale ranged from 0 for no conflict to 4 for very strong conflict. Examples of the eight conflict categories included response latency and hesitation as defined below:

"1. Response latency. Refers to the period following the examiner's presentation of an item before the child gives any response. A 'long' response latency is defined as ten seconds or more.

2. Hesitation. The child starts to respond to an item, but needs slight encouragement to complete the item" (p. 302).

Finally, "the child's total score on the six items for a parent was used as the index of the child's conflict about the parent's attitude toward him" (p. 279).

Evidence of reliability, validity, and standardization

The scale was administered to the same sample used in the Child Nurturance-Control Scale by Emmerich. Within this sample, both rater reliability and test-retest reliability were derived. In the first case, "the correlation between two judges' independent ratings of a sample of items was +.86." In the later instance, the test-retest reliability coefficients of 26 children were +.44 for the mother-child items and +.56 for the father-child items" (p. 279).

Bibliography

Emmerich, W. Parent identification in young children. Genetic Psychology Monograph, 1959, 60, 257-308.

TEST DATA WORKSHEET

Name of Test	Draw-A-Group Test
Author	A. Paul Hare and Rachel T. Hare
Age	6-10
Area (i.e. language development)	Social and personal adjustment of child.
Type of measure	Projective drawing technique
Source from which the test may be obtained	See reference

Description of Measure:

"The Draw-A-Group Test is designed to identify the status of the individuals in the group from leaders through followers to the isolates." The test is an attempt " . . . to demonstrate that the children in any one group can be ranked in a hierarchy of leaders, followers, and isolates on the basis of their drawings of the group."

"The test is administered in the classroom. The drawings are done with colored crayons on 18" x 24" paper. The instructions are as follows:

> Think of the children you like to play with most on the playground.
> Now think of the thing you like to do best with this group of children.
> Then draw a picture of your group doing the thing you like to do best.
> When you are through we will write down what is going on in your picture.. . .

As the children finish, information about their pictures is recorded by the experimenter including the names of the children in the picture (in the order drawn) and what is happening as told by the child." In half of the classes, the child was asked specifically to include himself in the picture.

The criteria for analyzing the drawings are the result of suggestions in the literature such as a range of colors used, use of dark or somber colors, use of space and relative size of figures, hostile or threatening aspects of figures, unusual characteristics of figures such as distortions or omissions, and bizarre or unusual objects. Additional criteria are used which seem meaningful relative to the nature of the instructions and of group activity in general. These are the number of figures in the drawing, the nature of the activity, the centrality of the subject in the activity, the inclusion of children from other classes, the inclusion of authority figures such as the teacher, partial drawings on the back or evidences of resistance to the task, and the general appropriateness of the drawing and the child's description of the activity to the instructions."

The authors have also provided detailed and specific examples of case interpretations based on these drawings (Hare and Hare, 1956).

Evidence of reliability, validity, and standardization:

"The sample for the study consists of 10 classes of children, five from the first, second, and third grades in two private day schools and five from two public schools in New England. The classes range from eight to 24 members giving a total N of 166. The children are from the middle and upper socio-economic levels and include several recognized emotional problems and physical handicaps. The ages range from six to 10 years."

The authors have not included any data on the test's reliability. Validity was determined by the correlations between the picture rankings by the clinicians and the teachers' rankings. "Positive correlations between the picture ranking and the teacher's ranking. . . /were/ obtained for nine of the 10 classes with a mean. . . of +.52 which is statistically significant. . .since the probability of a mean correlation this high or higher if the true correlation is zero is <.01. In addition for eight of the 10 classes the correlations were statistically significant with p<.01. and .05. Within these eight classes, the correlations ranged between .47 and .92. "This evidence supports the hypothesis that a child's drawing of his group is related to his position in the group." On the supposition of that ranking extreme distributions depending on class would be easier than ranking the entire class distribution, rank correlations were also computed ". . . for a collapsed ranking consisting of only the upper and lower thirds of each class. This operation. . ./increased/ the mean correlation to +.62. The difference between this mean correlation and the mean correlation for the entire ranking. . . /was statistically/ significant.. . ." (Hare and Hare, 1956, pp. 57-58).

Bibliography:

Hare, A. P. and Hare, Rachel T. The draw-a-group test. Journal of Genetic Psychology, 1956, 89, 51-59.

TEST DATA WORKSHEET

Name of Test	Fruit Distraction Test
Author	Santostefano, S.
Age	6-13
Area (i.e. language development)	"a measure of the constricted-flexible cognitive control"
Type of measure	Individually administrated performance test
Source from which the test may be obtained	See Santostefano (1964) in bibliography

Description of Measure:

"The Constricted-Flexible cognitive principle . . . concerns the manner in which a person deals with a stimulus field containing contradictory or intrusive information. Flexible control individuals, when dealing with a central task, can selectively withhold attention from intusive information and thereby are not disrupted by it. Constricted control persons, on the other hand, have difficulty selectively withholding attention from intrusive stimuli and their performance with the central task is disrupted."

"This test is made up of two cards (10" x 15"). Card 1 contains 50 drawings of apples (red), bananas (yellow), bunches of grapes (blue), heads of lettuce (green), (12 each of bananas and lettuce and 13 of grapes and apples) which are randomly distributed over the card in ten rows, five fruit in each row. Card 11 contains the same colored fruit arranged identically to the first card. However, in addition, immediately above, below, or to the side of each fruit is one achromatic line drawing of one of the following: chair, car, airplane, shoe, telephone, clock, (non-food object); bottle of milk, glass, spoon, ice cream cone, loaf of bread, cake (food objects). These achromatic line drawings are considered 'intrusive information' with respect to the central task (described below) and each appears four times on the card randomly distributed with the last two fruits having no intrusive information by them.

S was given a practice card containing five of the colored fruit and E made sure that S knew the names of the fruit and could correctly read aloud the names of the four primary colors from left to right. S was then asked to read aloud the colors of the fruit on Card 1 as rapidly as possible and without skipping. E recorded the reading time cumulatively by pairs of lines and reading errors. Errors consisted of omitting a color while reading, reading the name of a fruit rather than the color it bears, or reading an incorrect color in terms of the correct sequence. After S completed the first card, he was given Card 2 on which were the same colored fruit but with the surrounding 'distractions' described above. S was asked again to read aloud the colors of the fruit as fast as possible and to try not to pay attention to any of the pictures that were around the fruit. E recorded time to read the colors cumulatively by pairs of lines, and again recorded reading errors. After S completed the test, E removed the chart and asked S whether he happened to notice and could recall any of the pictures which were surrounding the fruit.

The test yields three measures: (a) Reading Time Distractibility Score (time to read Card 2 minus Card1); (b) Reading Errors Distractibility Score (total reading errors with Card 2 minus total reading errors with Card 1); and (c) the number of intrusive stimuli recalled" (Santostefano, 1964, p. 214).

Evidence of reliability, validity, and standardization:
Santostefano (1964) found significant differences among 3 groups (Brain damaged children, orphaned children, and public school children) on the variable of reading time differences between "normal" and "distracting" conditions. Differences in number of reading errors between the normal and distracting conditions were not significant for the three groups of children, although they were in the same direction as reading time differences. No significant differences were found between the groups on the number of "intrusive object" recalls, i.e. memories of distracting stimuli. Scores on this measure were not significantly related to intelligence. (Santostefano, 1964).

Bibliography:

Santostefano, S. J. Clin. Psychol., Cognitive controls and exceptional states in children, 1964, 213-218.

TEST DATA WORKSHEET

Name of Test: General Anxiety Scale for Children (GASC)

Author: Sarason, Seymour, et al.

Age: Grades 1-9

Area (i.e. Language development: Personality-Anxiety

Type of measure: Test--group

Source from which the test may be obtained: See (1) in bibliography.

Description of Measure:

The GASC is made up of 45 items of which 34 are designed to measure general anxiety and 11 constitute a lie scale. Examples of items in this scale are as follows:
"1. When you are away from home, do you worry about what might be happening at home? 2. Are you frightened by lightning and thunderstorms? 3. When you are home alone and someone knocks on the door, do you get a worried feeling? 4. Do you worry that you are going to get sick? 5. Do you worry about whether your father is going to get sick?

Evidence of reliability, validity, and standardization:

There are no reliability data given for the GASC. Several kinds of data on validity are given by Sarason. He describes correlations between the Test Anxiety Scale for Children and the General Anxiety Scale for Children from grades 1 to 6 in American schools and from 1 to 5 in English schools. As might be expected, the correlations between the two measures are lower in the lower grades, probably as a result of lower reliabilities in the first and second grades. The median correlation between TASC and GASC for American boys was .67, and for girls, .56. The girls in the study do not show the same pattern of low reliabilities in the lower grades as did the boys, which is probably explainable in terms of sex differences on GASC scores.

Low negative correlations are reported by Sarason between GASC and I.Q. scores and GASC and mean achievement scores.

"It is our opinion that even a cursory study of the GASC . . . would suggest that there are more items having to do with possible bodily harm than with the establishment and maintenance of social relationships or with how others perceive the child. In other words, the GASC seems to contain items more pertinent to the anxieties of boys than to the anxieties of girls" (Sarason, 1960).

Bibliography:

1. Sarason , Seymour B., et. al. Anxieties in Elementary School Children. Yale University, 1960.

TEST DATA WORKSHEET

Name of Test: Intolerance of Ambiguity Scale (Mussa)

Author: Rolf E. Mussa

Age: Sixth grade

Area (i.e. language development): Personality (rigidity)

Type of measure: Self-rating scale

Source from which the test may be obtained: See Reference in bibliography.

Description of Measure:
Mussa defines intolerance of ambiguity "as a general trait, characterized by need to structure an unstructured situation, to avoid ambiguous stimuli, by a tendency to resort to black-white solutions and to maintain a 'status quo' rather than to take a chance.... Obviously, the concept, intolerance of ambiguity, is closely related to rigidity." The 12 items comprising the Intolerance of Ambiguity Scale reflect the definition of this general trait. The score for the scale simply consists of the total number of items answered in the direction indicative of intolerance of ambiguity. Keys for making this judgment are provided in the article. The scale itself has been administered on a group basis by Mussa. Samples of the items from the scale follow: "The answer in parenthesis is scored as indicative of intolerance of ambiguity."

"1. I often wish people would be more definite about things (1). (agree).
2. I don't like to undertake any project unless I have a pretty good idea as to how it will turn out (1) (agree)."

Evidence of Reliability, Validity, and Standardization:
The scale was administered as part of a series of group tests to 280 sixth grade pupils in the public schools of a midwestern community of 80,000. "About half of the subjects came from experimental classes with special emphasis on the dynamics of human behavior while the other half came from regular classes, and served as controls for the experimental classes in another research study."

Though no evidence of reliability is provided for this scale, several indices suggest construct and/or concurrent validity

1. An r of .41 (N 232) between the Intolerance of Ambiguity Scale and the Children's Antidemocratic Attitude Scale indicates that the former general trait is positively related to ethnocentrism.

2. Children who scored above the mean (Highs) on reliable tests of social and physical causation achieve significantly lower scores on the scale than those who scored below the means (Lows). These statistically significant differences "held up" even among low and high subjects matched for I.Q.

scores.

3. The correlation between scores on the ambiguity scale and the I.Q. test was -.07.

Bibliography:
Munns, R. E. A comparison of "high causally" and "low causal[illegible]" oriented sixth grade children on personality variables indicative of mental health. Proceedings of the Iowa Academy of Science, 1959, 66, 388-394.

TEST DATA WORKSHEET

Name of Test	Irritability Scale
Author	Graham, Frances K., Matarazzo, Ruth G., and Caldwell, Bettye M.
Age	Infants (Neonates 1-5 days old)
Area (i.e. language development)	Physiological irritability
Type of measure	Rating scale to determine ". . . how sensitive the infant is to stimulation and not how loudly and lessly he cries" (Graham, et al., 1956, p. 13).
Source from which the test may be obtained	Authors

Description of Measure:
"The experimenters had considerable experience with the phenomena to be rated, and the ratings probably depend to a considerable extent upon 'expertness'" (Graham, et al., 1956, p. 12).

"In an attempt both to reduce the dependence on 'expert judgment' and to increase reliability, we analyzed the behavior upon which such ratings would be based and broke it down into separate components. The Ss were first rated in terms of these separate components or submeasures, and only afterward was an over-all rating made on that scale."

Several factors which contribute to this rating include (1) the intensity of stimuli which evoke crying; (2) the state of the infant; (3) the cry of the infant; and (4) the ease of quieting."

"A three-point scale, with numerical values of 0, 1, and 2, was provided for the rating of irritability." A score of 0 represented normal irritability, 1, a 'just perceptible' form of abnormal behavior and 2, the extreme and easily identified form of abnormal behavior. In anchoring the points, the definition for each classification include crying in response only to intense and external stimulation for 'normal irritability'; crying or 'fussing' in response to mild stimulation for 'just-perceptible irritability'; and crying in response to many mild stimuli for 'abnormal irritability'. (Graham et al., 1956, p. 13).

Evidence of Reliability, Validity, and Usability:
"Reliability of the procedures was measured by split-half correlation, test-retest agreement and interscorer agreement as applicable. All tests appeared to be satisfactorily reliable" (Graham, 1956, p. 32). Split-half reliabilities for the pain threshold scores for both traumatized and nontraumatized ranged from .82 to .97. The other subtest scores could not be satisfactorily divided into comparable halves. Test-retest reliabilities for the Pain, Maturation, and Vision scales were from .62-.69. On the Irritability and Vision Scales, the test-retest agreements after 24 hours were respectively 75% and 86% of perfect agreement. Statistics on interscorer

agreement were as follows: (1) Maturation Scale (r=.97); (2) Vision Scale, (r=.90); (3) Irritability Scale (68% perfect agreement). Finally, there was 79% perfect agreement in the Tension Scale (Graham, 1956, pp. 21-22).

In addition to the reliability of individual raw scores, statistics are provided for differentiating groups and accurate identification of individuals as members of a particular group. In the former case, statistically significant differences (p of .05 or .01) between the normal and traumatized groups occurred on all five scales based on F test, t test, and Chi square (Graham, 1956, p. 24). In the latter instance, two pediatricians rated 25 cases of traumatized infants with respect to the severity of the trauma. ". . . the extent of the agreement was high and significant as estimated by a correlation ratio of .86 (Graham, 1956, p. 18).

Bibliography of Test Usage:

Graham, Frances K., Matarazzo, Ruth G., and Caldwell, Bettye M. "Behavioral Differences Between Normal and Traumatized Newborns...." Psychological Monographs, 1956, 70, No. 20, (Whole No. 428).

TEST DATA WORKSHEET

Name of Test — Life Situation Perception Test

Author — Ladonko, A.

Age — Approximately six

Area (i.e. language development) — Personality

Type of measure — Projective

Source from which the test may be obtained — See Ladonko (1962) in bibliography

Description of Measure

This measure is described by the author as "...a new method of percept-diagnosis which besides being easy to perform, offers supplementary information to the Rorschach test and gives the psychologist a still better view regarding some of the tested person's situational reactions."

Administration of the measure is simple. "After seating the patient comfortably, the testing psychologist proceeds: 'Now please look around yourself and try to make a description of what you see.' This explanation leaves the patient in complete freedom to interpret the environment as he chooses."

"Everything the tested person says is recorded, as is done with Rorschach responses. During this time, the psychologist must be completely silent until the patient runs out of answers or has to be interrupted in order to end the test. Then he says: 'This will be enough.' In case the patient has difficulties after the first answers, the psychologist may ask: 'Do you see anything else?' If there is an answer, it must be recorded. When the test is finished, the time employed for the description is written down. Later, the answers are transcribed on a scoring sheet, classified, scored and interpreted as if it were a Rorschach test."

Differences between scoring of the Rorschach test and this measure are described and modifications in scoring and interpretation are noted very briefly. Four new types of responses given by subjects taking this test, but not found in Rorschach protocols, are described and exemplified. They are: personal answers; affective answers; subjective answers; "WD" answers... Some personality correlates of each of the four response types are suggested. For example, affective answers "...indicate the influence of affects in the person's behavior. When there are five or more affective answers, affects dominate over all other motivations of behavior."

Evidence of reliability, validity, and standardization

None

Bibliography

Ladonko, A. The life situation perception test. Journal of Clinical Psychology, 1962, XVIII, 297-299.

TEST DATA WORKSHEET

Name of Test	Lipsitt Self-Concept Scale for Children
Author	Lipsitt, Lewis P.
Age	Fourth, fifth, and sixth grade children
Area (i.e. language development)	Self-concept
Type of measure	Self-rating Leikart-type scale
Source from which the test may be obtained	Lipsitt, L. P. Child Develpm., 1958, Vol. 29, 463-472.

Description of Measure:

"The self-concept scale contained the following 22 trait-descriptive adjectives, presented here in the order used in the scale: friendly, happy, kind, brave, honest, likeable, trusted, good, proud, lazy, loyal, cooperative, cheerful, thoughtful, popular, courteous, jealous, obedient, polite, bashful, clean, helpful. Each of these adjectives was prefaced by the phrase 'I am . . .' and was followed by a five-point rating scale. Nineteen were considered as positive or socially desirable attributes, while three were considered negative (italicized above). The rating categories, scored from 1 to 5, were entitled not at all, not very often, some of the time, most of the time, and all of the time. A score of 1 was received on an item if S checked the first category, a score of 5 if the last category were checked, except in the case of the three negative adjectives which were scored in inverse fashion. A score on the self-concept scale was obtained for each S by summing the ratings ascribed to himself on each item. Lower scores were presumed to reflect degree of self-disparagement.

The ideal-self scale contained the same adjectives but here each was prefaced by 'I would like to be . . . .' Again, S rated himself on each item on similar scales. The discrepancy scores employed in the present study consisted of the simple subtraction of the total self-concept score from the total ideal-self score."

The discrepancy score was used because it is customarily considered "... as reflecting degree of dissatisfaction with oneself."

Evidence of Reliability, Validity, and Standardization:

The scale was administered to 300 fourth, fifth, and sixth graders on the same day. The sample was subdivided as follows: fourth grade, 47 boys and 62 girls; fifth grade, 50 boys and 61 girls; sixth grade, 41 boys and 37 girls.

Test-retest reliabilities over a two week period ranged from .73 to .91 for the self-concept scale and .57 to .72 for the discrepancy score. In the former scale all the correlation coefficients obtained statistical significance beyond the .001 level. In the latter scale, four of the six correlations obtained significance at the .001 level, and two at the .01 level. No reliable grade or sex difference of mean self concept scores were evident. However, ". . . for all six grade sex combinations, the self-concept measure correlated significantly with performance on the CMAS Children's Manifest Anxiety Scale, by comparable correlations for the discrepancy scores

were somewhat less and in somewhat less and in some cases not reliable" (Lipsitt, 1958).

Bibliography:

Lipsitt, L. P. A self-concept scale for children and its relationship to the children's form of the Minifest Anxiety Scale. Child Develp., 1958, Vol. 29, 463-472.

TEST DATA WORKSHEET

Name of Test: Miniature Situations Test

Author: Santostefano, S.

Age: 6 to adults

Area (i.e. language development): Personality

Type of measure: Individually-administered performance test

Source from which the test may be obtained: See Santostefano (1957) and (1962) in bibliography.

Description of Measure:

"The Miniature Situations Test (MST) invites children to act upon one of two objects in order to assess overt, coping responses (as opposed to verbal, self-report responses) indicative of personality functioning. For example, the child chooses between breaking a light bulb versus watering a plant and between drinking from a bottle versus a cup. In contrast to the naturalistic method of situational testing, the MST technique imposes some degree of restriction on the freedom of expression allowed the personality in order to (a) produce unequivocally identifiable responses executed within a few seconds and thereby available in manageable units for study, and (b) provide circumscribed and identical stimuli to all subjects, minimizing variance created by incidental stimuli in the situation" (Santostefano, 1965, p. 418).

"Tasks were devised which required Ss to act upon stimulus material and which satisfied the following criteria: (a) the stimulus material should be administered on the top of a desk; (b) only a single act, which could be executed within a few seconds, should be sufficient to complete each task; (c) the meaning of a task should be agreed upon by competent judges; (d) the nature of a task should reduce the likelihood that the significance of the act performed would be apparent to S.

The tasks were combined into pairs constituting miniature situational tests. Tasks were paired which required about the same amount of effort to complete, and also to insure that as wide a variety as possible of combinations of aspects of personality would be represented in the situations. The battery of tests employed in this study consisted of 41 situations of pairs of tasks" (Santostefano, 1960, p. 373).

A factor analysis of the measure resulted in the following factors:

"Factor I seems to to discriminate persons who prefer to aggress overtly and forcefully from those who are more inhibited and more reluctant to exhibit forceful aggression.

Factor II seems to differentiate persons who avoid unfamiliar or unknown stimuli which are viewed by them as potentially injurious, threatening, and difficult to handle from those who approach these stimuli.

The upper members of the situations defining Factor III suggest a displaying of the self in conspicuous, solo verbal behavior; the lower members suggest a preference for passive, inconspicuous behavior.

With the exception of situation 11, the upper members of Factor IV contain symbolic acts of aggression as contrasted with the overt, forceful aggression represented in Factor I. The behavior of the lower members uniformly represents order and neatness.

Because situation 10 also contributes to Factor IV, the description of Factor V is considered to be tentative. By timing and hand-cuffing E (situations 14 and 15), S seems to exercise some degree of control over E, while the alternates suggest compliance and acceptance of control exercised by others" (Santostefano, 1960, pp.375-376).

Evidence of reliability, validity, and standardization:
"The MST performance of brain damaged, orphaned, and public school children was examined to determine if differences in their behavioral responses would be consistent with personality characteristics associated with these groups, thus providing support for the construct validity of the method. A number of the situations discriminated among the groups in directions which provided such support. For example, brain damaged children showed more aggression (tearing paper versus repairing torn paper with scotch tape) and a greater need for structure (tracing a design versus coping it free hand). Orphaned children, on the other hand, showed that oral-regressive considerations dominated their encounters with objects (e.g., they chose significantly more often to drink water from a baby's bottle versus a cup" (Santostefano, 1965, p. 421).

In the same study, Santostefano states, "It should be noted that no frequencies were observed to occur in directions contrary to expected differences between the groups."

In general, the results provide some support for the validity of the constructs presumed to be tapped by several of the situations and present implications for personality research and clinical assessment.

Bibliography:

Santostefano, S. Forced-choice acts as personality measures. Unpublished doctoral dissertation, Pennsylvania State University, 1957.

Santostefano, S. An exploration of performance measures of personality. J. clin. Psychol., 1960, 16, 373-377.

Santostefano, S. Miniature situations as a way of interviewing children. Merrill-Palmer Quart. Behav. Devel., 1962, 8, 261-269.

Santostefano, S. Construct validity of the miniature situations test: I. the performance of public school, orphaned and brain damaged children. J. clin. Psychol., 1965, 21, 418-421.

TEST DATA WORKSHEET

Name of Test	Mooseheart Wishes and Fears Inventory
Author	Mooseheart Laboratory for Child Research Mooseheart, Illinois
Age	4-16 years
Area (i.e. language development	Personality
Type of measure	Projective technique
Source from which the test may be obtained	Single mimeographed copy may be obtained from C. Richard Elich, Child Guidance Clinic, Mooseheart, Illinois

Description of Measure:
The information obtained on the Wishes and Fears Inventory ". . . is used as a projection of the child's personality and is so interpreted." The inventory consists of a series of questions which have no correct answers whatsoever and hence assume a projective nature. "While the form of the questions varies, replies to questions on each of these items and reasons for each of the responses are requested:

3 Wishes
2 Positive Identifications
2 Negative Identifications
2 Desired Activities
2 Undesired Activities
2 Changes Desired (in oneself)
2 Fears
1 Earliest Recollection

The inventory comes in two forms, Form A designed for younger children, and Form B for older students. In a former, the questions ". . . may be read verbatim if desired, and the child's responses noted under each heading. The child should be asked 'Why?' he chose that response since the reason for his choice will often give the clinican more insight into the child's personality than will the choice itself."

Examples of the questions on this form are given below:
"Wishes: With young children this can be presented as the gifts of a magician. E.G. "Do you know what a magician is, a magic man? Today imagine that I am a magician, can give you anything in all the world. Of all the things that a magician can give you what would you wish for first?"

2.

3.

Pos. Identification:
E.G. "I am still a magician who can make you any body you want to be. Now of all the people you have ever hears of, people you have seen in the movies, in the funny papers, hears on the radio, read about or know them; maybe they are alive or dead, mayber they are real or imaginary, what person would you most want to be

2.

3.

Neg. Identification:
E.G. "I an still a magician, and because I like you I don't want to make any mistakes. Of all the people you have ever heard of, know, seen in the movies, if you had to choose someone to be like who is the last person you would choose?"
. . . . . "Why would you choose them last?"

2.

3.

"Form B is a record blank and may be used with the questions given on Form A or with simplified questions. . . (as below):

1. If you could have anything you wished for, what would be your wish? Why? What else would you wish for? (Three answers).

2. Who would you most like to be like? Why? (Two answers).

3. Who wouldn't you like to be like? Why? (Two answers).

In using and interpreting the Inventory, two broad generalizations appear. First, "The Inventory has been found valuable in use with children between 4 and 16 years of age and yields especially rich material in the age range 7 to 13 years." Secondly, the recommendation is to obtain data from two general sources of information. One suggestion is that the test behavior of the S be noted in terms of several dimensions which are outlined (e.g., vagueness of response, suspicion and resistance, etc.). A second source of findings stems from the responses on the Inventory which may be classified in terms of wishes indicated by the well adjusted or the maladjusted. Guidelines for making this decision are outlined.

As a final comment, some work has been done with the scale to determine "age trends and sex differences in the wishes, identification, activities, and fears of children" (Winker, 1949).

Evidence of reliability, validity, and standardization:
The scale has been used with 750 children at the Mooseheart Laboratory for Child Research (Winker, 1949, p. 191). "The Inventory (however) was purely exploratory and . . . no statistics (exist) on reliability and validity" (Personal Communication).

Bibliography:

Kilch, C.R. Personal Communication. July 6, 1966

Mooseheart Laboratory for Child Research. Mooseheart, Ill.: The Wishes and Fears Inventory. Undated (mimeographed).

Winker, J.B. Age trends and sex differences in the wishes, identification, activities, and fears of children. Child Develpm., 1949, 20, 191-200.

TEST DATA WORKSHEET

Name of Test: Mummery Scale of Ascendant Behavior

Author: Dorothy V. Mummery

Age: 3 to 5

Area (i.e. language development): Personality (social acceptability or social maturity)

Type of measure: Rating scale

Source from which the test may be obtained: Mummery, 1963

Description of Measure:
Mummery's Scale is based on the following definitions of ascendant behavior, "ascendant behavior is any kind of behavior by which an individual attains or maintains mastery of a social situation or attempts to attain or maintain mastery so that he is in control of his own activities and can carry out his purposes. At the preschool ages, attempts at mastery of a situation include:

1. Attempts to secure materials he wants from his companions.
2. Attempts to direct or influence the behavior of companions.
3. Attempts to defend himself, his possessions, and activities and to resist mastery includes:
4. Success in the above three types of attempts."

In accordance with this definition, this author developed 79 categories of children's behavior classifiable under six headings. The headings and examples of the behavioral categories are given below:

I. Verbal methods of securing play materials. (15 items, 1-15). The child uses a declarative statement to express a desire for a toy which his companion has, or which lies near his companion.

a. 'Want that over ther.'
b. 'Hey, I need that scoop.'

. . . . . .

II. Physical methods of securing play materials. (9 items, 16-24). Child takes a toy from the possession of his companion; i.e., he immediately picks up a toy which his companion has been using but has just put down.

a. A. puts shovel down while he is emptying his dish. B. quickly picks it up.

. . . . . .

III. Verbal methods of directing companion (21 items, 25-45). Statements in which he suggests a certain mutual activity by the word 'Let's,' 'We,' or similar expressions.

a. 'Let's make a road.'
i. 'Shall we build a bigger hill?'

. . . . . .

IV. Physical methods of directing companion (8 items, 46-53). Child attempts to force or inhibit companion's activity by a non-verbal threat of force; e.g., doubles up fists as if to strike.

a. A. is making a loud noise which B does not like. B doubles up his fist as if to strike him, but withdraws it.'

. . . . . .

V. Verbal responses. (13 items, 54-66). The child flatly and emphatically refuses to give up the toy or to follow the direction. Differs from 54 mainly in manner and tone; he may even yell or scream.

b. 'Stop that!'

c. 'No! I won't'

. . . . . .

VI. Physical responses. (13 items, 67-79). A child defends his possession of a toy by calmly taking the toy from the hand of the child who has just succeeded in getting it from him, but who is now offering no active resistance; i.e., he simply 'lifts' it from his now unresisting campanion.

a. The moment A. puts his shovel down, B. picks it up. A. takes it from his hand and B. lets him, giving no physical sign of resistance whatever." (Mummery, 1947, pp. 42-46).

The exact procedure in recording ascendant behavior was to pair children for playing games, record their behavior verbatum from the moment play began for a five minute period in a controlled play situation, and then to complete the record, "by adding the number of instances falling under each item for each child and recording the totals in the space provided on the blank" (Mummery, 1947).

The details of conducting the observations, recording the responses, and deriving the two weighted scores onto one unweighted score (the Jack Score) have been described in full. The precise toys used in the controlled play have also been delineated (Mummery, 1943; 1947).

The later publication, Stott and Mummery (1956), the 79 categories were regarded as representing three main behavior areas: "1. gaining possession of coveted objects, 2. directing or influencing the activities of another, and 3. defending possessions and activities" (Stott and Mummery, 1956, p. 112).

Each of the 70 categories "were identified and carefully defined and delineated. And each of the items was illustrated by specific examples. On the basis of Thurston's equal-appearing interval technique, 41 experts were asked to rate each item on a scale ranging from 1 (most unacceptable) to 11 (most acceptable). The scale value for each item was then based on the median judgment by the 41 specialists (Mummery, 1947). The directions to the judges for evaluating each item have been described in considerable detail (Mummery, 1943).

Evidence of reliability, validity, and standardization:
Test reliability and validity were established on initial scores of 42 preschool children ranging in age "from 42 to 61 months with a mean of 49 months. Test scores indicate that the group represented a definite selection upward in intelligence, the mean I.Q. being 119. The distribution of the subjects among the 7 divisions of the Minnesota Classification of Occupational Status prepared by Goodenough showed that 71 percent fell in the first two classifications as compared to 7.2 percent in the general population. The group thus represented a selection upward in socio-economic status as well as intelligence."

The reliability with which the device measures ascendant behavior was determined by correlating scores on pairings 1 and 3 against 2 and 4 in the series of 5 pairings. These coefficients were .70, .62, and .61, which, by application of the Spearman-Brown formula, became .85, .80, and .80 for the Jack, acceptable, and combined ascendance scores respectively."

By two different procedures, reliability of observers was also determined. "By the first method, reliability was expressed as the correlation between Observer A and Oberver B on the total number of instances recorded for each of the categories during 58 consecutive five-minute pairings. In order to reserve the three-and-four year-old children for the main part of the study, a group of 23 five-year-old children attending the Junior Primary of the University of Iowa Elementary School was used in these observations. This coefficient was .98.

Observer reliability was recalculated on the same data by the method of correlation between weighted scores assigned to each child in each pairing from the behavior items obtained by two observers recording simultaneously. These reliabilities expressed as coefficients of correlation on scores in the last 26 consecutive pairings, are .92 and .91 for the acceptable and combined scores respectively. Since the main data of the study are based on 5 pairings for each child, these coefficients, based on single pairings, would seem to represent 'minimum' reliabilities." (Mummery, 1947, pp. 54-55).

Validity of the data was derived from two different methods. First, "the opinions of the experts was sibned scale-value to the different categories of behavior in itself constitutes the validity of the weightings for acceptibility and unacceptability that are used in this study" (Mummery, 1947, p. 57). Secondly, attempts were made to correlate weighted scores by observers to the teacher ratings of behavior on specially constructed five-point scales. Samples of these scales and instructions to the raters have been presented elsewhere (Mummery, 1943). In general, the teachers' scales related to the frequency of ascendant behavior of their child. (Scale A), to the frequency of the success (Scale B), and to the methods he used in attaining and maintaining ascendant behavior (Scale C). These correlations between experimental frequencies of success and teachers ratings of success, though in a positive direction, were too low for reliability. Consequently, Mummery concluded as follows: "One of two conclusions may be accepted: 1) either teachers are unable to rate children on the type of methods used, or 2) the methods used by children vary more with the situation than does the amount of ascendant displayed by them" (Mummery, 1947, p. 59).

Bibliography:

Mummery, Dorothy V. An analytical study of ascendant behavior of preschool children. Iowa City, Iowa: State University of Iowa, unpublished dissertation, 1943.

Mummery, Dorothy V. An analytical study of ascendant behavior of preschool children, Child Develpm., 1947, 18, 40-81.

Stott, L.H., & Mummery, Dorothy V. Adult attitudes toward ascendant behavior in young children, Merrill-Palmer Quarterly, 1956, 2, 110-120.

TEST DATA WORKSHEET

Name of Test: Perception Score Sheet

Author: Combs, A. W., & Soper, D. W.

Age: Kindergarten - first grade

Area (i.e. language development): Observation schedule

Type of measure: Perceptions of self and perceptions of others

Source from which the test may be obtained: See Combs, A. W., & Soper, D. W. in bibliography

Description of Measure:
This is a frame of reference to be used by specially trained personnel to evaluate the self-perceptions of children and their perceptions of other significant persons in their environment. There are 39 items or areas on which each child is rated, using a five-point scale. Most of the items are described and clarified, as in the following example:

"I. Perceptions of Self

SA--Self Generally

1. Adequate--Inadequate
This category refers to the child's overall feeling of adequacy. It represents a global feeling of being fulfilled, enough, sufficient, as contrasted to feeling ineffectual, lacking or the like. It should be recognized that it is possible for a child to feel generally adequate while at the same time possibly feeling inadequate with respect to some individual items on our check list.

SAI--Self as Instrument

This section has to do with various categories describing how the individual perceives himself as an instrument for carrying out his purposes."

"Adults helpful--Hindering

Adults are need fulfilling or need frustrating or depriving. Adults are seen as a source of satisfaction of one's wants, hopes, desires. Adults are facilitating to one's goals and purposes."

The items are grouped as follows:

I. Perceptions of self
- Self generally
- Self as instrument
- Self with other children
- Self with adults
- Self with teachers
- Self and the school curriculum

II. Perceptions of others
Perceptions of children
Perception of adults
Perception of teachers
Perception of school

This observation schedule seems to be appropriate for most elementary school children, although Combs and Soper used it only with kindergarten and first graders.

Evidence of reliability, validity, and standardization:
Scores of the kindergarten group were factor analyzed. "An examination of the communalities of the 39 items...shows only six items with communalities of less than .70 and indicates a high level of reliability for the scales used in this research."

The strongest factor identified was labeled "The General Adequacy Factor." "This is an all-inclusive factor including every item on the Perceptual Score Sheet with the exception of 'enough in music' and 'enough in art.' This factor alone accounts for 67 percent of the total variance of the 39 items. Its strength and pervasiveness is further emphasized by the fact that the next strongest factor (II) accounts for only 9 percent of the total variance. Factor I, by itself, accounts for more of the total variance than the sum of all other identifiable factors in the analysis.

Bibliography:

Combs, A.W., & Soper, D.W. The relationship of child perceptions to achievement and behavior in the early school years. Cooperative Research Project No. 814. University of Florida, 1963.

TEST DATA WORKSHEET

Name of Test: Psychiatric Follow-Up Coding (Rubin)

Author: Eli Rubin

Age: Kindergarten, first, and second grade

Area (i.e. language development): Personality (Social and emotional adjustment)

Type of measure: Psychiatric rating scale

Source from which the test may be obtained: Author

Description of Measure:
According to Rubin, "151 children from kindergarten ough second grade were re-
ferred to the Wyandotte Michigan special class proj as candidates for place-
ment in a special class program..." (Rubin, et al., 66, p. 48).

"A psychiatric interview, along with the social history and the parent interview, was part of the initial screening examinations for every subject. The method of assessing change in psychiatric status was based on a comparison by a child psychiatrist of the initial psychiatric evaluation with his psychiatric interview conducted at the time of exit, and again at the time of follow-up. At the times of reassessment, the phychiatrist based his judgments on the psychiatric interview exclusively, in dependent of any other knowledge of the child's interim adjustment. His comparative judgments then served as the basis for the psychiatric follow-up coding, consisting of sixteen separate scales (nine-point scales)" (Rubin, 1966, p. 6).

These scales provided ratings of the following areas of behavior. (1) Spontaneity and the interview situation, (2) Cooperativeness, (3) Impulsivity, (4) Reality-orientation (5) reality testings, (6) hyperkinetic sydromes behavior science (7) hyperkinetic syndrome--physical signs (8) Neuropsychiatric symptoms, (9) feelings of self-confidence and self worth, (10) child's perception of stress in peer relations, (13) emotional maturity, (14) group belongingness, (15) capacity, for emotional relationships with others, and (16) overall progress in emotional adjustment. (Psychiatric Coding, n.d.). An example of the psychiatric coding is given below:

"CODING SAMPLE --PSYCHIATRIC FOLLOW-UP

Impulsivity: This scale evaluates the degree to which the patient has changed in the control of impulsive motor acts, gestures, or behavior as observed in the interview.

1. Marked decrease in impulsivity
2. Moderate decrease in impulsivity
3. Minimal decrease in impulsivity
4. No change in impulsivity
5. Minimal increase in impulsivity
6. Moderate increase in impulsivity
7. Marked increase in impulsivity
0 No information
Confidence rating" (Rubin, et al., 1966, p. 69).

Evidence of reliability, validity, and standardization:
No interrater reliability measurement was attempted . . . but each of the 16 scales was provided with an opportunity for a confidence rating on a four-point scale. Only those scales that are rated less than four (no confidence) were accepted for inclusion." (Rubin, et al., 1966, p. 69). The scale for the physician's confidence in rating, together with a score for each level of confidence, is as follows: complete confidence, 1; moderate confidence, 2; minimal confidence, 3; and no confidence, 4 (Psychiatric coding, n.d.).

Though no statistical data provided for the scale, it may be of particular research value for several reasons. Thus, Rubin and his colleagues (1966) consider these scales "available to reliable measurement." Moreover, the Coding scales of this type are relatively rare."

Bibliography:

Rubin, E. Z., Simson, C. B., & Tetwee, M. C. Emotionally handicapped children and the elementary school. Detroit: Wayne State University Press, 1966.

Rubin, E. Z. Psychiatric coding (n.d., ditto).

TEST DATA WORKSHEET

Name of Test	St. Louis Symptom Inventory
Author	Glidewell, J. C., Domke, Herbert R., Mensh, Ivan N., & Kantor, Mildred B.
Age	Validating data has been obtained only for white children in the third grade, though the inventory could be readily applied to early elementary school children.
Area (i.e. language development)	School Adjustment
Type of Measure	Symptom inventory with parents and/or teachers as informants.
Source from Which the Test May be Obtained	Journal of Educational Research, 1963, 56, 508-515.

Description of Measure:

"The information about the presenting symptoms of the child was collected through a private, home interview with the mother of the child. The mother was asked a series of questions about various characteristics of the family and the family background. Following this she was introduced to a series of questions which constitute the symptom inventory and take this form: 'Does Johnny have any trouble ....(eating, sleeping, getting along with other children, etc.)?' Affirmative responses were followed by probes into a) the specific nature of the difficulty, b) the frequency, c) the duration, and d) the severity. This line of questioning was continued through twenty-one areas of difficulty.... The number of symptoms is determined simply by counting the number of symptoms reported by the mother. The range of possibility extends from the minimum of reporting no presenting symptoms in the child to the maximum of twenty-one."

The major purpose of the inventory was ". . . to test the effectiveness with which a mother's symptom report can be used to identify, in advance, those children who would be referred by a teacher have at least sub-clinically disturbed." In other words, the inventory served as a screening instrument for emotional disturbance.

Evidence of Reliability, Validity, and Standardization:

A sample of the study was composed ". . . of 830 white families, each having at least one child in the third grade in one of fifteen schools located in St. Louis County. The area served by the schools varied widely with respect to social characteristics. The schools were selected on the basis of willingness of school administration to participate in the research program and on their ability to supply, from a single school, two third grade classrooms necessary for the research designed. The third grade was selected because at this grade level the children were expected to have relatively few transient symptoms of disturbance. In the third grade, most children have completed the initial behavioral adjustments to the move from school to home. At the same time, they have not yet begun to make the transition to pre-adolescence."

"For this study, ratings of general adjustment were made by the children's teachers

on a four-point rating scale. The adjustment categories were derived from those employed in an early study by Ullmann (14).

1. Well adjusted: A happy child who is well adjusted in his relationships with others and in his accomplishments.

2. No significant problems: A child who gets along reasonably well and accomplishes reasonably well the things that usually go with his age and level of development.

3. Sub-clinically disturbed: A child who is not so happy as he might be: has moderate difficulties getting on; growing up represents something of a struggle.

4. Clinically disturbed: A child who has, or, at his present rate is likely to have serious problems of adjustment, and needs clinical help because of such problems.

Six percent of the sample (n=49) were given ratings by psychiatric social workers based on professional findings. These independent ratings were exactly alike for 53% of the 49 cases and within one point in 92% of the cases. There was 80% agreement as to whether or not a child showed at least a mild disturbance. This same trend appeared in the previous study...., where it was found that the teachers' opinions about the emotional state of the children agreed exactly with the opinion of the professional mental health personnel in 86% of the cases" (Glidewell, Domke, and Kantor, 1963).

In brief, reliability data related to independent ratings by psychiatric social workers and an agreement by teachers and professional mental health personnel. A third source of reliability was the consistency of the mothers' reports. In this connection, a sample of 184 mothers had indicated about ". . . a seventy-percent agreement between first- and second-year reports" (Glidewell, et al., 1959, p. 24). But one note of precaution needs to be introduced; namely, the extent of the relationship between mothers' reports and teachers' ratings as one moved up the social class scale. Thus, these correlations were .70 for upper class, .33 for upper-middle, .20 for lower-middle, and .17 for lower-class. The criteria for social class determination were: 1) occupation of the head of the family; 2) education of the head of the family; 3) gross family income." In this larger study, the correlation for the total data was about .20, ". . . a significant shrinkage from the previous correlation of .50... (Glidewell, et al., 1959).

"In summary, upper-class mothers agreed well with teachers' ratings of both adjustment and disturbance, middle-class mothers agreed with teachers' ratings of disturbance but not of adjustment, and lower-class mothers agreed least with ratings of both adjustment and disturbance" (Glidewell, et al., 1959).

In a large sample (N of 830), analysis of variance yielded a significant difference between number of symptoms reported by parents and teachers' ratings of adjustment with p. 0001 (Glidewell, et al., 1963). There were no significant differences between sex groups and the number of symptoms reported by the mother and the analysis of variance "held up" separately for all social classes (Glidewell, et al., 1963).

On the basis of this evidence, the authors concluded ". . . that there is a reliable and positive relationship between the number of symptons reported by a child's mother and a degree of maladjustment reported by a child's teacher (Glidewell, et al., 1963).

A second source of validity received support from the predictive efficiency of the sympton inventory as a screening instrument for emotional disturbance. Three criteria affect this efficiency. First, the number of false negatives should be minimal; that is, the instrument should not miss selecting children who are actually disturbed. This criterion has also been labeled the "minimum missed-case rate." Secondly, the false positive rate should also be minimal; that is, the designation of children as disturbed who are not maladjusted. Thirdly, the number of children selected for further examination should be optimal; that is, adjusted to the diagnostic facilities available within the agency doing the screening. If the minimum missed-case rate is the principal criterion, the best cut-off point is between 0 and 1 sympton. The missed-case rate of the percentage of false negatives in this case is only 9% but the false positive rate is 82% and 84% of the children referred will require further study. If a minimum of false positive constitutes the main concern, the critical score should be seven or more symptons. Almost all (99%) of such children would be found to be disturbed but the rate of false negatives would be very high (88%). If one wishes to maximize the success in predicting both presence and absence of disturbance, the critical cut-off score should be near the middle of the range; that is, after three symptons (Glidewell, et al., 1963).

In short, the authors have summarized the validity data for this inventory as follows: "The question about the possible relationship between mother' reports and teachers' ratings is given as clear-cut an answer as correlational research can give. The existence of the relationship was demonstrated in a pilot study; it was confirmed in this large-sample study; and it has been confirmed on several subdivisions of the large sample. The several subdivisions confirm its existence separately for both boys and girls, and separately for four social classes. It appears quite clear that, at least for samples in St. Louis County, the greater the number of symptons the mother reports, the greater is the likelihood that the child will be rated disturbed by the teacher" (Glidewell, et al., 1959, pp. 126-127).

"With respect to the second question about screening effectiveness, the current findings are quite comparable to previous small sample findings...."

The current study gives 68% valid decisions at the cut off point of 3 symptons. "Altogether, these findings indicate screening efficiency better than many psychiatric screening techniques...." (Glidewell, et al., 1963).

Bibliography:

Gildea, Margaret; Domke, H.R.; Mensh, I.N.; Buchmueller, A.D.; Glidewell, J.C., & Kanter, M.B. Community mental health research: findings after three years. Amer. J. Psychiat., 1958, CXIV, 970-976.

Glidewell, J.C.; Mensh, I.M.; Gildea, Margaret. "Behavior symptons in children and degree of sickness." Amer. J. Psychiat., 1957, 114, 47-53.

Glidewell, J.C., Gildea, Margaret, C.L., Domke, H.R., & Kantor, M.B. Behavior symptons in children and adjustment in public school. Hum. Organization, 1959, 18, 123-130.

Glidewell, J.C., Domke, H.R., & Kantor, M.B. Screening in schools for behavior disorders: use of mothers' report of symptons. J. educ. Res., 1963, 56, 508-515.

Kantor, M.B. Some consequences of residential and social mobility for adjustment of children in Mildred B. Kantor (Ed.). Mobility and Mental Health. Springfield, Illinois: Charles C. Thomas, 1965, pp. 86-122.

TEST DATA WORKSHEET

Name of Test — School TAT

Author — Mary Engel, Ph.D.

Age — Elementary school age

Area (i.e. language development personality

Type of measure — Test--individual-projective

Source from which the test may be obtained

Description of Measure:
This is a picture-story test in which all picture-card stimuli are school-oriented, showing school-age children in various ambiguous situations. We constructed a conceptual continuum, representing more or less mature ways of coming to terms with the demands of the school and rated responses to the School TAT accordingly. The School TAT consists of five pictures, four of which have equivalent versions for boys and girls. They portray children in various situations that often occur in the course of a school day. They depict children on a school bus; youngsters in an art class, with one child apparently yelling or yawning; a child at the blackboard with an adult beside him; a youngster racing down a school corridor with a clock overhead; and a classroom scene with all children but one raising their hands." The examiner classifies the stories into the following categories: Rebellion, Platitude, Constriction, Anxiety, Punishment, Magic Solution, Never Again, Problem Solving, and Fanciful. These categories are described briefly.

Evidence of reliability, validity, and useability:
"Inter-rater reliability is r=.81, based on the independent rating of 126 stories from pilot subjects by the investigators and a psychologist not associated with the project. Two judges not associated with the project replicate the ratings of these data, working from written definitions of the categories, to the extent of r=.83."

Bibliography of test usage

Engel, Mary, Cohen, Roberta B., and Sanfilippo, Janet. Children tell stories about school—an exploration of the differences in the elementary grades. Paper read at Amer. Psychol. Ass., Los Angeles, California, 1964.

TEST DATA WORKSHEET

Name of Test — The Structured Interview Technique

Author — Pauline G. Vorhaus

Age — Upper Elementary to Adult

Area (i.e. language development) — Personality-general

Type of measure — Structured Interview

Source from which the test may be obtained — See bibliography

Description of Measure

While this measure is not specifically set up for use with children, it and similar structured interview techniques following a short picture projective measure are used by numerous psychologists in working with children.

"The procedure which comprises The Structured Interview may be summarized as follows:

1. The subject is asked to draw a person. In the unusual instances where the drawing is of the opposite sex, he is then asked to make an own-sex drawing.
2. After he has completed his own-sex drawing, he is asked the series of questions given below concerning the own-sex figure he has drawn. These responses are recorded and numbered on the left side of a vertically folded paper; this being Part I of the Structured Interview.
3. He is then asked the same questions about himself as he was first asked about the figure he had drawn. These responses are recorded and numbered on the right side of the same vertically folded paper; comprising Part II of the Structured Interview."

This measure consists of the 32 structured interview questions so called, although some of them are in fact, areas of inquiry, for example, relationship to parents, home attitudes, health, appetite, etc. Under "health" there are specific questions such as, "Is he healthy? If not what is wrong? How was his health in childhood?" In general, the questions are designed to bring out the subject's own needs, the pressures impinging on him, his responses to specific types of usually distressful situations. There is no scoring scheme provided for the measure. It is designed to be interpreted clinically, although it would be possible for the researcher to impose a scoring scheme on most of the items.

Evidence of reliability, validity, and useability

There is no evidence given for test validity or reliability. There is reference made to the Thematic Apperception Test, with the suggestion that, like the TAT, this measure gives "good leads or working hypotheses to be verified by other methods." (1) The discussion of reliability, like that of validation, refers to Murray's Thematic Apperception Manual, and suggests that the same statements of validity and reliability as are made for the TAT can be made for this measure. Murray is then quoted as stating, "... we should not expect the repeat reliability to be high..."

Bibliography

Harrower, Molly; Vorhaus, Pauline; Roamn, Melvin; and Bauman, Gerald. Creative Variations in the Projective Techniques: Charles C. Thomas, Springfield, Illinois, 1960.

TEST DATA WORKSHEET

Name of Test: Teacher Rating Scale (Rubin)

Author: Eli Rubin

Age: Kindergarten through second grade

Area (i.e. language development): Social and emotional adjustment in classroom

Type of measure: Teacher rating scale

Source from which the test may be obtained: Author

Description of Measure:
"A Teacher's Rating Scale was devised to give a rating of the social and emotional adjustment characteristic of . . (the emotionally disturbed) child in the classroom as seen by the teacher ... This . ... (scale) assessed a variety of behaviors, "including distractibility, aggressiveness, anxiety, tension-seeking, curiosity, communication with others, and action to criticism.... Assessment of these characteristics was made under two conditions for the majority of items: in the teacher's fear, and when engaged in independent activity For a few items, observations were made of behavior outside the classroom.

The items below illustrate the scales and the scoring system used (Rubin, et al., 1966, pp. 64-65).

"CODING SAMPLE --TEACHER'S RATING SCALE OF ADJUSTMENT
Scale 13
1. Child very seldom gets involved in disputes, quarrels, or fights with other children
3.
5. Child quarrels and fights with other children as much as is expected of his age.
7.
9. Child's contacts with others very often results in arguments, quarreling, fighting (regardless of who started it)."

Evidence of Reliability, Validity, and Standardization:
"On the TeacherRating Scales, interrater reliability coefficients were obtained for 79 scales and range from .26 to .94 (Rubin, 1966). "Sixty-six per cent of the interrater correlations calculated on the ... scores from this instrument were above .75. With exception of four scales -- ... 'affect of criticism by the teacher; ' . . 'persistence in independent activity; ' . 'joining aggressive activities in the teacher's fear; ... and ' emotional reactions to criticism from the children; ' --where the correlations were below .50, this scale can be considered to provide a consistent measure of classroom adjustment" (Rubin, et al., 1966, p. 65). If the items would show the lowest reliability were eliminated, "the coefficients then range from .60 to .94 with a mean of .78 (Rubin, 1966).

As regards validity, Rubin (1966) has contended that the scale "appeared to have considerable value in the evaluation of change in ...(his) study" (Rubin, 1966).

Reference to change refers to improvement in social and emotional adjustment as a result of placement in a special class for emotionally disturbed children. "A total of 151 children from kindergarten through second grade were referred to the Wyandotte (Michigan) special class project as candidates for placement in the special class program... " (Rubin, 1966, p. 48).

Bibliography:

Rubin, E. Z., Simpson, C.B., & Beetwee, M.C. Emotionally handicapped children in the elementary school. Detroit, Michigan: Wayne State University Press, 1966.

Rubin, E. Z. Personal communication, Nov. 9, 1966.

TEST DATA WORKSHEET

Name of Test	Teacher's Rating Scale
Author	Sarason, Seymour, et al.
Age	Elementary School Children
Area (i.e. language development)	Personality-Anxiety
Type of measure	Rating Scales
Sources from which the test may be obtained	See bibliography

Description of Measure

This is a rating scale to be used by teachers, designed to measure anxiety in elementary school children. Each of the 17 items is rated on the five point scale, in which a score of one means that the behavior described in each item occurs "very often." (A child is very often like this; it is very characteristic of him.) A score of two means that the behavior occurs fairly often (the child is frequently like this; it is fairly characteristic of him.) A score of five means that the behavior almost never occurs. (The child is hardly ever like this; it is not characteristic of him.)

Examples of the items on which the child is rated using the afore mentioned scale are as follows: "1. Does the child perform less well in school than your evaluation of his intelligence would lead you to expect? 2. Does the child's voice tremble when he is asked to recite? 3. Does the child become upset when he is told that the answer which he has given is wrong? 4. Does the child's illness or physical complaints tend to coincide with test days or class recitations?"

Evidence of reliability, validity, and standardization

Sarason (1) correlated scores on the Test Anxiety Scale for Children and the Teacher Rating Scale and found that the correlations ranged from .09 to .34. Most of the correlations were significant at or beyond the .05 level. Sarason (1) further found fairly high negative correlations between teacher rating scores and mean achievement, with a strong tendency for the magnitude of the correlations to drop as grade level increased. It is also a negative but low correlation between teachers' ratings and I.Q.

Bibliography

Sarason, Seymour B., et. al. Anxiety in Elementary School Children. Yale University, 1960.

TEST DATA WORKSHEET

Name of Test	Toronto Infant Security Scale
Author	Dr. Betty Margaret Flint
Age	Birth to two years of age
Area (i.e. language development)	Mental health of infants
Type of Measure	". . . a diagonostic check-list which attempts to assess the mental health (security) of infants...." (1959, p. 27).
Source from Which the Test may be Obtained	University of Toronto Press

Description of Measure:
"Because the scales attempt to measure the 'usual' state of a child rather than how he is at the moment, reliance on parental evaluation is necessarily heavy" (1959, p. 38). As in the Vineland, then, a parent as an informant is necessary in order to administer the test. Scores are available for the acceptance of dependence (D) and the acceptance of effort on the secure items (E) and the rejection of dependents (D) and effort (E) on avoidance mechanisms (Deputy agent items and regressive items). Endorsements in each category are expressed in percentages by multiplying the proportion of test items (endorsed) for the total number of applicable items by 100 (1959, p. 36). The scale itself is based on the theory that the truly secure baby is indicated by the items reflecting dependence on adults and effort in areas where the child has sufficient skill to help himself (1959, pp. 30-31).

"The items are grouped according to four age levels — 0-6 months, 6-12 months, 12-18 months, and 18-24 months — and, in effect, constitute four scales covering the total age range" (1959, p. 27). Directions for scoring include such matters as a child's lack of experience and uncertainties in the judgment of the commission concerning the applicability of items. Directions are also available on the use of the findings from the scale for guidance. In view of the complexities in administering and scoring the scale, it would appear that its use should be confined to trained clinicians." The information collected by a person who either trained in its use and familiar with the "Security Theory", or by a person experienced in the observation of the infant behavior who has attempted to understand the theoretical background of the scale" (Manual).

"Ordinarily a test should take half an hour" (Manual).

Evidence of Reliability, Validity, and Useability:
Mean and median scores were calculated for poorly adjusted and well-adjusted children on each of the four scales. The graphic evidence indicates that the median scores for well-adjusted infants on all four scales were considerably higher than those for the poorly adjusted infants (1959, pp. 44-45). The N for each group was twenty (well-adjusted infants) and thirteen for the poorly adjusted children (1959, p. 41). The poorly adjusted children were living in foster homes and the well-adjusted with their parents. In the process of developing the items for the scale

the author used the following sequence. Records of factual behavior for 150 infants under care of a local agency were examined. Descriptions of behavior were then listed according to their reflection of security or insecurity. Subsequently agreement with the senior psychologist at the clinic was obtained with respect to the categorization of the majority of items. Lastly, items for the final revision were selected on the following basis: (1) initial adjustment to a new or changing situation, (2) reflections of security, and (3) indications of insecurity. Reliability coefficients of the four scales were based on the Pearson-r correlation for test-retest scores from the groups as a whole irrespective of clinical class identification. These coefficients range from .46-.91. The author concludes from these results that ". . . the test-retest reliability of the Infant Security Scale is satisfactory and, with the exception of level 4, test 2 increased with age" (1949, p. 45). The author contends that consistency of results on the same children over time should not be expected in view of their intense responsiveness to their environment. At present she is attempting to obtain data from interrelative-reliability (1966).

Bibliography of Test Usage:

Flint, Betty M. The Security of Infants. Toronto: University of Toronto Press, 1959.
Flint, Betty M. Manual For Infant Security Scale. Unpublished material (ditto).
Personal communication. March 30, 1966.

TEST DATA WORKSHEET

Name of Test	Zeligs' Annoyances Test
Author	Zeligs, Rose
Age	Sixth grade
Area (i.e. language development)	Sources of annoyance
Type of measure	Questionnaire
Source from which the test may be obtained	See Zeligs (1966).

Description of Measure:
This test is designed to identify the personal, social, and environmental "factors" annoying to children. The author suggests (Zeligs, 1945) indirectly that the test may measure adjustment or irritability.

The following material from Zelig (1966) gives information on the administration and scoring of the measures. "DIRECTIONS FOR PREPARING AND ADMINISTERING ZELIGS' ANNOYANCES TESTS

TEST 1. PERSONAL FACTORS ANNOYING TO CHILDREN
List items found in Journal of Genetic Psychology, 1962, 101, pages 256-259 for the boys and 260-265 for the girls. (from tables) title: Children's Attitudes Toward Annoyances

Like Don't Mind Don't like Hate Hate much

On top of the page have child give:

Name Grade Boy or Girl Date School

Then write:
DIRECTIONS: We would like to know how children feel about certain things. Put a circle around the word or words that tell best how you feel about each thing on the list. If you like the thing mentioned put a circle around the word like; if you don't care either way, put a circle around don't mind; If you don't like the thing mentioned, put a circle around the words don't like; if you hate the thing mentioned, put a circle around the word hate; and if you hate much the thing mentioned put a circle around the words hate much. Answer everything on the list. Be sure to give your true feelings.

TEST 2. SOCIAL FACTORS ANNOYING TO CHILDREN
List items found in tables in Journal of Applied Psychology, 1945, 29, 75-82, making separate tests for boys and for girls. Proceed as for Test 1, above.

TEST 3. ENVIRONMENTAL FACTORS ANNOYING TO CHILDREN
List items found in tables in Sociology & Social Research, 1941, 25, 549-556., making separate tests for boys and for girls. Proceed as for Test 1, above.

The three tests can be administrated at the same time as it does not take long to encircle the words, and the faster they answer the more reliable the results will be.

TO SCORE. Add total number of items encircled under each category of feeling. Multiply the number under like by 2; under don't mind by 4; under don't like by 6; under hate by 8; and under hate much by 10. Add the totals and divide by 10 to get the score.

The total score divided by the number of items in the test gives the score in per cents. The total score for the entire test of our study of all categories was 56.4% for the boys, 60.4% for the girls, and 58.4% for all subjects. The higher the score the more easily annoyed the subject is.

The score for an individual is the per cent of the highest possible scores i.e. if a subject marks hate much for every item his score would be the total number of items and the per cent would be 100 per cent" (Zeligs).

Evidence of reliability, validity, and standardization:

Some standardization figures are given for sixth grade boys and girls in the articles cited by Zeligs.

Bibliography:

Zeligs, Rose. Environmental factors annoying to children. Sociol. soc. Res., 1941, 25, 549-556.

Zeligs, Rose. Social factors annoying to children. J. Appl. Psychol., 1945, 29, 75-82.

Zeligs, Rose. Children's attitudes toward annoyances. J. Appl. Psychol., 1962, 101, 255-266.

Zeligs, Rose. Directions for preparing and administering Zeligs' annoyances test. Personal communication, 1966.

MEASURES OF ATTITUDES

TEST DATA WORKSHEET

Name of Test	Bronfenbrenner Parent Behavior Questionnaire (BPBQ)
Author	Urie Bronfenbrenner, Professor of Psychology Cornell University, Ithaca, New York
Age	Fourth through sixth grade
Area (i.e. language development)	Attitude Scale (child's perception of how his parents treat him).
Type of measure	Questionnaire
Source from which the test may be obtained	Siegelman, 1965

Description of Measure:
The BPBQ requires that the child answer questions about how his parents act towards him. The questionnaire apparently has particular research value from the phenomenological standpoint since this view stresses that only the child's perception of his parent behavior toward him affects his personality development.

"The BPBQ consists of 45 statements concerning parental behavior. The same 45 statements are used for mother and for father. The children are asked to indicate the extent to which the statements in the questionnaire are true of how their parents act toward them. The subject selects one of the following five choices for each of the first 25 items; in every case, in most cases, sometimes, seldom, never. The last 20 statements have the choices of: almost, every day, about once a week, about once a month, only once or twice a year, never. The scoring ranges from 1 (in every case or almost every day) to 5 (never), so that a low final score indicates affirmation of certain parental behavior and a high score denial of some parental tendency. Fifteen variables are purportedly measured by the BPBQ, and there are three statements for each variable. The range of scores for a single variable, therefore, is 3 to 15. The three items for any one variable are separated from each other by seven or eight statements describing the remaining variables."

Examples of three variables with their representative statements are given below:
"1. Nurturance: I can talk with her (him) about everything. Comforts me and helps me when I have troubles. Is there for me when I need her (him).
2. Affective Reward: Says nice things about me to other people. Is very affectionate with me. Praises me when I have done something good.
3. Instrumental Companionship: Teachers me things which I want to learn. Helps me with hobbies or handiwork. Helps me with schoolwork when I don't understand something" (Siegelman, 1965, p. 164-165).

Evidence of reliability, validity, and standardization:
'Fourth, fifth, and sixth-grade boys (N 81) and girls (N 131), from a low socioeconomic area in New York City completed the BPBQ. The fifth- and sixth grade classes were in one elementary school, and the fourth-grade classes were in a

second elementary school nearby. Approximately 65 per cent of the samples were Puerto-Rican, 25 per cent Negro, and 10 per cent Italian. The children in the fifth and sixth grades were in classes ranging from the 'brightest' to the 'slowest' in each grade."

Based on generalized Kuder-Richardson Formula 20 reliabilities. The following data have been reported by Siegelman: "The reliabilities range from .26 to .83 for male fathers, from .23 to .70 for male mother, from .55 to .88 for female father, and from .32 to .75 for female mother. The mean reliabilities for all BPBQ scales are .58 for male father, .45 for male mother, .68 for female father, and .51 for female mother. The numerous low reliabilities are not surprising since each scale is composed of only three items. The reliabilities for female father, in this respect, are unusually high."

In addition factor analysis yielded data on the internal consistency of the scales. "The first three unrotated factors accounted for 62 per cent of the total variance of male fathers, 54 per cent for male mothers, 57 per cent for female fathers, and 50 per cent for female mothers. Each of the remaining three factors considered for boys and girls accounted for less than nine per cent of the total variance."

These BPBQ factors were those of loving, punishment, (with strong suggestions of rejection), and demanding...." Some validity for these factors may be suggested by the fact that increasing numbers of studies ... suggest that loving, rejecting, demanding and antonomous ways of acting represent fundamental and vivacious qualities of parental behavior." The mean of these factors have been described in full by the author. The author has also suggested a method of merging scales to increase the reliability of a measurement on the BPBQ. "On the basis of the rotated factors, the scales with high-factor saturations for a given factor were combined to form a single score. The single 'factor score' combined raw scores for the variables with high-factor loadings for each subject." The details of the procedure as well as the statistical justification or rational have been presented in full. In any case, "factor-score reliabilities based on the merged range from .70 to .91.... these internal-consistency coefficients are generally higher than the reliabilities of the individual scales. The single-scale reliability of 82 per cent were below .70. The number of items used to calculate the factor score reliability ranged from 12 to 18, contrasted to three items associated with each individual scale."

Bibliography:

Siegelman, M. Evaluation of Bronfenbrenner's questionnaire for children concerning parental behavior. Child Development, 1965, 36, 163-174.

TEST DATA WORKSHEET

Name of Test	A Child Attitude Inventory for Problem Solving (Crutchfield, Covington)
Author	Richard S. Crutchfield & Martin V. Covington
Age	Fifth and sixth grades
Area (i.e. language development)	Attitude scale
Type of measure	Group administered paper and pencil inventory
Source from which the test may be obtained	

Description of Measure

"CAPS consists of two scales. Scale I is designed to indicate the child's beliefs about the nature of the problem-solving process and his attitudes toward certain crucial aspects of problem solving, such as the expression of novel or unusual ideas. A number of related themes are treated, including the child's conception of the inneteness or unchangeability of one's problem-solving ability, the desirability of supressing rather than expressing novel ideas, the wisdom of persisting in the face of a problem that others have failed to solve, and the value of generating many ideas. The scale consists of 30 such true-false items.

"Scale II, also consisting of 30 true-false items, is intended to assess the child's feelings about his own ability to succeed in problem-solving situations. Questions concerning some of the typical sources of childhood anxiety about thinking are represented. This includes the fear of having one's ideas held up for ridicule, anxiety about not understanding how to go about solving problems, and the fear that one is not capable of effective thought" (Covington, n.d.).

Examples from each scale follows:

Scale I

"1. Yes No A problem like the one about the TV cable and the pipe is probably too hard for anyone in the 5th grade to solve.

2. Yes No There is probably only one answer to a problem like this one."

Scale II

"Now we would like to know how you, yourself, might feel when working on problems like the one of getting the TV cable through the pipe. Remember, these questions are about how you think and feel, so there are no right or wrong answers. Circle the word "Yes" if you agree with the question. Circle the word "No" if you disagree. Answer every one, even if it seems hard to decide.

1. Yes No Do you feel that other children in class know more about what to do in working on a problem like this than you do?

2. Yes No Would you like to work on a problem like this one?" (Thinking about Thinking, n.d.).

Potential uses CAPS have also been discussed in detail (Covington, n.d.).

Evidence of reliability, validity, and standardization

An initial form of CAPS consisting of 40 true-false items in Scale I and 25 true-false in Scale II, was "tried out on a total of 123 fifth-grade and 65 sixth-grade Ss..." The current scales were developed on a basis of item analyses of the initial forms. "The present form of CAPS was administered on two separate occasions to seven fifth-grade and five sixth-grade classrooms from the...public schools, representing a total of 325 children from four different elementary schools. A five week period intervened between the two administrations." Means, standard deviations, reliability coefficients, have been classified according to grade level with sexes combined. For grade five the reliability coefficient on the two test administrations were respectively .70 and .65 for Scales I and II. In the sixth-grade, the corresponding coefficients were .67 and .64. A number of additional findings gives some support for the validity for the CAPS.

1. On the basis of item analyses on an earlier form of CAPS, the present version of the scales includes highly differentiating revised items.

2. A positive but modest correlation between the two subscales (r of .35 for 325 Ss) indicates that the amount of overlap is minimal and that the two tests measure independent functions.

3. The scale scores show a significant and positive correlation with the IQ results on the California Mental Maturity Scale (r of .33, $p < .01$) but negative correlations exist between these data and two anxiety measures by Sarason--Tests Anxiety for Children and General Anxiety Scale for Children.

Bibliography

1. Covington, M. V., A child attitude inventory for problem solving, (n.d., ditto).
2. Crutchfield, R. S., & Covington, M. V. Thinking about Thinking, (monolith, n.d.).

TEST DATA WORKSHEET

Name of Test	The G-W Method of Paired and Projective Questionnaire (PDPQ)
Author	J.W. Getzels & J.J. Walsh
Age	8-13, whites and negroes
Area (i.e. language development)	Attitude structure and socialization
Type of measure	40 incomplete sentences
Source from which the test may be obtained	(See reference)

Description of Measure:
The authors provide empirical evidence that the method of Paired Direct and Projective Questionnaire can profitably by applied to the systematic investigation of overt and covert levels of attitude structure (Getzels and Walsh, 1958, pp. 1-15). For this purpose, ". . . a class of 48 freshmen and sophmores in a women's college in New England were utilized as the subjects for the experiment" (Getzels and Walsh, 1958, p. 7). The authors also investigated the development of levels of attitudes as a measure of socialization with variations as a function of such factors as age, sex, sibling position, and social class status as previously described. In this study, they conceived of socialization as a two-fold process. "On the one hand, the child learns to renounce, suppress, or redirect drives and behavioral impulses that are at variance with proper social standards. On the other, he learns to bring his observed behavior into conformity with the norms and expectations of the society in which he lives."

In other words, the socialized child has learned that "'in every day life it is often necessary to mask the true self, to present to the world the appearance of conventional fitness'" .... (Getzels and Walsh, 1958, p. 15). The principal measure of socialization was the Index of Differentiation which was computed by the formula

"ID=$\frac{P - D}{P}$ where P represents the number of negative completions on the projective questionnaire and D the same score on the direct questions. The sample utilized for the study of ID as a measure of socialization included 8 elementary schools, consisting of private and parochial institutions with a wide range of social classes, as well as 428 boys and 485 girls range in age from 8 to 13. In addition the breakdown of the sample reveals the following: "Negro children constitute 11% of the sample, and the sample is divided as to religious preference as follows: Catholic 42%; Jewish 6%; Protestant 47%; and 5% of the sample that admit no religious affiliation" (Getzels and Walsh, 1958, p. 17). The test itself consist of 40 incomplete sentences. In the direct form, the items are phrased in the first person. The projective technique phrases the sentences in the third person and the items are disguised as a verbal speed test. "The items are written

so as to conform the three major requirements: First, they must be of such a nature and structure as to facilitate a dichotomous classification of the responses, usually as positive (socially acceptable) or negative (socially unacceptable) .... Second, the items are to be free of affect-laden words that might exert undue influence in the nature of their response. Thus, words like 'good,' 'bad,' 'angry,' 'sad' — words that might demand uniformly positive or negative responses—are avoided. Third, the items should as far as possible present behavioral situations corresponding to areas identified in theoretical or empirical studies as having differential values for varying subsections of the population" (Getzels and Walsh, 1958, pp. 4-5). The projective and direct tests include separate forms for boys and girls. In administrating the test, the authors use the following procedures: "The projective instrument was administrated first, and the direct test was administrated after a two-week interval. At the end of the second test session the children completed a personal data sheet which provided data concerning age, sibling position, and other variables needed in the analysis of the data." (Getzels and Walsh, 1958, p. 18). These writers feel that the PDPQ meets a real need in attitudinal research because of a ". . . lack of an efficient technique for determining 'internal' and 'external' reactions to the same stimulus examining the relationship between them" (Getzels and Walsh, 1958, p. 1). The PDPQ bridges this gap between levels of reaction. The basic assumption that this ID measure this gap is". . . that when a subject is pressed to attribute behavior to others, and he perceives the situation as devoid of personal reference, he tends to reveal levels of behavior he might otherwise be disposed to distort or censor" (Getzels and Walsh, 1958, p. 3).

Finally examples of items on the direct and projective techniques follows:

". . . I think that most bosses are ....."

". . . Beverly thinks that most bosses are ....."

". . . Chuck thinks that most bosses are ....."

Evidence of reliability, validity, and useability:
The interscorer reliability ranges were respectively .97-.98 for the direct scores, .95-.96 for the projective scores, and .90-.95 for the discrepancy scores. With respect to interscorer reliability the authors state: "This problem was investigated by correlating the scores assigned to a random sample of 20 pairs of tests, each test containing 38 items, with the scores assigned to these tests by two other independent scores" (Getzels and Walsh, 1958, p. 5). The nature of the sample on which these scores were based is, however, not clear. The authors use the "Index of Differentiation" which is a measure of the discrepancy of negative responses on the direct and projective forms of the questionnaire study socialization. They predicted that if this Index of Differentiation (ID) represents the extent of socialization four hypotheses should receive confirmation: "1. The mean Index of Differentiation should increase as a direct function of age. 2. The mean Index of Differentiation should be greater for girls than for boys in the same age group. 3. The mean Index of Differentiation should be greater for the only child than for children with siblings. 4. The mean Index of Differentiation for children of the higher or middle socioeconomic classes should be significantly different from that of the lower socioeconomic classes." The empirical data varified these hypotheaized relationships in every case and high statistical levels of confidence. The authors therefore conclude: "This congruence between the empirical data provided by the

PDPQ and the predicted relationships derived from attitude and socialization theory attests to the usefulness of the techniques" (Getzels and Walsh, 1958, p. 25).

Bibliography of test usage:

Getzels, J.W., and J.J. Walsh. "The method of paired direct and projective questionnaire in the study of attitude structure and socialization. Psychological Monographs, 1958, 72, No. 1 (Whole No. 454).

TEST DATA WORKSHEET

Name of Test — Maternal Care Checklist

Author — Rheingold, Harriet

Age — Infancy

Area (i.e. language development) — Maternal care, mothering

Type of measure — Observation schedule

Source from which the test may be obtained — See Rheingold (1960) in bibliography

Description of Measure

This measure is described as "a method of measuring maternal care in terms of the actual operations the mother performs in caring for the infant... The method employs the technique of time-sampling maternal care as it occurs in its natural setting."

"On the basis of observation and trial...a checklist composed of 42 items was drawn up, 30 for mothering activities and 12 for infant activities... Of the 30 mothering items six recorded the location of S in his environment, four the number of people in his room, and four the number of people within six feet of him.

"The remaining 16 items recorded activities more closely related to mothering. 'Caretaking' was an all-inclusive item which was checked if at the moment of observation someone was doing something for S while in proximity to him. 'N of caretakers' recorded the number of different caretakers who cared for S during the eight-hour period of observation. The last 14 items were sufficient to cover all the caretaking operations performed by people in caring for three-month old infants at the level of analysis selected for study here."

The score is the frequency of occurrence of each item over a specifiedkperiod of time.

Evidence of reliability, validity, and standardization

The median percentage of inter-observer agreement in scoring all items was 89.8, with a range of 16.7 to 100. Observer agreement was also measured by correlating the frequencies obtained for each item by different observers. The median r was .97, range .35 to 1.00.

Evidence on validity of this observation frame of reference is given by Rheingold's comparison of the maternal care given to five home infants and five infants in an institution. Significant differences were found on 23 of the 42 items. The differences were in what most child researchers would consider the expected direction.

Bibliography

Rheingold, Harriet L. The measurement of maternal care. Child Development, 1960, 31, 565-575.

TEST DATA WORKSHEET

Name of Test: Maternal Values Questionnaire

Author: Cox, Henrietta

Age: Parents

Area (i.e. language development): Maternal values

Type of measure: Likert-type scale, forced choice scale, and ranked traits

Source from which the test may be obtained: See Cox, Henrietta in bibliography.

Description of Measure:
The author secribes this as a measuring instrument with "three quite independent parts." A likert-type scale consists of 5 subscales with names and numbers of items in each scale as follows:

Emotional Tone (demonstrativeness or non-demonstrativeness of affection).	
Autonomy-Control of the child's behavior.	6 items
Achievement-orientation in childhood, and adult life.	6 items
Mastery or Powerlessness over the environment.	6 items
Conceptions of Human Nature (trust or distrust in interpersonal relations.)	6 items
Discipline Practice (Not intended to constitute a scale).	5 items

In addition, three items relating to marital adjustment were included, but were also not intended to constitute a scale.

The items are taken from various other existing scales and combined into new ones as noted above.

The subject marks each item by circling either SA (strongly agree), A (agree), D(disagree), or SD (strongly disagree). Examples of items are given below.

1. Children should be allowed to learn through their own experience rather than being told what to do all the time.
9. Mothers ought to punish children who misbehave by taking away something they like.
10. A Child should be taught from infancy to take the greatest pride in doing things well.
19. Parents ought to show their children they love them by hugging, kissing, and playing with them often.
25. Nowadays the wise parent will teach the child to live for today and let tomorrow take care of itself.
31. Children should learn that most people can be trusted.

Part II, the forced-choice scale, parallels the Likert scale in content. This scale is given in the appendix of Cox's thesis. A trait list constituted the

third scale. These latter two measures were not used in Cox's thesis, so no data on reliability or validity are given.

Evidence of Reliability, Validity, and Standardisation:
Each of the five scales discriminated significantly among social classes, but none of the scales discriminate among races or religion, although some of the individual items do. Each of the 30 items (five scales, six items each) in the total scale is analyzed for discriminability between social classes, races and religions.

Bibliography:
Cox, Henrietta Sheppard. "Social Classes as Subcultures: Variations in Value-Orientations in Selected Areas of Mother-Child Behavior." Unpublished Master Thesis. Washington University, St. Louis, Missouri. June, 1964.

TEST DATA WORKSHEET

Name of Test: Parental Nurturance-Control Attitude Scale (Emmerich)

Author: Walter Emmerich

Age: Parents of preschool children

Area (i.e. language development): Attitudes

Type of measure: Open-ended questionnaire

Source from which the test may be obtained (Emmerich, 1959)

Description of Measure:
The Parental Nurturance-Control Attitude by Emmerich was based on a projective questionnaire consisting of eight hypothetical situations in which the parents were asked how they would deal with their child. The items contained in the self-administering parental questionnaire resembled those presented to the children in doll-play interview. For example, it's one was as follows:
"1. You and your child are in the food store and (she won't leave. She sees some candy she wants. What do you do?" (p. 303).
The parents' responses were then rated on a nurturance-control scale similar to those scales used to evaluate the children's doll-play interview responses, "except that the specific examples of the major scale points were expressed in more 'adult' terms." The scale, then, rated the parental responses with scores ranging form +3.0 (strong nurturance) to -3.0 (strong control). The description of the category strong nurturance is given below:

+3.0 Strong Nurturance. Parent facilitates child's accomplishment of a goal to a greater extent than the child requested.
1. Buys several toys
2. Reads for a long time
3. Makes special effort to prepare or secure food child wants
4. Plays with child or lets him play indefinitely
5. Allows child free reign in hitting dog
6. Gives child special privileges beyond caring for medical problem
7. Lets child stay up indefinitely
8. Encourages child to get dirty (p. 300).

Evidence of reliability, validity, and standardization:
Within the sample groups described for Emmerich's Child Nurturance-Control Scale, "Twenty-five... mothers and 23 fathers completed and returned the questionnaires." "The following procedure was used in sampling questionnaire items to be rated for reliability purposes: (a) 90 responses to the first five items were selected;

(b) the sample was stratified so that there would be 18 ratings made of each of the five items; (c) half of the ratings were of the mothers' responses, and half were of the fathers' responses; (d) the particular subjects for whom a particular item-parent pair would be rated was determined randomly. The correlation between two judges' independent ratings of a sample of items was found to be +.87. The responses to questionnaire Items 6, 7, and 8 were so frequently rated +2.0 on the nurturance-control scale that these items did not discriminate among parents. Consequently, the parents' nurturance-control scores were based upon the first five items only". (p. 280).

As an external criterion of validation, teachers' ratings of the mothers' nurturance-control toward their children were correlated with the sum of the mothers' nurturance-control scores on the first five questionnaire items. "The correlation for the 25 mothers who responded to the questionnaires were found to be +.51 (p. .01).

Bibliography:
Emmerich, W. Parent identification in young children. Genet. Psychol. Monogr., 1959, 60, 257-308.

TEST DATA WORKSHEET

Name of Test: Parental Practices Inventory

Author: Rue L. Cromwell

Age: Parents

Area (i.e. language development): Child rearing practices

Type of measure: Self-report and self-administering record from parents "... what works and what does not seem to work in the care, handling, and disciplining of ... [Children]."

Source from which the test may be obtained

Description of Measure:
In this self-report and self-administering inventory of child-rearing practices, parents are requested to complete every item with an answer that is closest to the truth. They are urged to describe actual practices which they have used in a specific situation. The inventory itself consists of items in six general areas. The areas, and a summarizing description of each, are given below.

I. IDENTIFYING INFORMATION. The parent is requested to answer such items as her relationship to the child, the number of children in the home, and the names and ordinal positions of the children. The entire section includes four items.

II. DESCRIPTION OF ALL CHILDREN IN THE HOME. Eight items comprised a section which requests the parent to list the ordinal position of the child with whom she did the best job in child rearing, the best adjusted child, the child with the closest relationship to her, etc.

III. DESCRIPTION OF MY PRACTICES AS A PARENT. This part of the inventory includes 11 items. In 11 behavioral situations, the parent is requested to select the best way of handling it for herself among six-alternative responses. A sample item follows:

"7. When we parents disagree, this child knows it
_____nearly always.
_____sometimes.
_____almost never (We try not to let him know)" (Inventory, p. 5).

IV. DESCRIPTION OF THIS CHILD'S EXPERIENCES. The emphasis of this description is to urge parents to respond "... in terms of what has been done during most of the child's life." The subscale consists of 12 items. Sample items are given below.

"1. The person who has done most of the raising of this child is
_____his mother.
_____his father.

_____his grandparents.
_____older children in the family.
_____a relative.
_____other. For example_____.

2. When this child is punished, the person punishing him is usually
_____angry and upset.
_____calm and rational" (Inventory, p. 6).

V. HOME PRACTICES. In this section, parents must answer 11 items relating to child-care practices. Not only must parents complete each blank but they are also directed to check off the alternative responses in terms of priority. As the directions state: "put a 1 for the best choice, a 2 for the second best choice, etc., until every blank is filled in." A sample item follows:

"When this child is punished physically (spanked, slapped, etc.) he
_____cries.
_____talked back.
_____clams up.
_____becomes timid or afraid.
_____behaves even worse.
_____tries to correct his action if possible.
_____says he is sorry, but doesn't mean it.
_____says he is sorry and seems to mean it.
_____seeks reassurance or love.
_____acts like he had never been punished at all" (Inventory, p. 9).

VI. HOME VALUES. The scale employs a force-technique in which parents must select the most appropriate response in every case from among 60 pairs of statements. As the directions state: "For each pair, put a check by the statement which is first in importance to you." A sample item is as follows:

"1. _____Having him get a good education.
_____Having him always be on his best behavior around company"
(Inventory, p. 12).

Following the six scales, the parent who completed the inventory is required to answer two questions which serve as a check on validity. On these questions, parents merely need to check off the most appropriate alternative. First, the parent is requested to react to the extent to which the questions and answers gave a true picture of her home and child from among three alternatives (very close, fairly close, not very close). Secondly, she is asked to indicate the proportion of answers on which she was completely objective and frank from among four choices (all, most, about half, and only a few).

In addition to the inventory itself, Cromwell has developed a code book for the inventory with complete directions on the scoring of the variables.

Evidence of reliability, validity, and useability:
Two validity questions which follow the inventory have already been described. Furthermore, the code book gives complete directions on the ascertainment of the protocols in general regarding their validity. A check on validity also is available in terms of the consistency for parental responses. To the reviewers knowledge, however, reliability data are absent; standardization on any research or normative group has not been undertaken; and further validity data are lacking.

In a word, the instrument is a research tool in its preliminary stages.

Availability of alternate forms, permissibility of duplication:

Bibliography of test usage:

Cromwell, Rue L. Parent Practices Inventory, Unpublished (Mimeographed). January, 1966.

TEST DATA WORKSHEET

Name of Test	Sex-Role Attitude Test (Srat) — Adult Form
Author	Leonard M. Lansky
Age	Adults
Area (i.e. language development)	Attitudes toward sex-role choices
Type of measure	Questionnaire
Source from which the test may be obtained	See McKay, 1964, in bibliography

Description of Measure:
The SRAT ". . . consists of 50 items each of which ask the parent to judge either a preschool boy's (girl's) or a parent's reaction to a preschool boy's (girl's) preference for or decision between a pair of objects, games, or activities" (Lansky, 1966).

The parent is given three "sets" which are as follows: Set 1: HAPPY, PLEASED, DELIGHTED, CONTENT, PROUD, Set 2: NEUTRAL or IN BETWEEN THE FEELINGS OF SET 1 and SET 3, that is, NEITHER HAPPY NOR UNHAPPY, NEITHER PLEASED OR ANGRY, etc.

Set 3: ANGRY, UNHAPPY, SAD, DISCONTENT, CROSS, DISAPPOINTED. Thus, to each item the parent responds with a choice of one of three points along a variable which may be described as something like satisfaction-dissatisfaction. For example Item 1 from the scale which parents would use to rate a daughter goes as follows:

If a girl had a choice between playing with a toy gun and playing with a doll, how would her mother feel if her daughter wanted to play with both the toy gun and the doll (circle one)?.....Set 1 Set 2 Set 3.
The parent then indicates his or her degree of pleasure or displeasure by circling one of these. There is a comparable item in the scale for evaluating parents attitude toward boys' sex-role choices. Each camparable item is similar except for the changing of the sex of the "actor." Items 2 and 3 of the scale are given below as further exa es of the types of items used. If a girl had a choice between dressing up as an Indian princess and dressing up as an Indian chief, how would her mother feel if her daughter wanted to dress up as an Indian princess (circle one)? Set 1 Set 2 Set 3
If a girl had a choice between playing with toy tools for building and playing with toy things for cooking, which would her mother like her daughter to play with the most (circle one)? building tools cooking utensils

"There were five pairs of sex-linked activities: (1) playing with a gun or doll; (2) dressing up as an Indian chief or princess; (3) playing with toy tools for building or toy things for cooking; (4) playing on swings with some boys or some girls his (her) age; (5) playing with a toy shaving kit or a toy cosmetic kit."

Ten variables are measured. Examples are "father's attitude toward a same sex choice, "mother's attitudes toward a cross-sex choice," and "father's permissiveness towards the boy's wish to sometimes enjoy a same-sex and other times an opposite-sex choice."

Each item is scored 0, 1, or 2, with 0 meaning that the parent responded favorably to the item.

Evidence of reliability, validity, and standardization:
No data are given on reliability of the measure. Some evidence on construct validity is provided by one of Lansky's studies. The researcher interested in using his measure should consult this paper (Lansky, 1966).

Bibliography:

Lansky, Leonard M. "The family structure also affects the model: II. sex-role attitudes in parents of preschool children." Revision of paper read at Am. Orthopsychiat. Assn., San Francisco, 1966.

McKay, G.H. "Some dimensions of sex-typing in kindergarten children: an exploratory study." Cambridge, Mass. Unpublished Ed.D. dissertation, 1964, Harvard University.

TEST DATA WORKSHEET

Name of Test	Teachers' Rating Scale of Parental Nurturance-Control (Emmerich)
Author	Walter Emmerich
Age	Teachers of preschool children
Area (i.e. language development)	Nurturance-control toward child by parents
Type of measure	Rating scale
Source from which the test may be obtained	(Emmerich, 1959)

Description of Measure

The Teachers' Rating Scale of Parental Nurturance-Control is comprised of two subscales--a scale for rating the mother's nurturance toward the child when he was dependent on her (as "the child's wanting the mother to play with him. . .") and the scale rating the mother's control of the child when he was deviant (as the child's "refusing to go home and continuing to play"). In both cases, the scores range from a +3.0 to a -3.0. Respective examples from these subscales are strong nurturance for a child when he is dependent on the mother and strong maternal control when a child is deviant:

> "+3 Strong Nurturance. Mother usually facilitates child's accomplishment of a goal to a greater extent than the child requested. e.g., mother volunteers to play with the child in several activities. . .
>
> "-3 Strong Control. Mother usually strongly interferes with child's goal, e.g., scolds or physically punishes the child for making the request." (p. 304).

Evidence or reliability, validity, and standardization:

Since five nursery schools were involved in the sample, presumably the same number of teachers rated the behavior, of the 25 mothers who responded to the parental questionnaire. The teachers' ratings comprising the scale, as previously noted, consisted of two subscales--the mother's nurturance-control toward the child when he is dependent and the same maternal behavior when he is deviant. The respective teacher agreement reliabilities for these subscales were +.62 and +.41. Moreover, the correlation between the sum of the teachers' scores on the two subscales with the sum of the mother's nurturance-control scales for the first five questionnaire items was +.51 ($p < .01$). Emmerich has indicated, however, that he did not consider the teacher agreement reliability to be very high.

Bibliography:

Emmerich, W. Parent identification in young children. Genet. Psychol. Monogr., 1959, 60, 257-308.

TEST DATA WORKSHEET

Name of Test — The Traditional Family Ideology Scale

Author — Phyllis Elaine Huffman

Age — Adults

Area (i.e. language development) — Family ideology

Type of measure — Likert-type scale

Source from which the test may be obtained — See Huffman, Phyllis in bibliography

Description of Measure

The Traditional Family Ideology (TFI) is composed of 40 items, divided into six sub-scales as follows:

Conventional values--12 items
Over-idealization of parents--6 items
Power-relationships--9 items
Pseudo-masculinity and femininity--8 items
Discipline--7 items
Rejection of id--6 items

Some items are included in more than one subscale. There are 40 different items, of which six are used in two subscales and one is used in three subscales. Examples of the items are as follows:

It helps the child in the long run if he is made to conform to his parents' ideas.

Women can be too bright for their own good.

It is only natural and right for each person to think that his family is better than any other.

The saying, "Mother knows best" still holds more than a grain of truth.

The unmarried mother is morally a greater failure than the unmarried father.

"Six choices were allowed: +3 for strong agreement, +2 for moderate agreement, +1 for slight agreement, -1 for slight disagreement, -2 for moderate disagreement, -3 for strong disagreement" (Huffman, 1950).

Evidence of reliability, validity, and standardization

"Reliability of the scale, based on odd-even correlation corrected by the Spearman-Brown formula, was .84" (Huffman, 1950, p. 32). Item means and an index of discrimination power are given for each item.

The TFI scale correlated .65 with the Adorno ethnocentrism scale and .73 with the Adorno F (Facism) scale when given to a group of 109 students in adult evening psychology classes.

There are suggested relationships between TFI scores and religious preference and attendance at church. Persons whose religion and politics were different from those of their parents tended strongly to have lower TFI scores than subjects

having the same religion as their parents. There was no relationship apparent of TFI with the subject's political affiliation. The relationship between TFI and occupation was suggestive. Additional data are given by Huffman on TFI scores and projective question responses.

Bibliography

Adorno, T. W. and Others. The Authoritarian Personality, New York, Harper and Brothers, 1950.

Huffman, Phyllis Elaine. Authoritarian personality and family ideology: A scale for the measurement of traditional family ideology. Unpublished masters thesis, Western Reserve University, 1950.

TEST DATA WORKSHEET

Name of Test — Unnamed--A measure of Piaget's "reciprocity" concept

Author — Hartley, E. L., and Gondor, E. I.

Age — 5-14

Area (i.e. language development — Reciprocity (maturity is social perspective)

Type of measure — Test--individual (Picture test with interrogation)

Source from which the test may be obtained — Unpublished--see bibliography

Description of Measure:
The examiner presents a series of six pictures to the subject and then proceeds with the interrogation, which for FIGURE I goes as follows:

Three men were arguing. (Pointing to them in turn) One was American, one was English, one was Chinese. They were talking about flags; which is the prettiest flag. (Pointing) This is the Chinese flag, this is the American flag, and this is the English flag. The three men were Chinese, American, and English. (A) Now what do you think they said? (B) And who was right? (C) Why?

Evidence of reliability, validity, and useability:
Using subjects in grades K-8, Hartley and Gondor founded a progression from Stage I to Stage II to Stage III as the children become older. Stage I "involves" absolute judgment, no sensitivity to relativity of values;" Stage II involves "symmetrical" responses, with basic sociocentrism maintained; Stage III "involves a recognition that others may be motivated as we are and derive value judgments from a perspective which differs from ours." Examples are given of responses which are typical of children in each of the Stages, I, II, and III. These examples are usable as general scoring criteria.

Bibliography of test usage

Hartley, E.L. and Gondor, E.I. The development of children's ideas about their community. Unpublished manuscript, New York Research Council of the City College of New York, 1957

MEASURES OF SOCIAL INTERACTION AND SKILLS

TEST DATA WORKSHEET

Name of Test	Check Sheet of Opportunity in Human Relations
Author	Cunningham, Ruth and Associates
Age	Upper elementary and older
Area (i.e. language development)	Opportunity to participate in human relations.
Type of measure	Questionnaire
Source from which the test may be obtained	See Cunningham, Ruth and Associates in Bibliography

Description of Measure:
"The major purposes of this Check Sheet . . . are to discover the range of activities in which an individual engages, and the values he places on them. There are also opportunities for the individual to indicate his opinion of the degree of adult control, to show to whom he would go for help with personal problems, and to give his opinion of himself as a group member. The responses on this instrument were helpful in gaining insight into the relation between children's experience and individual differences, age, sex, and between opportunity for an experience and its value." The measure consists of 40 activities which the subject is to evaluate in two ways: 1. How often he participates in these activities, and he checks a column to indicate either frequently (almost every day), sometimes (once a week or so), seldom or never (not more than once a month), and ?. For the same item, the subject also checks how valuable he perceives the experience to be for him, and checks it as either important, O.K.-good but not important, unimportant, or ?. Samples of items are the following:

Item 1: Be a member of a team for some sport.
Item 4: Work or play with people of different religious beliefs.
Item 8: Work or play with people who have considerably less money than my family.
Item 17: Meet people who come from outside my community.
Item 24: Discuss boy-girl relations.
Item 31: Work for pay.

The scale consists of five additional items asking the subject to whom he would go in the family with a personal problem and to whom he would go outside the family with such a problem, how he feels about adult control in his family, and in school, one item having to do with self evaluation which he checks words or phrases which he feels describes him, one item in which he describes the three things he wants most to improve about himself, and finally he lists the three best friends in his room, home room, group, or section at school. The reading level for this measure is appropriate for the more able upper elementary children and beyond.

Evidence of reliability, validity, and standardization:

None.

Bibliography:

Cunningham, Ruth. Understanding group behavior of boys and girls. New York: Teachers College, Columbia University, 1951.

TEST DATA WORKSHEET

Name of Test: Classroom Social Distance Scale

Author: Cunningham, Ruth, and Associates

Age: Upper elementary

Area (i.e. language development): Social acceptance-rejection

Type of measure: Nomination questionnaire

Source from which the test may be obtained: See Cunningham (1951) in bibliography.

Description of Measure:
The instructions and method for using this scale are simple and straight forward. Each child rates every other child in the room on a five point scale as follows:

"1. Would like to have him as one of my best friends.
2. Would like to have him in my group but not as a close friend.
3. Would like to be with him once in a while but not often or for long at a time.
4. Don't mind his being in our room but I don't want to have anything to do with him.
5. Wish he weren't in our room."

Cunningham (1951, p. 171-72) states, regarding this scale, "It allows for a reaction, on a five-point scale, of each youngster to every other in the group. This instrument is most useful, in our opinion, in studying responses of each child, when it is used by a teacher who knows his group. However, it is possible, by assigning numerical values to the five items on the scale, to arrive at two types of social distance scores: one, a self-social-distance score indicating the degree of acceptance or rejection of the group by an individual; and two, a group-social-distance score, indicating the degree of acceptance or rejection of an individual by the group."

Evidence of reliability, validity, and standardization:
Cunningham reports correlations between scores for group-social distance (how other rated him) and chronological age, I.Q., and socio-economic status. She also reports correlations between self-social-distance (how he rated others) and chronological age, I.Q., and socio-economic status. The subjects were 32 fourth and fifth grade children, and the results were as follows:

Factors Correlated	Correlation
"Group-social-distance and	
Chronological age	+.036
Intelligence quotient	+.480
Socio-economic status	-.337
Self-social-distance and	
Chronological age	-.345
Intelligence quotient	-.026
Socio-economic status	-.09"

Cunningham observes that the child with a high-group social distance score tended to be the child who on the basis of her Social Analysis of the Classroom (described elsewhere in this book) could be described as follows:

"He is aggressive and displays a high degree of initiative, especially in suggesting interesting things for the group to do; he is cheerful, jolly, enjoying everything he does; he sees a joke easily and can laugh even when the joke is on him; and he is friendly and goes out of his way to be nice to his colleagues.

By the same token, someone who is judged to be timid, untidy, and afraid, is frequently sad or unhappy; who is unfriendly or too bashful and shy, will tend, in this group, to receive a low score on the Classroom Social Distance Scale."

<u>Bibliography</u>:

Cunningham, Ruth. Understanding group behavior of boys and girls. New York: Teachers College, Columbia University, 1951.

TEST DATA WORKSHEET

Name of Test	Hackett's Sociometric Technique
Author	Wally Reichenberg-Hackett
Age	Preschool children
Area 9i.e. language development)	Social interaction in classrooms
Type of measure	Sociometric technique
Source from which the test may be obtained	(Hacket, 1963)

Description of Measure:

Hackett's Sociometric Technique was based on behavioral observations of the child in play sessions entirely. Its purpose was to reveal the child's social status within the group. As an expository device, it may delineated in terms of an external approach, and the specific ingredients of the method.

General Description

The measuring instrument rested on the premise that a child's "sociometric choices could be determined through systematic observations and analyses of play behavior. On the basis of recorded observations, a sociogram consisting of a graphic representation of a child's social contacts within a given time could be constructed. The graph brought into focus a child's interactions, his relations with the group or any child in the group, and the type of social contact he initiated or invited. This sociogram illustrated his status and permitted quantitative asswell as qualitative analysis of social behavior."

The Method

The essential ingredients of the method included the pertinent observations recorded in the record, the units of behavior to be described, the type of contacts the child made, and the categories employed to classify the various types of behavior exhibited by the child.

Pertinent Observations

"The pertinent observations were made on three consecutive mornings during 'free play' periods. . . . Each observation period lasted 45 minutes. The observer followed the child throughout this period, using shorthand to describe all social contacts, and recording verbatim all verbal exchanges. Monologues, singing, and laughter were included in the record. All three reports, compiled by the same observer, were then combined into one running account of two and a quarter hours of observation."

Units of Behavior

The running report was analyzed in terms of "episodes." "An episode represented the smallest discernible unit of action. . . . For purposes of the sociogram, however, only those episodes which were relevant to a social situation or an

change in a social situation were designated with consecutive Arabic numbers. When, for example, a child spoke to another child requesting a toy, this was considered an episode."

Types of Contact

Differentiating recordings were made for "'outgoing' contacts, as when the subject approached another child, and 'incoming' contacts, those initiated by the other person." A third recording was made for verbal and physical contacts.

Categories of Behavior

From the materials in the running accounts of the child's interactions, categories for various types of behavior were developed. The development of these categories preceded according to certain sequential steps as follows:

1. "As a first step, the various types of outgoing and incoming contacts, child or adult, were noted."

2. " The second step pertained to the subject's communications. These verbalizations were categorized under several headings." For example, the provision of information by the child served as one category: "One hundred miles means how fast you can go.. . . I am building a garage for this truck".

3. "The third step consisted of an appraisal and summary of incidents typical of different forms of the child's social behavior within the play situation." The complete and detailed description of this appraisal system has been presented. In general, the seven behavioral systems employed for the child's social behavior were as follows: (I self-assertiveness, (II) aggressive behavior, (III) hostility, (IV) submissive, (V) domineering, (VI) cooperative, and (VIII) sympathetic.

Each behavioral unit included a definition and behavioral examples form actual social interactions between children. Illustrations from both aspects of the behavioral unit are presented below :

"7. SELF-ASSERTIVENESS
(to be clearly distinguished from aggression)

The establishment of self-identy through free expression of feelings, desires, opinions. The child feels sure of himself, he knows his rights and feels free to let others know about this, expresses himself naturally and directly. This may include commands and demands.

Examples:

The children had finished rest period and folded their rugs. A student reached down to get a child's rug and fold it for her. The child quickly grabbed the rug and told the student that she could do it herself."

"The final step routinely used, pertained to an investigation and summary of the child's moods: laughing, smiling, crying, screaming, whining, and pouting were consisted under this heading."

The authors have also recommended that this method of sampling a child's social contacts be used twice a year to review social adjustment of individual children.

Evidence of reliability, validity, and standardization:

No direct data on samples, norms, validity, or reliability have been provided for the use of this instrument. Some evidence has been presented which indicates, however, that the scoring is objective in that it can be reduced to enumerative procedures. Case histories have also been described which illustrate the application and value of the sociometric technique. The rationale for this instrument as explained by the author, would seem to provide some content validity.

Bibliography:

Reichenberg-Hackett, W. A sociometric technique for preschool children and its use in the study of individual child behavior. Journal of Humanistic Psychology, 1963, 3, 44-59.

TEST DATA WORKSHEET

Name of Test — Mother-Child Interaction Test

Author — Zunich, M.

Age — Nursery School Age

Area (i.e. language development) — Mothers' behavior toward children

Type of measure — Observation frame of reference

Source from which the test may be obtained — See Zunich, M. in bibliography

Description of Measure:
This is a frame of reference for use in observing mother-child "interaction," although the 17 categories refer primarily to the behavior of the mother toward the child, and involves little if any reciprocal behavior on the part of the child. Each category has a name, a short description of the kind of behavior to be observed, and an example. Thus, the first category is "Being Un-cooperative—e.g., Mother ignores the child's stimulation. Ex: Mother continues to read magazine when child addresses her." Other categories are "Giving Permission—e.g., Mother consents to child's proposed activity. Ex: 'Yes, you may use the towel." "Observing Attentively — e.g., Mother noticeably directs her attention to the child and/or child's activity by silently watching. Ex: Mother watches child as the child plays with the stove."

The total list of categories used is as follows: Being Un-cooperative, Contacting, Criticizing, Directing, Giving Permission, Giving Praise or Affection, Helping, Interferring, Interferring by Structurizing, Lending Cooperation, Observing Attentively, Playing Interactively, Reassuring, Remaining Out of Contact, Restricting, Structurizing, Teaching.

Evidence of reliability, validity, and standardization:
"Reliability was measured by calculating percentages of agreement between two observers recording maternal behavior simultaneously during thirty-minute periods."

The observers were trained to reach a certain level of agreement in their observations before actually beginning any collection of data.

Forty mothers and their children were observed each in five thirty minute observation periods by two observers, and the percentage of agreement between the observers for the 17 categories ranged from 81 to 100%, with the median agreement percentage of 88%.

Correlations were computed between the Mother's attitudes, as measured by the PARI, and their observed behavior toward their children in the 17 categories mentioned. Twelve of the 272 correlations were significant at the .05 level or beyond. The

authors point out, in view of the large number of correlations done and the possibility of a number of them being significant by chance, that "Although logical relationships are suggested by the significant values of the correlation coefficients between the six attitude subscales and six behavior categories the interpretation of these atatistical relationshps should be undertaken with caution." "Of the seventeen behavior categories, only two showed a significant relationship with the sex of the child at the .05 level when the median test was employed. The mothers of daughters evidenced more Contacting than did mothers of sons. However, mothers of sons showed more Observing Attentively behavior than did mothers with daughters."

Bibliography:

Zunich, M. Relationship between maternal behavior and attitudes toward children. Journ. gen. Psych., 1962, 100, 155-165.

TEST DATA WORKSHEET

Name of Test: SD Scale for Children

Author: Patricia W. Lunneborg and Clifford E. Lunneborg

Age:

Area (i.e. Language development): Social Desirability

Type of Measure Test: Questionnaire

Source from Which the Test May be Obtained: See Bibliography

Description of Measure:
"Originally, 89 items were drawn from the MMPI, Edwards Personal Preference Schedule, and other sources and rewritten for fourth-through sixth-grade children. The criteria employed in rewriting were that, in E's opinion, (a) the item had an obviously socially desirable response, and (b) the item was one for which it was expected that children answering without distortion would display variability and not all give the same answer, e.g., although the SD response to "I like to take a bath" was considered to be "yes", it was assumed some children would say "yes", some "no" if they could answer uninfluenced by test-taking attitudes ..." "... this SD item pool was judged "suitable" for the population intended by three elementary school officials, and relatively free of anxiety content by a panel of 10 graduate psychology students. These judges independently selected the response (yes-no) to the original 89 statements that "anxious children would give and the non-anxious would not", using a common definition of anxiety. Items were eliminated when there was significant agreement among the judges as to an "anxious" response.

From the resulting 78-item pool, and 20 item SD scale was derived.

Evidence of Reliability, Validity, and Useability:
"SD scores were significantly higher for girls than boys (t 4.78; p .01) and declined regularly with grade for each sex. However, the only significant mean difference for girls and boys combined was between the fourth and sixth grades (t 2.35; p .05). The split-half reliability coefficient for the SD scale, corrected for test lenght, is .83".

Bibliography of Test Usage:

Lunneborg, Patricia W. and Lunneborg, C. E. The Relationship of Social Desirability to Other Test-Taking Attitudes in Children. J. Clin. Psychol., 1964, 20, 473-477.

TEST DATA WORKSHEET

Name of Test	Slobodian Reading Observation Record (ROR)
Author	June Jenkinson Slobodian
Age	Teachers of elementary school children
Area (i.e. language development)	Teacher-pupil interaction in the setting of the reading group
Type of measure	Structured observation schedule
Source from which the test may be obtained	

Description of Measure:
The purpose of the Reading Observation Record was to "guide the observation and systematic recording of teacher-pupil verbal interactions in the setting of reading instructions...." The ROR "was developed following procedures outlined by Medley and Mitzel (1963). The ROR makes possible the sequential recording of verbal discourse segments ... and elements within these segments." Consequently, the schedule provides "data on the number of call-on response units, acceptance-response units, rejection-response units, interruption-response units, interruption-acceptance units, and interruption-rejection units." The ROR also provides data on the sequence of interaction, as well as the "time spent by teachers on various standard reading instructional activities."

Examples of the major headings within the ROR and the definitions for each are given below:

"2. Headings — The major headings are listed below with definitions for each:

a. Talk - Presentation of information, structuring by use of statements, possibly leading or transferring to questioning. Literally any statement which is not in response to a question or in response to teacher request for information falls in the 'talk' category.

b. Calls on/Questions - A teacher or student asking a question or requiring by command a response from a student or teacher is marked in this column.

c. Responds - Answers given to questions or replies made when called on are indicated by marking the appropriate box under this column.

d. Interruption - A sudden change of topic which occurs before the appropriate series of actions to close the unit of action is complete. Entrance of the reading group for in-

formation, etc. by a student outside the group. Teacher stopping the chain of discussion to speak to child either inside or outside the group for purposes of maintaining order or giving direction. Correcting of oral reading by supplying word missed, or by making the student reread for self-correction." (Manual, ROR).

Evidence of reliability, validity, and standardization:
on two different occasions, interrater agreement for the classification for the observation made reached a level of 95%. In one case, agreement was between the author and an assistant with no training in rating (Slovodian, 1966). In another instance, five experienced special reading teachers were trained as observers for a specific study. "All reached a criterion of 95 per cent agreement on all observations made during training" (Davis, & Slovodian, 1966). No data on norms or direct evidence of validity have been provided.

Bibliography:

Davis, O.L., & Slovodian, June J. Teacher behavior toward boys and girls in first grade reading instruction. Paper presented at American Educational Research Ass., Chicago, February, 1966.

Slovodian, June J. Personal communication, March 8, 1966.

TEST DATA WORKSHEET

Name of Test: Social Analysis of the Classroom

Author: Cunningham, Ruth

Age: Elementary School Age

Area (i.e. language development): Social Behavior

Type of Measure: Nomination Questionnaire

Source from Which the Test May be Obtained: See Cunningham 1951 in Bibliography

Description of Measure:
"The Social Analysis of the Classroom . . . offered 37 descriptions of behavior which each pupil could apply to as many or as few of his classmates as he thought fit" (Cunningham, 1961). The children are instructed to read each of the statements and to write down the names of the persons whom the descriptions fit. He is to write "myself" if he thinks the description fits him. Some sample items from the measure are given below:

Item 1: Here is someone who finds it hard to sit still in class; he (or she) moves around in his (or her) seat or gets up and walks around.

Item 4: Here is someone who doesn't like to talk very much, is very quiet, even when nearly everyone else is talking.

Item 12: This is someone who never seems to have a good time, who never seems to enjoy very much anything he (or she) does.

Item 13: This is someone who is always cheerful, jolly, and good-natured, who laughs and smiles a good deal.

Item 25: Here is someone who can enjoy a joke and see the fun in it even when the joke is on himself (or herself).

The reading level of this measure suggests that it is more appropriate for upper level elementary children and older.

Evidences of Reliability, Validity, and Standardization:
"Some items of the Social Analysis of the Classroom . . . seemed peculiarly descriptive of those who received high Group Social Distance scores, in that they were frequently assigned to such individuals in this group but were never used to describe those having the five lowest scores on this scale" (Cunningham, 1951, p. 179). Thus, these items seem to be significantly related to scores on the Social Distance Scales.

Bibliography:

Cunningham, Ruth. Understanding Group Behavior of Boys and Girls. New York: Teachers College, Columbia University, 1951.

MEASURES OF PERCEPTUAL SKILLS

TEST DATA WORKSHEET

Name of Test: Boston University Speech Sound Discrimination-Picture Test

Author: Wilbert Pronovost and Charles Dumbleton

Age: Kindergarten and first grade

Area (i.e. language development): Articulatory discrimination

Type of measure: Picture-type sound discrimination test

Source from which the test may be obtained: Dumbleton, C., et al., An analysis of the relationships between speech and reading abilities of four hundred and twenty five first grade children. Unpublished Master's thesis, Boston University, 1952.

Description of Measure:
The Boston University Speech Sound Discrimination Test requires children to identify paired word sounds correctly and to discriminate between unlike sounds. It is essentially a revision of a similar scale originally developed by Mansur 91950) and partially validated by Haroian (1951). "The revision took the form of eliminating some word pairs, adding other different pairs, using fewer of the possible combinations of each pair, and revising the picture sheet to obtain random positioning of the pairs with only three word pair pictures." (Pronovost and Dumbleton, 1953, p. 258). These arrangements on a test page included "... one 'unlike' pairing and two 'like' pairings" [e.g., cat-cat, bat-bat, cat-bat]. "The position of the particular word pairs was selected at random in order to eliminate the factor of pattern responses. Simple line drawings were used for the pictures, which were reproduced by means of multilith.. . .

In administering the test, only two of the possible combinations of each pair of words were used: the 'unlike' pairing as pictured on the picture sheet, and one of the 'like' pairings determined at random. The order of presentation of the items was determined at random also. On each half of the test, one-half of the 'unlike' pairs was presented, and one-half of the 'like' pairs--the picture sheets being used twice in the same order.. . .

On the present test the child, in order to respond to a presented 'unlike' pair, must make only a 'like-different' judgment. In response to a presented 'like' pair, he must make a 'like-different' judgment and an identification."

The complete test consists of 72 paired words. The auditory characteristics of the paired sounds consisted of vowels (e.g., pen-pin), semi-vowels and nasals (e.g., lock-rock), plosives (e.g., cat-bat), fricatives (e.g., vase-face), blends (e.g., grass-glass), and miscellaneous (e.g., pan-sand).
As an example of the procedure, a description of the "familiarization" follows:

"4. Use picture No. 1 (cat-bat) for familiarization instruction as follows: 'We are going to play a game with these pictures. Each page has three pictures on it like this. On this one there is a picture of a cat and a bat, a bat and a bat, and a cat and a cat. When I say two words, you point to the right picture. If I should say cat-bat, you would point to this one. Which one would you point to if I said bat-bat? Now listen carefully, because I am only going to say it once. Point to cat-cat.'"

Complete instructions for scoring and administering the test are available. Though the majority of the children complete the test in 10 to 15 minutes, there is no time limit for administering the scale (Pronovost and Dumbleton, 1953).

"The test was originally designed to test speech sound discrimination of young children with misarticulations of speech sounds. The correlations of discrimination ability and articulatory proficiency are low and we do not use the test except with the children with a large number of misarticulations. Dr. Bernard Schlanger of Ohio State University has used the test with retarded children. Several psychologists have used the test as a gross auditory discrimination test for the testing knowledge of the same-different concept. It has been used in Head Start programs and with neurologically impaired children, but I have not seen any of the data (Pronovost, 1966).

Evidence of reliability, validity, and standardization:
"The number of correct responses was assigned as the test score. The range of scores for the 434 children tested was from 72 to 16; the standard deviation was 6.56, the mean was 65.5, and the mode was 70. These measures indicate a distribution which is strongly skewed negatively." Further, about ". . . 10% of the children attained scores which were one or more standard deviations below the mean. In view of the skewedness of the distribution of the results of this test, one would be justified in assuming that any socre one or more standard deviations below the mean would indicate poor ability in speech and sound discrimination."

"Reliability of the test:" The product-moment correlation coefficient between the two halves of the test, with correction by the Spearman-Brown formula for attenuation, was .88. "This coefficient is sufficiently high for the test to be used with confidence." Moreover, the ". . . order of difficulty obtained for the items was similar to those found by Mansure. . . /1950/ and Haroian . . .. /1951/.

Validity appears to rest essentially on item analysis based on the per cent of correct responses on each item by upper and lower quartile groups. On the strength of this procedure with the phi coefficient, 65 items were significant at the .01 level of confidence or better. These coefficients range from .17 to .41 (Pronovost and Dumbleton, 1953, p. 263). The authors, however, give the following precautions regarding the use of the tests.

/1./ "The instrument is valid and reliable, but subject to some limitations. It does not test on a high enough level of speech sound discrimination to be used in a definitive study of the abilities of the normal population. When used as diagnostic instrument, the instrument is valuable as an indicator of the adequacy of a child's speech sound discrimination abilities. The inability to find usable word pairs which can be pictured easily, however, makes it impossible to test each difficult sound in various combinations.

Evidence of reliability, validity, and standardization:

/2./ The limit on the validity of the instrument in individual use is the articulation of the person administering the test. If the tester prolongs or emphasizes the sound which differentiates the words of a pair, the child's score will be higher than it should be. . . . Greater validity could be obtained if a tape recorded test were used under controlled acoustical conditions" (Pronovost and Dumbleton, 1953, p. 266).

Bibliography:

Dumbleton, C. et al. An analysis of the relationships between speech and reading abilities of four hundred and twenty-five first grade children. Unpublished Master's Thesis, Boston University, 1952.

Haroian, R. D. Preliminary validation of Mansur's speech sound discrimination test in the kindergarten and first grade. Unpublished Master's Thesis, Boston University, 1951.

Mansur, R.W. The construction of a picture test for speech sound discrimination. Unpublished Master's Thesis, Boston University, 1950.

Pronovost, W. Personal Communications, May 17, 1966.

Pronovost, W., Dumbleton, C. A picture-type speech sound discrimination test. J. Speech & Hearing Disorders, 1953, 18, 258-266.

TEST DATA WORKSHEET

Name of Test	Elkind's Ambiguous Pictures
Author	David Elkind
Age	6-11
Area (i.e. language development)	Perceptual development
Type of measure	Performance scale consisting of ambiguous black and white figures

Source from which the test may be obtained "Photoprints of each of the pictures...have been deposited with the American Documentation Institute. Order Document No. 8154 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.25 for photoprints or $1.25 for 35mm microfilm. Make checks payable to: Chief, Photoduplication Service, Library of Congress" (Elkind, 1964, p. 1391).

Description of Measure

"The materials consisted of two sets of ambiguous pictures printed in black and white and mounted on 8 x 11½ inch tag board sheets. For one set of figures cardboard shields were available which were so cut that when they were placed over the ambiguous figure the hidden (reversed) figure became immediately apparent..." (Elkind, et al., 1965a). Standardized directions are available for the administration of the test, its scoring, timing, and sequence of presentation" (Elkind, 1964). These directions are also available for obtaining a criterion score of learning (Elkind, et al., 1965b).

The test has been used to obtain normative data in perceptual learning and development (Elkind, 1964) to check Piaget's hypothesis that complex perceptual phenomena--"such as illusions, size constancy, and figure-ground reversals,--"...derived from the interaction of maturation and experience" (Elkind, et al., 1962); to investigate "the effects of perceptual training on unmatched samples of brain-injured and familial retarded children" (Elkind, et al., 1965a); and to study Piaget's theory of decentration as it relates to reading performance. This theory holds that, with increasing age, "...the child's perception becomes progressively decentered in a sense that it is gradually freed from its earlier domination of field effects" such as, "good form," "continuity," and "closure" (Elkind, et al., 1965b).

Samples of the scored figures on each card for both Set A and Set B are listed below:

	"Set A	"Set B
Card 1	A butterfly Face (left) Face (right)	A leaf Face (left) Face (right)
Card 2	A tree A cat	A tree (or head) Face (left, bottom) Face (left, top) Face (right, bottom) Face (right, top)" (Elkind, 1964, p. 1394).

Evidence of reliability, validity, and standardization

The test has been administered for various purposes and to subject groups as follows: (1) to obtain normative data: total N of 135, age range of 6-11, lower middle-class children, N within each age range 18-27 (Elkind, 1964);

(2) Piaget's theory on perceptual complexity: N of 69, range of age levels between 6 and 8, children attending a university school (Elkind, et al., 1962);

(3) brain-injured (BI) and familial retardates (FR): 49 BI, average C.A. of 93.37 months, 39 FR, average C.A. of 101 months (1965a); and

(4) perceptual decentration and reading performance: 60 children in grades 3-6, half of which were competent readers, with the remaining half retarded one or two years in reading (Elkind, et al., 1965b).

"Results of the...normative study has shown that: (a) the two sets of ambiguous pictures have comparable means and S.D. at each age level; (b) the correlations between the two sets are positive and tend to increase with age; (c) with the exception of the sixth-grade group, the mean ambiguous picture scores tend to increase regularly with age; (d) there is a low but positive correlation between scores obtained on the ambiguous pictures and those obtained on a widely used group test of general intelligence." In general, the correlation coefficients between these two tests (A and B) of ambiguous pictures range from .32 to .71. The correlation between the combined scores (Score of Set A plus score of Set B) and the Kuhlman-Anderson Intelligence Test range from .44 to .67 for ages 8-11. At age 7, the small r was -.13 and at age 6 it was .19 (Elkind, 1964).

The results from the use of this test have also revealed the following:

(1) In the case at least of cerebral-palsied injured youngsters, "...the BI child requires a higher level of mental ability than the FR child who reached the same level of perceptual achievement (Elkind, et al., 1965a).

(2) All children (age range 6-8) improve greatly with special training. But, "... the influence of maturation was suggested by the finding that 6- and 7-year-old children required more intense teaching and reached a lower level of performance than did the 8-year-old children." These findings are regarded as consistent with the cross-sectional observations reported by Piaget on which he based his theory on interactional perception (Elkind, et al., 1962). In the comparison of test findings between slow (N of 30) and competent readers (N of 30), who were matched on crucial variables, the results indicated that "...slower readers were significantly less adept than average readers of comparable age, intelligence, and sex. Slow readers had lower pre-training scores, required more and more direct clues to reach the learning criterion during training, and transferred the effects of training to a significantly lesser extent than did the average readers." The authors interpreted the data in this fashion. "While these results do not prove the role of perceptual decentration in reading, nor the validity of the logical analysis of this process, they do provide evidence for the construct validity of the notion of decentration and for the fruitfulness of the logical model of perception for generating testable hypotheses" (Elkind, et al., 1965b, p. 55).

Bibliography

Elkind, D. Ambiguous pictures for study of perceptual development and learning. Child Development, 1964, 35, 1391-1396.

Elkind, D., Koegler, R. R., and Go, Elsie. Effects of perceptual training at three age levels. Science, 1962, 137, 755-756.

Elkind, D., Koegler, R. R., Go, Elsie, and Van Doornick, W. Effects of perceptual training on unmatched samples of brain-injured and familial retarded children. Journal of Abnormal Psychology, 1965, 70, 107-110.

Elkind, D., Larsen, Margaret, and Van Doornick, W. Perceptual decentration learning and performance in slow and average readers. Journal of Educational Psychology, 1965, 56, 50-56.

TEST DATA WORKSHEET

Name of Test	Hidden Figures Test
Author	Kagan, J. et al.
Age	Elementary School Children
Area (i.e. language development)	Visual pattern analysis
Type of measure	Individually administered picture-type test
Source from which the test may be obtained	Examples of items are given in Kagan, J., et al., 1964.

Description of Measure

"HFT was devised especially for this study and consisted of one practice and 11 test items. Each item included a figure card that illustrated a familiar object (e.g., Mickey Mouse, a cowboy, a rose), and a hidden figure card in which the familiar object was embedded (i.e., blended into a patterned background)" (Kagan, J., et al., 1964).

The administration of the test is described by Kagan as follows: "The E said, 'Here is a picture. What is this?' On the rare occasion when the child did not respond correctly E said, 'It is a _____.' Then the hidden figure card was placed in front of the child to the right of the figure card and S was told, 'Now you find the _____ in this picture. Point to the _____ when you find it.' If S selected incorrectly, he was told that he had not found the figure and was encouraged to look for it again. A maximum of 60 seconds was allowed for each item. The number of incorrect guesses, response time to the correct selection, and success or failure on each item were recorded. The items differed markedly in difficulty, and a weighted score for each child's performance was computed using weighted solution times for each item" (Kagan, J. et al., 1964).

Evidence of reliability, validity, and standardization

"The odd-even intratest reliability coefficients, corrected by the Spearman-Brown formula, were .78 for the first grade and .41 for the second grade" (Kagan, J. et al., 1964).

Some evidence of construct validity is suggested in the study by Kagan which hypothesized a strong relationship between analytic responses, as measured by the Conceptual Style Test, and in inhibition of incorrect solution hypotheses on the Hidden Figures Test. This hypothesis was supported.

Bibliography

Kagan, Jerome, et al., Information processing in the child. Psychological Monographs, 1964, 78, No. 1 (Whole No. 578).

TEST DATA WORKSHEET

Name of Test	Irwin-Jenson Sound Discrimination Test
Author	Orvis C. Irwin and Paul J. Jensen
Age	6-17 for two groups of exceptional children: (1) cerebral palsy and (2) educable mentally retarded
Area (i.e. language development)	Speech (Sound Discrimination)
Type of measure	Individually administrated objective rating scale
Source from which the test may be obtained	See reference.

Description of Measure:
The Irwin-Jensen Sound-Discrimination Test consists of 30 pairs of words which are orally presented to the child by the examiner on a one-to-one basis. The child is simply requested to indicate if the words are the same or different. Three possible scores are available: The total number of correct responses, the total number of errors, and the number of no responses. In calculating the scores, five which are pairs of identical words, should not be used. These simply serve to indicate if the child is listening. The 25 word-pairs, then, are unlike words.

In administrating the test, Irwin and Jensen give two precautions. First, they suggest that the child be tested for hearing. Secondly, they recommend that a check be made by the examiner to determine if the child is mature enough to understand the difference in the meaning of "same" and "different" (Irwin and Jensen, 1963a). A parallel form (Form B) with adequate reliability and validity has also been developed (Irwin, 1963b).

Evidence of reliability, validity, and standardization:
Forms A and B have been administrated to both cerebral palsied and mentally retarded children. In the former group, Form A was administrated to "153 children with cerebral palsy from six through sixteen years of age with the mean chronological age of 10.3 years. The mean mental age was 6.3 years, the range was 2.8 to 17.0 years. They were from speech centers, public schools, and hospitals in eleven states" (Irwin and Jensen, 1963b). On the other hand, Form B "was administrated to 260 cerebral palsied children in two geographical regions." The children were from eleven states primarily from the wet and midwest areas of the U.S. "The age range was from six to seventeen years with a mean chronological age of 10.9 years. The mean mental age was 6.8 years and the range was 1.7 to 15.1 years" (Irwin and Jensen, 1963b).

Reliability of Form A rests on observer agreement and internal consistency. Thus, two observers who recorded the responses of 65 cerebral palsied children with speech defects were able to obtain a mean overall agreement of 96%.

With mentally retarded children, Form A was administrated to 357 mentally retarded children (1965a). Form B was administrated to "a total of 347 mentally retarded

children from 6 to 17 years of age." (1965b). Means and standard deviations for all four groups, subdivided by 6, have been provided. However, no significant differences between sex groups have been found for either form.

The reliability of Form A by means of a Kuder-Richardson Formula was (r=.87.

"A further indication of the reliability of the test is afforded by an analysis of variance in which the two sexes in the two geographical areas are entered. The result ... showed no significant differences among the means. Consequently, each group mean varifies the other group mean." To Irwin and Jensen this finding constitutes "another demonstration of the internal consistency or of the reliability of the test" (1963a).

Validity of Form A depends on a method of item selection and external criteria.

Method of item selection. Item analyses of a preliminary pool of 50 items was made on the basis of three criteria: (1) the power of the items that differentiate among examinees, (2) the uniqueness of the items or their lack of interrelatedness, and (3) the difficulty level of the items. The mean and methods for each step have been described in full. On the basis of these three criteria, a final scale of 25 items was developed.

External criteria. The validity of Form A was determined by the method of extreme groups. In this connection, the significant differences in mean scores for the test were found for extreme groups and chronological age levels, mental age levels, and speech and language ability as judged by speech therapists (1963a).

In Form B, the final 25 items were also selected from the 3 criteria described above. Validity likewise was derived from the identical three extreme group comparisons previously mentioned. Further, the Kuder-Richardson coefficient of reliability was nearly identical (r of .88). Finally, parallel form reliability is indicated by an r of .90 with homogenous means and variances (1963b).

The use of the final version of the test wi h mentally retarded children yielded similar results. Thus, Form A (1965a) and Form B (1965b) differentiated between extreme groups classified on the basis of chronological age, mental age, and educational status (educable and trainable children). Further, the respected correlations of Form A (1965a) and Form B (1965b) with the Templin Sound Discrimination Test were respectively .83 and .73 (p .01). Again, with both forms, Kuder-Richardson reliability coefficients were very similar to those found for cerebral palsy children.

As a final note, it may interest some workers to know that Irwin and Hammill (1965c) had done an item analysis of Form A with mentally retarded children.

Bibliography:

Irwin, O.C., & Jensen, P.J. A test of sound discrimination for use with cerebral palsied children. Cerebral Palsy Review, 1963a, 24, 5-13.

Irwin, O.C., & Jensen P.J. A parallel test of sound discrimination for use with cerebral palsied children. Cerebral Palsy Review, 1963b, 24, 3-10

Irwin, O.C., & Hammill, D.D. A comparison of sound discrimination of mentally retarded and cerebral palsied children. Form A. Cerebral Palsy Review, 1965a 26, 3-6.

Irwin, O.C., & Hammill, D.D. A second comparison of sound discrimination of cerebral palsied and mentally retarded children. Form B. Cerebral Palsy Review, 1965b, 26, 3-6.

Irwin, O.C., & Hammill, D.D. An item analysis of a sound discrimination test, Form A, for use with mentally retarded children. Cerebral Palsy Journal, 1965c, 26, 9-11.

TEST DATA WORKSHEET

Name of Test: Pain Threshold Test

Author: Graham, F.K., Matarazzo, R.G., and Caldwell, B.M.

Age: Infants (Neonates 1-5 days old)

Area (i.e. language development): Impairment of sensory function

Source from which the test may be obtained: Authors

Description of Measure:
Essentially the Scale measured an infant's reaction to an electronic stimulator. "The apparatus delivers shocks whose intensity, duration, and frequency can be varied within the stimulating range for skin and peripheral nerves. For the present work, duration and frequency were held constant so that the shock consisted of a two-second impulse at a frequency of 14 per second which could be varied in intensity from 50 to 530 volts. Eleven voltages covering this range and corresponding to the points on a control dial were selected as stimuli. The average difference between stimuli was 50 volts. A resistance of one-half megohm was in series in the circuit, thus reducing the effect of individual skin resistance. We also made separate measurements of skin resistance in a number of Ss to check that, in fact, the resistance did not determine the response.

The infant was stimulated while lying on his back with the legs in the usual flexed position. The indifferent electrode, covered with moist gauze, was placed in the center of the back against the skin, and the shirt tied so that it remained in position. A circular area, approximately 3/4 inches in diameter, just below the knees of both legs was treated with vaseline. These areas were stimulated alternately by a silver ball electrode of 1 mm. diameter. An effort was made to avoid stimulating the same spot successively, especially if the skin reddened. There was apparently sufficient spread of stimulation so that differences in spot sensitivity did not introduce undue variable error."

"The specific response required was a movement of the stimulated leg or the foot of that leg, with or without other bodily movement. . . ."

"Four threshold determinations were made, using descending-ascending-descending-ascending order of presentation, and the average of these four series was taken as the threshold...." Assurances were made that the infant was wide awake before stimulation was applied and the series did not begin until he responded to two consecutive stimuli. Consequently the stimuli of varying voltage intensities were applied. "The series was continued until the infant did not respond to the last stimulus evoking a response and the first stimulus which did not evoke a response" (Graham et al., 1956, pp. 4-5).

Evidence of reliability, validity, and useability:
"Reliability of the procedures was measured by split-half correlation, test-retest

agreement and interscorer agreement as applicable. All tests appeared to be satisfactorily reliable" (Graham, 1956, p. 32). Split-half reliabilities for the pain threshold scores for both traumatized and nontraumatized ranged from .82 to .97. The other subtest scores could not be satisfactorily divided into comparable halves. Test-retest reliabilities for the Pain, Maturation, and Vision scales were from .62-.69. On the Irritability and Vision scales, the test-retest agreements after 24 hours were respectively 75% and 86% of perfect agreement. Statistics on interscorer agreement were as follows: (1) Maturation Scale (r=.97; (2) Vision Scale, (r=.90); (3) Irritability Scale (68% perfect agreement). Finally, there was 79% perfect agreement in the Tension Scale (Graham, 1956, pp. 21-22).

In addition to the reliability of individual raw scores, statistics are provided for differentiating groups and accurate identification of individuals as members of a particular group. In the former case, statistically significant differences (p of .05 or .01) between the normal and traumatized groups occurred on all five scales based on F test, t test, and Chi square (Graham, 1956, p. 24). In the latter instance, two pediatricians rated 25 cases of traumatized infants with respect to the severity of the trauma. ". . . the extent of the agreement was high and significant as estimated by a correlation ratio of .85 (Graham, 1956, p. 18).

Bibliography of test usage:

1. Graham, Frances K.; Matarazzo, Ruth G.; and Caldwell, Bettye M. "Behavioral Differences Between Normal and Traumatized Newborns...." Psychological Monographs, 1956, 70, No. 20, (Whole No. 428).

TEST DATA WORKSHEET

Name of Test: Right-Left Discrimination

Author: Arthur L. Benton

Age: Five to adult

Area (i e. language development): Right left discrimination

Type of measure: Performance Test

Source from which the test may be obtained: See (1) in bibliography.

Description of Measure:
This test is designed to measure the ability of children and adults to discriminate between right and left. There are two forms of the test: "Form A, which requires the subject to execute localizing movements to command, consists of 32 items. It assesses six aspects of right-left discrimination, the specific tasks being as follows:

1. With the eyes open, pointing to single lateral body parts.
2. With the eyes open, execution of double crossed and uncrossed commands.
3. With the eyes closed, pointing to single lateral body parts.
4. With the eyes closed, execution of double crossed and uncrossed commands.
5. Pointing to lateral body parts on a schematic, frontview representation of a person.
6. Execution of double crossed and uncrossed commands involving lateral body parts of both the subject and the schematic representation." (1).

Form A requires no verbal responses of the subject, and no emphasis is placed on speed. "Form V, which requires the subject to name lateral body parts, consists of 36 items. Designed to be the "verbal response" analogy of Form A, it assesses virtually the same aspects of right-left discrimination, the specific tasks being as follows: 1. With the eyes open, naming single lateral body parts. 2. With the eyes open, and after the examiner has placed one of the subjects hands on an ipsilateral or contralateral body part, identifying both the hand and the indicated body part in terms of right and left. 3. With the eyes closed, naming single lateral body parts. 4. With the eyes closed, and after the examiner has placed one of the subjects hands on an ipsilateral or contralateral body part, identifying both the hand and indicated body part in terms of right and left. 5. Naming lateral body parts on a schematic, front view representation of a person. 6. Identifying the hand and body part in terms of left and right on front-view representations which depict crossed and uncrossed localizations" (1).

Form V therefore requires the subject to give a verbal response. "Comparisons of a patient's relative level of performance on the two forms have yielded particularly interesting findings in the aphasic disorders" (1).

Form A consists of 32 items of which the following are examples:
1. Show me your right eye.
2. Touch your right knee with your right hand.
3. Point to the man's left ear.
4. Put your left hand on the man's right shoulder.

All of the above are done with the eyes open and with the eyes closed
1. Show me your left leg
2. Touch your right eye with your left hand.

Form V of the battery includes items like the following: With the eyes open, the examiner touches the subjects left hand and says "which hand is this?" The examiner then points to a picture and says 'which ear is this?' The examiner places the subject's right hand on the subject's right ear and asks "which hand is on which ear?" Now look at these pictures (Picture 2 for example, shows a person with the right hand on the left eye) "which hand is on which eye?"

With the eyes closed or blindfolded, the examiner touches the subject's left hand and says "which hand is this?" With the subject's eyes still closed, the examiner gives the same instructions as previously, placing one of the subject's hands on one of his ears and asking him which hand is on which ear?" The same items are used but in reverse order, presumably to eliminate the possibility of the subject's memorizing the order and correct responses.

According to Benton, "a review of the clinical literature shows clearly that right-left discrimination and finger agnosia typically do not appear as isolated phenomena. Instead, they occur in association with one or more of a considerable variety of other behavioral deficits." Some of these deficits are mental deterioration, aphasic disturbances, disturbances in body awareness, distrubances in writing, agraphia, distrubances in calculations, distrubances in reading, visual disorientation, and others.

Evidence of Reliability, Validity, and Standardization:
"Corrected odd-even reliability has been found to be .88-.92 in samples of normal children. Retest reliability, as estimated by the correlation coefficient between scores on equivalent forms given 20 minutes apart, was found to be .72. Stability of performance over a period of time, as estimated by the correlation coefficient between scores on equivalent forms given 10 weeks apart, was found to be .67. Immediate practice effect, as assessed by a comparison of mean scores on equivalent forms given 20 minutes apart, was found to be practically nil." (1) Norms are given for children ages six through nine (1, pp. 27-28).

Bibliography:
Benton, A.L., and Cohen, B.D. (1955). Right-left discrimination and finger localization in normal and brain-injured subjects. Proceedings of the Iowa Academy of Science, 62: 447-451.

TEST DATA WORKSHEET

Name of Test: Strauss-Werner Marble Board Test (Goldenberg version)

Author: Samuel Goldenberg

Age: Elementary children with Binet Vocabulary IQ scores of 80 or above (Goldenberg, 1955, p. 162).

Area (i.e. language development): Psychological diagnosis of organic brain disorders in children

Source from which the test may be obtained: (Goldenberg, 1955, p. 144-164)

Description of Measure:
The Marble Board Test, which measures the reproduction of visually perceived stimuli, " ... involves the use of two identical square boards, each having ten rows of ten holes. The examiner places a design or configuration on one board by putting marbles in certain of the holes and then the subject is asked to reproduce the design on his board. A record of the location and sequence of placements is made for each pattern" (Goldenberg, 1955, p. 152). The specific dimensions for producing these boards for anyone interested in using the test have been specified by Strauss and Lehtinen (1947).

In his version of the Strauss-Werner Marble Board Test, Goldenberg has developed a very objective scoring system relating to accuracy, method or approach, and organization. Examples and definition of each scoring procedure have been provided (1955, pp. 215-219). Examples of items for each type of scores are as follows:

"MARBLE BOARD SCORING GUIDE

Accuracy:
A. Perfectly accurate--score five
   1. Reproduction like the model in all respects, except
   2. Position of configuration on the board may vary from the model in the number of empty rows of holes above, below or to the sides of the pattern.

. . . . . .

Method:
A. Continuous
Consecutive placement of adjacent marbles proceeding in a single direction of movement:
   1. Along the configuration outline (composed of one or more component subforms) to completion of subform;

   2. Along outlines of component subforms, each completed with a single direction of movement.

. . . . . .

Organization: (patterns 3, 4, 5, 6)
A. Articulate
Each subform is reproduced in sequence; one is complete before beginning on the other; overlapping forms must overlap" (Goldenberg, 1955, p. 215-219).

Evidence of reliability, validity, and standardization:
The Marble Board Test was administrated as part of a battery of tests "to three groups of children: 26 children with clear diagnosis of brain injuries, 32 emotionally disturbed children and 32 normal children. All of these subjects had Binet vocabulary IQ scores of 80 or above.

Discriminant function analysis reveals that "the 'approach' aspect of the Marble Board Test was found to differentiate the brain injuried individuals at the same level of efficiency as did the Ellis test." The Ellis Visual Designs, in turn, "contributed nearly all the maxium discrimination which could be obtained with the nine test measures included." This statistical procedure was applied to the best combination of tests for detecting brain-injuried children with respect to Efficiency Score for the test variables.

"The Efficiency Scor is the difference between the proportion of brain-injuried children correctly identified according to their test performance and the proportion of the non-injuried subjects incorrectly identified (false positives)."

In short, "the most adequate combination of test for the purpose of diagnosis of brain injury in children, from among those studied, is composed of the Ellis Visual Designs and the Marble Board Test. There was only negligible gain in differentiation of individuals as a result of using additional measures of test performance" (Goldenberg, 1955, p. 163-164).

The reliability of this approach or method score was checked by the following procedure: "Two experienced clinical psychologists were trained in the use of the scoring rules in several practice sessions and then they and the author independently scored 104 Marble Board performances produced by 13 subjects randomly selected from the first 40 non-injuried subjects tested" (Goldenberg, 1957). The 88 per cent agreement reported for the Approach score (Goldenberg, 1955, p. 167) was for these 104 performances. On the basis of this same procedure, other statistical data indicated 91 per cent agreement in scoring Accuracy, and 77 per cent complete agreement on 52 performances in scoring Organization (Goldenberg, 1967).

Finally, in a subsequent communication on the scoring system Goldenberg (1957) has used tentative cut-off score of 3 or more ratings of incoherent to identify brain-injuried children between the ages of about 7 to 13. This is the "cutting point which Strauss had previously found effective ...."

Bibliography:

Goldenberg, S. A I: Scoring guide to Marble Board Test and Ellis Visual Design Test. IN A.A. Strauss & N.C. Kephart (Ed.), Psychol. & ed. of the brain-injuried child. Vol. 2. Progress in theory and clinic. New York: Grune & Stratton, 1955, pp. 215-222.

Bibliography (continued)

Goldenberg, S. Personal communication. April 2, 1957.

Strauss, A.A., Lehtinen, Laura. Psycholpathology and Education of the brain-injuried child. New York: Grune & Stratton, 1957.

TEST DATA WORKSHEET

Name of Test: Vision Scale

Author: Graham, Frances K., Matarazzo, Ruth G., and Caldwell, Bettye M.

Age: Infants (Neonates 1-5 days old)

Area (i.e. language development): Visual maturation

Type of measure: Visual fixation, pursuit, and eye movement

Source from which the test may be obtained: Authors

Description of Measure

"The stimulus is presented with the infant in a supine position. Either the examiner's hand, a ball, or a metal tape measure 1 and 3/8 inches in diameter was used. No further effort was made to standardize the stimulus, since we were attempting to determine whether the capacity was present under the most favorable conditions..." Several aspects of the visual response are considered in making a classification: (a) presence or absence of a kind of response (such as fixation or pursuit); (b) the ease of eliciting the response; (c) the direction of eye movement (horizontal or vertical); (d) the distance the eye moves... Each response (or item), with the exception of the zero point, is given an arbitrary weight of one, so that the number of items passed is the S's score" (Graham, et al., 1956, pp. 11-12).

No credit is given if no response of fixation or pursuit occurs in any one of carefully described abnormal features is observed. Illustrations of criteria to follow in allowing credits for the ten items are below:

"Item 1. No fixation or pursuit but none of the above abnormalities is observed.
Item 2. Fixation is brief or obtained with difficulty...
Item 3. Fixation is clearly present and easily elicited, i.e., occurs in a majority of the trials.
Item 4. Horizontal pursuit is questionably present..." (Graham, et al., 1956, p. 12).

Evidence of reliability, validity, and standardization

"Reliability of the procedures was measured by split-half correlation, test-retest agreement and interscorer agreement as applicable. All tests appeared to be satisfactorily reliable" (Graham, 1956, p. 32). Split-half reliabilities for the pain threshold scores for both traumatized and nontraumatized ranged from .82 to .97. The other subtest scores could not be satisfactorily divided into comparable halves. Test-retest reliabilities for the Pain, Maturation, and Vision scales were from .62 - .69. On the Irritability and Vision Scales, the test-retest agreements after 24 hours were respectively 75% and 86% of perfect agreement. Statistics on inter-scorer agreement were as follows: (1) Maturation Scale (r=.97); (2) Vision Scale, (r=.90); (3) Irritability Scale (68% perfect agreement). Finally, there was 79% perfect agreement in the Tension Scale (Graham, 1956, pp. 21-22).

In addition to the reliability of individual raw scores, statistics are provided for differentiating groups and accurate identification of individuals as members of a particular group. In the former case, statistically significant differences (p of .05 or .01) between the normal and traumatized groups occurred on all five scales based on F test, t test, and Chi square (Graham, 1956, p. 24). In the latter instance, two pediatricians rated 25 cases of traumatized infants with

respect to the severity of the trauma." ...the extent of the agreement was high and significant as estimated by a correlation ratio of .86" (Graham, 1956, p. 18).

Bibliography
Graham, Frances K., Matarazzo, Ruth G., and Caldwell, Bettye M. "Behavioral Differences Between Normal and Traumatized Newborns..." Psychological Monographs, 1956, 70, No. 20 (Whole No. 428).

MISCELLANEOUS

TEST DATA WORKSHEET

Name of Test Bicycle Safety--Performance and Skills Test

Author Miller, B. W.

Age Eight and older

Area (i.e. language development) Bicycle skill

Type of measure Performance

Source from which the test may be obtained See Miller (1962) in bibliography or write National Safety Council, 425 N. Michigan Avenue, Chicago, Illinois 60611

Description of Measure

The author describes 12 tests which measure skills on the bicycle. They are designed to measure specific skills such as balance, pedaling, braking, hand signaling, dismounting, and fine control. The 12 tests are as follows:

1. Balance Test (Straight Line)
2. Pedaling and Braking
3. Straight Line Test
4. Signaling, Mounting and Dismounting
5. Single Obstacle Test
6. Double Obstacle Test
7. Double Zig-Zag Obstacle Test
8. Figure Eight Steering
9. Figure Eight Balance Test
10. Turning Around
11. Emergency Turn and Stop
12. Cruising Test

An example of one of the tests is Test I, Balance Test (straight line). "Purpose: To test the delicate balance of the rider, the most important skill technique." Following this paragraph is a diagram showing the length and width of the pathway which the cyclist must negotiate. "Procedure: The rider starts from a standstill with the front wheel at one end of the lane and very slowly rides through the lane in not less than 30 seconds, with neither tire touching the lines on either side."

"Standards for Success;
( ) 1. Touching neither foot to the ground.
( ) 2. Going distance in more than 30 seconds.
( ) 3. Having neither wheel touch either line.
( ) 4. Not using brake excessively.
( ) 5. Expending not more than average amount of energy" (Miller, 1962).

Several diagrams are shown giving the layout of obstacle courses and other skill tests.

"The tests may be conducted on a success-failure or a point basis. The former is recommended for teaching purposes, the latter for contests" (Miller, 1962).

Miller states or shows in his summary score sheet on the last page of his article a possible point total of 70. It appears that this should be 75, since in his summary score sheet he gives it a total possible of 15 for the final test, the cruising test, whereas in the description of that test he has a total possible of 20. This would make the total possible score on this test 75.

Evidence of reliability, validity, and standardization

None

Bibliography

Miller, B. W. Skill tests for pedal pushers. Safety Council Magazine, 1962.

TEST DATA WORKSHEET

Name of Test: Blatt's Evaluation Check List for Classes

Author: Burton Blatt

Age: Elementary EMH classes

Area (i.e. language development): Evaluation of programs and services in EMH classes

Type of measure: Objective check list questionnaire

Source from which the test may be ovtained: (Blatt, 1966)

Description of Measure

The schedule for describing and evaluating special classes for EMH children by Blatt is an objective check list in which an observer and/or rater checks for the quality or presence of certain features and services available for a given program. This objective evaluation pertains to six areas: (a) details of observed lesson(s) and general learning, (b) summary rating of teacher in learning environments, (c) summary rating of children in learning environments, (d) summary rating of diagnosic and placement procedures and extent of consultative and supervising practice, (e) summary rating of quality of observed lesson(s) and curriculum emphasis, (f) per cent of daily distribution of time (teacher's estimate). Place is also provided for the observer to record additional comments and impressions.

Area A is given below for illustrative purposes:

"A. DETAILS OF OBSERVED LESSON(s) AND GENERAL LEARNING ENVIRONMENT

1. Materials for Observed Lesson (double check) if materials were adequate for lesson)
   a. None
   b. Textbooks, workbooks, and other commerical resources
   c. Commercial, teacher prepared material, and/or teacher-pupil materials" (Blatt, 1966, p. 61).

Evidence of reliability, validity, and standardization

Blatt has provided no objective data regarding the standardization, reliability, and validity of the check list. Two procedures, however, would seem to provide some content or logical validity for the evaluation schedule. First, he contends that "perusal of the literature failed to locate an instrument meeting all the following requirements:

1. It should be designed for use with special classes for the mentally retarded.
2. It should describe what is taking place among parents, administrators, teachers, and children in these classes.
3. It should pose questions that are answerable in one observation session.
4. It should be easily managed by the observer and be relatively free of ambiguous questions.
5. It should make provision for both quantitative and qualitative evaluations of the class" (Blatt, 1966, p. 59).

Blatt believed his instrument meet these requirements. Secondly, the manner in which he developed the items lead one to the conslusion of content validity. " 'Promising" questions culling from the literature, plus a variety of suggestions elicited from colleagues, simulated the construction of a prototype schedule. The schedule was tested in five special classes and first revised during the early spring of 1962. Two later revisions were required in order to reduce ambiguity, duplication, and complexity in the instrument."

Blatt has made the following suggestions to potential consumers of this schedule for its improvement: "This schedule is suggested for consideration by professionally prepared and qualified special educators who are either concerned with the task of developing such an instrument for the purpose of solving a research problem or are responsible for the evaluation, supervision, and development of these programs in a community. It is possible that, as further refinements are made and more formal validity and reliability data become available, the schedule will prove worthy of the serious function for which it was designed. It is expected that conscientious 'feedback' from those experimenting with it is crucial if this goal is to be realized" (Blatt, 1966, p. 59-60).

Bibliography

Blatt, B. The Intellectually Disfranchised: Impoverished Learners and their Teachers. Boston: Dept. of Mental Health, 1966.

TEST DATA WORKSHEET

Name of Test: Conceptions of Religious Denominations (Elkind)

Author: David Elkind

Age: 6 to 12

Area (i.e. language development): Conception of Religious identity

Type of measure: Semi-clinical interviews

Source from which the test may be obtained: See reference by Elkind.

Description of Measure:
The Conceptions of Religious Denominations by Elkind consist of six structured questions which explore the child's thinking on religious identity in four areas of conceptualization "(a) a conception of the extent of membership in the childs' religious groups; (b) a conception of the external size of which religious groups members may be recognized; (c) a conception of the common property or properties shared by all religious group members, and finally (d) a conception of the possibility of multiple (non-religious group membership)."

"The method implied was the semi-clinical interview provided by Piaget .... this method requires a set of novel questions--designed to elicit spontaneous thought--which serves the setting point for an interview type discussion aimed at clarifying a child's initial responses." Examples of the items in the interview are the two questions dealing with the extent of religious group membership: (a) are you a ..... , is your family ....., are all boys and girls in the world .....? and (b) can a cat or a dog be a .....?" (Elkind, 1964).

"After giving their yes or no responses, children were asked to explain their answers. The explanations, rather than the yes and no answers, were the primary data of the study" (Elkind, 1963, p. 293).

"The results of the interview were evaluated by mean of Piaget's ... criteria for determining a true developmental sequence. These criteria were: a) uniformity of responses among children at one or more adjacent age levels; b) the presence of adherences (remnants of ideas appropriate to young children among the older children) and anticipations (suggestions of ideas appropriate to older children among the younger subjects); c) movement with age in the direction of more abstract and adult-like conceptualizations)".

In line with the discovery of definite stages in the conceptualization of religious identity among children, Elkind has discussed the implications of these findings in full for religious education (1964, pp. 645-646).

Evidence of Reliability, Validity, and Standardization:
"Close to eight hundred children participated in the investigations, and with the exception of some of the Protestant children all the subjects were individually interviewed by the writer. Of the 790 children seen, 210 were Jewish, 280 were Catholic and 300 were Congregational Protestant. The Jewish children ranged in

age from 5-11; the Catholic children from 6-12 and the Protestant children from 5-14. With the exception of the 5-6 year old subjects among the Protestant children there were at least thirty children at each age level within the age ranges indicated."

In general, Elkind discovered three differentiated stages in a child's conceptualization of religious identity. The description of the stages, together with definitions and illustrations, are given below.

a) Stage 1. At the first stage (usually age 5-7) children had only a vague undifferentiated conception of their religious identity.

Mar (5-1) What is a Protestant? 'I don't know.' Really? 'Well, maybe it's something that makes you feel happy.' Can a dog or a cat be a Protestant? 'Yes, 'Why? 'They could fight among themselves. My boy friend (sic) is a Protestant and he fights.'"

b) Stage 2. Children at the second stage (usually 7-10) had a differentiated but concrete conception of their denomination.

Kay (8-11) What is a Protestant? 'He belongs to a Protestant family' Can a dog or a cat be a Protestant? 'Yes.' How is that/ 'Sometimes they follow you to church and because it's your dog and cat.'

Pete (8-4) What is a Protestant? 'He gets bap-a-tized.' Can a dog or a cat be a Protestant? 'Yes,' Why? 'Because everybody in our house is Christian.'

Jo (7-11) What is a Protestant? 'They go to different churches.' Can a dog or a cat be a Protestant? 'No.' Why no? 'Because they don't like animals in church.'"

c) Stage 3. Third stage children (usually ages 11-14) had a differentiated and abstract conception of their ~~denomination~~.

"Faith (12-1) What is a Protestant? 'A faithful believer in God and doesn't believe in the Pope.' Can a dog or a cat be a Protestant? 'No, they are God's animals, they cannot think or/know God.'" (Elkind, 1963, pp. 294-295).

The sources of supportive data for both reliability and validity, as these had been reported by Elkind, are presented below.

1. "...a psychologically naive but highly competent English teacher ... was given a description of the criteria for the stages for each of the four questions which showed a development sequence. She then went through the records and categorized all the 300 responses to each question into three stages. The comparison of her results with those of the writer showed an agreement of better than 90 per cent for the 300 categorizations for each of the four questions." The author therefore "... assumes that the categorization by stages had considerable reliability.
2. Piaget's concept of stages"... Presupposes that any given child is at about the same stage with respect to all aspects of his conceptual knowledge. If this were not true it would be impossible to assume, as Piaget ... does, that there is an underlying mental system that determines the whole of a child's conceptual performance at any given age.

To determine whether this assumption could be supported by the results of the present day, each child was given a homogeneity score. If all his answers (to

the four questions which elicited responses categorizable by stages) were the at the same stage, he was given an A score. If all his replies but one were at the same stage, he was given a B score. If two replies were at one stage and two at another, he was given a C score. Finally, if he gave replies at all three stages, he was given a D score.

...The degree of homogeneity of responses for individual children was considerable. Only 1 per cent of the children fell in the D category and only 16 per cent fell in C category. This amount of variation could easily be accounted for by anticipations and adherences of ideas from earlier and later stages." Thus, it would appear that the stages had some construct validity. Moreover, Elkind has presented some evidence indicating that the questions were of unequal difficulty. "The unequal familarity of the questions may therefore have played some part in producing the heterogeneity of replies from indivisual children, and thus reinforcing the conclusion that for any given child the level of his responses tends to be homogeneous (Elkind, 1963, pp. 302-303).

Bibliography:

Elkind, D. The child's conception of his religious denomination: II. The Catholic Child. Journal of Genetic Psychology, 1962, 101, 185-193.

Elkind, D. The child's conception of his religious denomination: III. The Protestant Child. Journal of Genetic Psychology, 1963, 103, 291-304.

Elkind, D. The child's conception of his religious identity. Lumen Vitae, 1964, XIX, 635-646.

TEST DATA WORKSHEET

Name of Test	Muscle Tension Scale
Author	Graham, F.K., Matarazzo, R.G., and Caldwell, B.M.
Age	Infants (Neonates 1-5 days old)
Area (i.e. language development)	Measure of muscle tonus and rigidity type of rating scale
Type of measure	rating scale
Source from which the test may be obtained	Authors

Description of Measure:
"The Muscle Tension Rating was designed to measure deviations in the direction of either increased flaccidity (lessened muscle tone) or increased rigidity. However, only amount of deviation and not direction was considered in statistical treatment of the data. Five submeasures were employed in making the rating.... The five submeasures required rating the following: 1, nature (flexed or extended) of the supine position which the legs assumed spontaneously; 2, resistance to limb displacement; 3, change in muscle tone in response to being pulled to a sitting posture; 4, amount of spontaneous activity; and 5, frequency of trembling of body parts and the stimuli evoking this response.... At the end of examination, the overall rating on muscle tension was made. A five-point scale, with numerical values from 2 to 2, was designed for the rating. The zero point represented the behavior of a normal infant and the endpoints (values 2 and 2) the extremes of flaccidity and rigidity, respectively, as seen in a grossly abnormal behavior. Each of the points to the submeasures was with respect to muscle tonus and rigidity, normal muscle tone, just perceptible flaccidity, and abnormal flaccidity. An illustration of the application of this five-point scale for abnormal flaccidity. An illustration of the application of this five-point scale for abnormal rigidity is given below:

"Characteristic leg position is extended, rigid and elevated. 2, The legs resist displacement and snap back to the initial position when released. 3, The muscles are already tense and do not change tone in response to pull- to-sitting. There is no head lag. 4, There may be any amount of spontaneous activity. 5, Trembling of more than one body part if frequent and may be spontaneous or in response to mild stimuli as well as to startle stimuli."

A composite record sheet for rating irritability and muscle tension facilitates the scoring for these scales (Graham et al., 1956, pp. 13-15).

Evidence of reliability, validity, and useability:
"Reliability of the procedures was measured by split-half correlation, test-retest agreement and interscorer agreement as applicable. All tests appeared to be satisfactorily reliable" (Graham, 1956, p. 32). Split-hald reliabilities for the pain threshold scores for both traumatized and nontraumatized ranged from .82 to .97.

The other subtest scores could not be satisfactorily divided into comparable halves. Test-retest reliabilities for the Pain, Maturation, and Vision scales were from .62-.69. On the Irritability and Vision Scales, the test-retest agreements after 24 hours were respectively 75% and 86% of perfect agreement. Statistics on interscorer agreement were as follows: (1) Maturation Scale (r=.97); (2) Vision Scale, (r=.90); (3) Irritability Scale (68% perfect agreement). Finally, there was 79% perfect agreement in the Tension Scale (Graham, 1956, pp. 21-22).

In addition to the reliability of individual raw scores, statistics are provided for differentiating groups and accurate identification of individuals as members of a particular group. In the former case, statistically significant differences (p of .05 or .01) between the normal and traumatized groups occurred on all five scales based on F test, t test, and Chi square (Graham, 1956, p. 24). In the latter instance, two pediatricians rated 25 cases of traumatized infants with respect to the severity of the trauma. ". . . the extent of the agreement was high and significant as estimated by a correlation ratio of .86 (Graham, 1956, p. 18).

Bibliography of test usage:

Graham, Frances K., Matarazzo, Ruth G., and Caldwell, Bettye M. "Behavioral Differences Between Normal and Traumatized Newborns...." Psychological Monographs, 1956, 70, No. 20, (Whole No. 428).

TEST DATA WORKSHEET

Name of Test	School Survey Form (Fouracre, Rooke, and Botwin)
Author	Maurice H. Fouracre, M. Leigh Rooke, & Perry Botwin
Age	Handicapped children of school age
Area (i.e. language development)	Prevalence of handicapped children
Type of measure	Five questionnaires
Source from which the test may be obtained	(Fouracre, et al., 1959)

Description of Measure

The School Survey Form was designed to obtain the prevalence of school-aged children with anyone of the following handicaps: crippling conditions, visual impairment, speech or hearing impairment, mental retardation, emotional disorder, and special health problems. Each set of questions was grouped according to types of disabilities. Moreover, the questions were designed for non-medically trained persons. With the exception of identifying data for the child rated, the questions simply required yes or no responses. An example for the set of questions for the verification of visual impairment is as follows:

Visual impairment: (Check the following only if they apply to the child named)

25. Is there a noticeable eye defect, i.e., strabismus (crossed eyes), nystagmus (excessive eye movement), etc.?
26. Is the child blind? (20/200 vision in better eye with maximum correction)
27. Does child have useable vision? (between 20/200 and 20/70 with maximum correction)
28. If answer to No. 27 is 'yes', does child wear glasses?
29. If child wears glasses, can he read ordinary print and see objects at reasonable distance with his glasses?
30. Does child have only one useable eye?" (Fouracre, et al., 1959, p. 278).

In the use of the questionnaire, it would probably be worthwhile to study its specific limitations as presented by the authors: "The questionnaire was designed to be self explanatory, and, for this reason, a minimum of printed instructions were included. Also, each disability category did not have a question about medical diagnosis, and, therefore, this check on the teacher presumed diagnosis is limited to only several areas of disability. To some extent, these could be construed as limitations. In three sections of the form, i.e., speech, hearing and emotional problems, some question was felt to exist concerning teacher understanding of the items or of their ability to detect observable symptoms with sufficient reliability. In individual instances, teachers may not have been familiar with the visual acuity limits specified, although the availability of nurses as consultants would be expected to compensate at least partly for this lack" (Fouracre, et al., 1959, 278).

In addition, these authors have also presented some general limitations of the questionnaire: "First, the study has been restricted to pupils enrolled in school. Second, it was basically a non-medically confirmed prevalence study which would lend itself to a more reliable identification of major handicapping conditions than of secondary disabilities" (Fouracre, et al., 1959, 14 and 15).

Evidence of reliability, validity, and standardization

The authors had provided no statistical data at all for this survey form. Quite possibly, however, some content validity for the questionnaire appears from the following statements: "The form used for reporting pupils assumed by teachers to have some type of disability was developed in 1955 for purposes of a state-wide census in another state and was later revised on the basis of that experience. The questionnaire was again revised for use in this study...questions on the form were designed for the understanding of non-medically trained persons and they were grouped according to types of disabilities.

Several questions, pertaining to known medical diagnoses, also were included, thereby giving some simple checks on the possible accuracy of the referrals based on presumptive diagnoses." Moreover, "trained and experienced judgement corroborated by common agreements independently reached by the two school systems concerned has thought to have lessened somewhat the element of error..." (Fouracre, et al., 1959, 11-13).

Bibliography

Fouracre, M. H., Rooke, M. Leigh, and Botwin, P. The Report of the Study on the Educational Needs of Physically Handicapped Children in Pittsburgh, Pennsylvania 1958-1959. Pittsburgh, University of Pittsburgh School of Education, 1959.

TEST DATA WORKSHEET

Name of Test	Spache Readability Formula
Author	George D. Spache
Age	Children and adults
Area (i.e. language development)	Reading level of material
Type of measure	Estimate of difficulty level of reading materials
Source from which the test may be obtained	See reference.

Description of Measure:
This readability formula, as other similar formulas, ". . . is intended to aid in the evaluation of reading material in preparation of written material" (Spache, 1964, p. 27). The scoring system is based on the elements of sentence length, proportion of hard words (those not on the Stone List) correction factors for average sentence length, and per cent of hard words (respectively .041 and .086), and the addition of a constant (.839). The sum of these figures then gives me the estimated grade level of difficulty for the reading material analyzed. Spache cites evidence indicating that this index should be based on an average of several samples. In this connection, the research indicated that ". . . samples from the beginning and end of each chapter was least accurate, ". . . that three samples would provide an estimate precise enough for most uses, while twelve or fifteen samples from a book would give a very careful evaluation. More than fifteen samples is unwarranted in achieving a more precise estimate" (Spache, 1964, p. 127). Accuracy of these samplings was gauged by comparing their results with findings of comprehensive samplings. Moreover, the suggestion by Spache is that each of these three samples be 100 words in length (Spache, 1964, p. 125). Spache further offers a number of interrelated comments relating to the purpose, uses, values, and drawbacks of readability indices.

Purpose: "Readability formulas are needed when finer discrimination of a probable reading difficulty are sought, as in providing reading materials for young children and for poor readers particularly. Teachers need and want materials which apparently differ by small degrees of difficulty when dealing with pupils of lesser reading skill.

When books have not been evaluated by expert opinion or other methods, as in the case of new trade books or whn a variety of books list are not available to the teacher, then readability formulas are of immediate, practical service. When the teacher is doubtful about the accuracy of the publishers' grade level designations, or the text seems inappropriate for other pupils, formulas provide a quick basis for reevaluations. For the estimation of reading difficulty of such publications as bulletins, pamphlets, newspapers, and magazines for which estimates are not

commonly available from other sources, the appropriate formulas are essential" (Spache, 1964, p. 24).

Uses: Readability formulas have been used for a variety of reasons from individuals and varied studies. In the educational field, teachers have applied these formulas to all types of material and content, including reading text, lectures, tests, written material, quote publishers and writers, particularly those producing books for children, have rapidly adopted formulas for the evaluation and grading of their products." Finally, both newspapers and industry have used these formulas for the simplification of materials as an aid in the communication process.

Values: Readability formulas are valid and justified for specific purposes. The three primary purposes are for arranging material in the order of their difficulty, for determining the comprehension level required by materials, and for the simplification of materials.

Drawbacks: ". . . there are certainly literary elements in factors of reading difficulty that the formulas do not measure. They do not reflect conceptual difficulties caused by very contextual meanings of words, idiomatic expressions or the ratio of abstract and concrete terms. No formula in current use accomplishes this type of distinction although several have attempted it. Secondly, the formulas do not evaluate the organizational character of materials, the matter of presentation or the degree of explanation.... Readability formulas do not reflect the difficulty of the content. Finally in Spache's estimation, these formulas ". . . do not reflect the difficulty of the content of the reading material. Nor, Obviously, can the formula predict the reader's interest in any content" (Spache, 1964, pp. 23-26).

Evidence of reliability, validity, and standardization:
In establishing the formula, the author used the following procedure: "the elements of sentence length and the proportion of hard words were selected as most indicative of reading difficulty in primary materials. This section was based on many earlier research studies which indicated these two elements as the best predictors of readability. School books in common use in our country were analyzed in terms of these two factors. He secured 224 samples of approximately 100 words each from 152 such books. The number of samples varied from one or two in preprimers to five or more in longer books. All the books were basal readers except for 23 books drawn from social science, health, and science. Each book was assigned a grade level designation of the publisher, i.e. preprimers 1.2, primers 1.5 first readers 1.8 and 1.9, second readers 2.1 and 2.7, and third readers 3.3 and 3.7." The final step included the laborous work of counting sentence length and the hard words in these books (Spache, 1964, p. 125).

Subsequent to this development of the formula the elements of reading difficulty were correlated with each other and with grade level designations. The following correlations were obtained: Sentence length with per cent of hard words (.563), sentence length with grade level (.751), per cent of hard

words with grade level (.683). "The multiple correlation coefficient obtained by combining the sentence length and per cent of hard words in predicting grade level of books is .818."

"The accuracy of this formula compares very favorable with that obtained from other readability formulas. The probable error estimate predicting the greater level of a book by this method is 3.3 months. In other words, in half the predictions the error of estimating the grade level will be less than this amount. In the remaining predictions, the error will be greater than three months. At the primary levels, where relatively fine degrees of discrimination are most desirable, this formula evidences a high degree of accuracy and should be distinctly useful."

According to the author, the validity of the formula should be based on this accuracy in indicating ". . . the level of pupil reading ability needed to read a book successfully." He cites one study which gives some supportive evidence for such validity. In this study, the author ". . . used the actual people performances in oral reading errors and comprehension in scaling a number of primary reading selections from basal readers. After he scales these selections, he compared his ranking with that obtained by our formula. The author ". . . found a rank correlation of .70 between the two scalings and found a mark relationship between the estimate based on our formula on actual pupil performance in reading books" (Spache, 1964, pp. 125-127).

Bibliography:

Spache, George D. Good Reading For Poor Readers. Champaign, Illinois: Garrard Publishing, 1964.

APPENDIX A

1. Academy of Child Psychiatry (Journal of the)
2. Acta Psychologica
3. American Journal of Mental Deficiency
4. American Journal of Orthopsychiatry
5. American Journal of Psychiatry
6. American Journal of Psychology
7. American Journal of Sociology
8. American Psychologist
9. Australia Journal of Psychology
10. Behavioral Science
11. British Journal of Educational Psychology
12. British Journal of Educational Studies
13. British Journal of Social and Clinical Psychology
14. Canadian Education and Research Digest
15. Education
16. Educational & Psychological Measurement
17. Educational Research (British)
18. Educational Research Bulletin
19. Elementary School Journal
20. Genetic Psychology Monographs
21. Indian Journal of Social Work
22. Japanese Psychological Research
23. Journal of Abnormal Psychology
24. Journal of Clinical Psychology
25. Journal of Educational Measurement
26. Journal of Educational Psychology

27. Journal of Educational Research

28. Journal of Educational Sociology

29. Journal of the Experimental Analysis of Behavior

30. Journal of Experimental Education

31. Journal of Experimental Psychology

32. Journal of Genetic Psychology

33. Journal of Health and Human Behavior

34. Journal of Humanistic Psychology

35. Journal of Mathematical Psychology

36. Journal of Personality

37. Journal of Personality and Social Psychology

38. Journal of Psychological Studies

39. Journal of Social Psychology

40. Journal of Verbal Learning and Verbal Behavior

41. Language Learning

42. Merrill-Palmer Quarterly

43. Psychological Bulletin

44. Psychological Reports

45. Research Relating to Children

46. Smith College Studies in Social Work