programs, materials and techniques

The Measurement of Pupil Progress and Selecting Instructional Materials

Carlene Van Etten, EdS, and Glen Van Etten, EdD

Several recent events have accentuated the pressure on special educators to document the effectiveness of their teaching. While the public has become more cautious with the use of their tax dollars, the courts and legislators are being more specific in their demands. The recently passed P.L. 94-142 requires teaching accountability in the mold as did the Mills case (Mills v. Bd. of Education) and MARC case (MARC v. Commonwealth of Massachusetts). Any arguments about the need for accountability are now moot. It appears to be the law of the land.

Beyond a regard for the law and its interpretation, educators have had a continual interest in measuring the progress of their charges. In the case of LD children, it is important that this be done not solely for the purpose of evaluating the children, but rather for teacher accountability in terms of improving instruction as well. Teachers of learning disability children are faced with the reality of highly variable performance from many of their students. Therefore, teachers need some type of measurement strategy in order to know if pupils are making adequate progress toward the program objective, and if not, how to modify the program so the objective can be achieved.

This article begins by describing present measurement approaches in education with a two-dimensional model. There are two primary reasons for presenting a model of measurement. First, the model is useful because it provides a method by which a body of information can be organized and presented. But more important, this model is presented as an attempt to provide a conceptual framework which will assist teachers in selecting instructional materials with appropriate measurement procedures. There is no extensive body of knowledge available to the teacher about any aspect of instructional materials. The bulk of the literature available in this area is mostly of the guideline variety. Teacher intuition and guidelines will always be a necessary part of materials selection. However, if we are to become more scientific in the selection of instructional materials, then we must begin to operate from conceptual models. This paper presents one such model. After the measurement model is presented, we use it to analyze the measurement procedures employed in three different instructional materials: Game Oriented Activities for Learning (Karnes 1972), DISTAR (Engelmann, Osborn, and Engelmann 1969), and the Edmark Reading Program.

THE MODEL: DIMENSIONS OF MEASUREMENT

FREQUENCY OF MEASUREMENT

Historically, pupil progress has been measured at given intervals throughout the school year. The most common periods corresponded with

the issuing of grades (i.e., every six weeks, every semester, etc.). The important measure, though, was the achievement test given at the end of the year. The results were then used to determine quality of the school program. This periodic measurement may be worthwhile when assessing the quality of the school, but is of little value when programming for individual children.

These achievement tests, unit mastery tests, behavior checklists, and competency checklists can be called noncontinuous measures. Noncontinuous measures are those taken after a given amount of instructional material has been covered (a unit or chapter) or after the elapse of a given period of time (every Friday, every six weeks, every year, once in a lifetime).

Noncontinuous measurement has certain disadvantages. The teacher has no way of knowing what the behavior was like between measures. There are no data to indicate whether the response being measured is variable or consistent. Valuable instructional time can be lost because adjustments in the curriculum must wait until the prescribed time has elapsed or the prescribed material has been covered.

Noncontinuous measurement also has some advantages. Periodic measurement of pupil skills is relatively less expensive in terms of teacher time and effort. For children making acceptable progress, periodic measurement is quite appropriate.

The opposite of noncontinuous measurement is continuous measurement. In continuous measurement, the child's response is recorded every time the environment provides an opportunity for responding. In other words, every response is recorded every time it happens.

Continuous measurement has certain disadvantages, too. It requires careful attention to the learning behavior of the child. Some teachers find management of the data collection overwhelming. Many teachers are not adequately prepared for the task, since few teacher training programs incorporate methods for the collection of data in their curriculum.

The advantages in terms of precision in measuring pupil progress are obvious. Continuous measurement provides a session-by-session record of change in pupil behavior. Programs and materials may be modified *immediately* as pupil behavior changes. Accountability is easy since the data are always available.

DIRECTNESS OF MEASUREMENT

Indirect Measurement. A very common indirect measure now used in special education is the IQ test. Since intelligence cannot be measured directly, behaviors that seem to be indicative of intelligence are sampled, and levels of intelligence are inferred.

Similar assumptions are made with other indirect measures such as the Graduate Record Examination, standardized achievement tests, and many diagnostic tests. Since all possible responses cannot be measured, critical skills are sampled. Success on the sampled items is interpreted to mean that the person taking the test has mastered most of the behaviors from which the sample is drawn.

Indirect measurement data has several disadvantages for the teacher of the learning disabled child. Since not all behaviors are measured, critical skills may be absent from the child's repertoire without the teacher knowing it. Also, the behaviors measured are not necessarily the exact behaviors that are taught. Failure to make correct responses may be caused by the child not being taught to generalize, rather than an inability to learn what was presented.

Indirect measurement is currently widely used in many instructional materials. It is often used with instructional programs designed to develop specific learning processes, such as auditory memory or visual memory, or with programs which have as their purpose the development of broad concepts, such as big/little or more/less. With programs of the nature just described, the attempt is not to teach a narrowly specified skill, but to teach a comprehensive concept. When the goal is to measure generalization of broad concepts or learning processes, indirect measurement is the measurement of choice. Future developments in the field of measurement may alter this conclusion. But the present "state of the art" affirms indirect measurement as a practical

measurement approach.

Direct Measurement. Direct measurement is based on the concept that what is taught is what should be measured. If spelling was taught, spelling is tested. According to the theory, the child should be tested on the very words he was taught, using the same mode of presentation through which he learned them. The ultimate in direct measurement is to record the actual learning behavior of the pupil.

Direct measurement procedures require careful specification of what is being taught. The curriculum must be carefully analyzed and constructed to insure that all critical skills are taught. Careful specification of expected pupil behavior is essential. The teacher is accountable only for what was presented. Behaviors that are not presented are not her responsibility.

TYPES OF MEASUREMENT WITHIN THE MODEL

In order for the dimensions of measurement to be functional they must be combined and considered together (Table 1). If one element of one dimension is combined with an element of the other dimension, four types of measurement can result. These four measurement types describe all of the procedures presently being used to measure pupil behavior and pupil change.

TABLE 1. The measurement model.

Frequency of Measurement	Directness of measurement: Indirect	Directness of measurement: Direct
	Type I Indirect and Noncontinuous	*Type III* Direct and Noncontinuous
	Type II Indirect and Continuous	*Type IV* Direct and Continuous

Type I Measurement (Indirect and Noncontinuous). Type I measurement has been the most common approach of the past three or four decades. This procedure measures areas of learning, not necessarily those behaviors that have been directly taught. Every reader of this article has experienced that sinking sensation that comes when we discover that what we studied was not what was tested. Type I does not necessarily measure what we learned. It is also possible that we knew the answer at one time, or that it will be learned tomorrow. But today we do not know the correct response.

Examples of Type I measurement would include standardized achievement tests given at the end of each school year, ACT given at the end of 12 years of school attendance, Graduate Record Exams given after four or more years of college, IQ tests, and aptitude tests. These tests are given at some prescribed time and are only generally related to what has actually been taught. Also, the source of the test taker's knowledge is unknown.

Type I measurement is of little immediate value to the teacher. It does not allow her to make daily curricular decisions or to know precisely what progress the pupil is making. If this type of measurement has any value, it is to the administration only.

Type II Measurement (Indirect and Continuous). Type II involves the continuous measurement of behavior that is not being specifically taught. For example, the teacher would be measuring knowledge of sight words while teaching phonetic sounds or vice versa. Every teacher will immediately recognize this measurement as totally nonfunctional. Researchers, who are always hunting a question to answer, may take longer. In fact, they may attempt to design a study that incorporates the concept.

It may sound ludicrous to *continuously* measure what is not being taught. Why, then, do we so easily accept the *occasional* measurement of what is not being taught (Type I measurement: indirect and noncontinuous)?

Type III Measurement (Direct and Noncontinuous). This type of measurement is becoming more popular today. It has always been an integral part of classroom teaching. Every elementary teacher has taught math and then given a test to determine how many of the basic facts are answered correctly. The old familiar Friday spelling test is a common use of the Type III measurement. Students were tested on what they studied (direct) after a given interval (every Friday).

A very popular measurement concept of the Type III variety is the criterion test. This recent development in educational measurement attempts to define the criterion level to be reached before a pupil can move to the next sequential step in the program. The criterion test is direct; it measures what has been taught, often using the exact materials from the instructional program. The measurement is noncontinuous; it is taken after a given amount of instructional material has been covered by the pupil.

Other examples of direct and noncontinuous measurement include unit mastery tests, chapter review tests, skills checklists, time-ruled checklists, behavior competency checklists, inventories, etc. These instruments measure directly what is taught by recording pupil behavior, as in the case of tests, or by teacher observation, as in the case of checklists.

Type IV Measurement (Direct and Continuous). The most sensitive measurement is Type IV. Type IV measurement involves recording the response of the child every time there is an opportunity for the behavior to occur. Direct and continuous measurement actually allows the teacher to record the learning pattern and progress of the pupil. Every correct or incorrect response is recorded. It is even possible to record the sequence of correct and incorrect responding. By utilizing these very precise measurements, functional relationships between environmental events and pupil behavior can be established (Lovitt 1976, White and Liberty 1976, White and Haring 1976).

Type IV measurement requires careful management of learning and teaching programs. The topography of the response must be carefully specified. At the present time, few prepared materials utilizing Type IV measurement are available for use by classroom teachers. Lovitt and Hansen (1976) have modified basal reading programs so that direct and continuous measurement is possible. White and Haring (1976) have also provided concrete guidelines for the development of materials and programs.

In order to utilize this level of measurement, teachers must master the technology of applied behavior analysis (Lovitt 1975a,b). Teachers must devise their own data collection procedures. Hopefully, many programs utilizing this necessary precision will soon be available from commercial publishers.

MEASUREMENT STRATEGIES YOU CAN USE

We now turn to applying these types of measurement to achieve the goal of educational measurement — to determine the effect of an educational program upon the learning behavior of a student. The measurement types just described may be used in four different measurement strategies: posttest only, pre-post testing, progress testing, and continuous measurement. Within each strategy, it is possible to use more than one of the four measurement types identified in Figure 1. We discuss the most appropriate use of each strategy, since the teacher must determine which is the best approach to use in measuring particular changes in student behavior.

POSTTEST ONLY

In the posttest-only strategy, a test is given at the end of a given unit of instruction, often to measure the degree to which a body of knowledge or group of skills has been mastered. In this strategy, the evaluation is not concerned with the knowledge or skills that the learner brings to the instructional session. The only interest is that the learner leave the session with a given amount of knowledge.

This strategy is utilized by many high school and university instructors. Students are tested at the end of a given unit of time or after the completion of a given amount of work to determine to what extent they have mastered the content of the material. It is entirely possible that some of the students had mastered the content of the work before the instruction occurred. This would seem to be a major disadvantage of this strategy. Valuable learning time is lost.

A second disadvantage is also inherent in this strategy. Since measurement is taken only at the completion of a given amount of instruction, there is no way to adjust the instruction to the individual needs of a student. If mastery is not accomplished, the only alternative is to

start over at the beginning. There is no way to determine at what point the learner began to fail.

The posttest-only strategy can be used effectively. In some situations, the posttest-only strategy is often the most convenient and practical approach available. Posttest only is a useful strategy when the primary requirement is a summative statement about the final level of achievement. An excellent example is typing skill. A typing teacher might want to pretest to determine if all students are at approximately the same levels. However, the major emphasis is on final attainment of so many words per minute. Driver's training courses provide another example of emphasis on final outcome. Posttesting is of extreme importance in this instance. Other examples can be found, especially in the vocational training areas, where posttesting only is the most practical approach.

Whenever constraints of time and money limit the use of more comprehensive measurement strategies, posttest only provides at least one opportunity for examination of the learning process. When specific conditions prevail or certain subjects are being taught, the posttest-only strategy may be either necessary or the most appropriate measure to employ.

Within this strategy, either Type I or Type III measurement is possible. The instruments used are usually achievement tests, criterion tests, standardized tests, teacher-made mastery tests, and checklists. Using this strategy there is no way to determine the effect of a given program, since entering skills are not assessed. It is possible that the program had no effect and that all of the skills demonstrated were presented at the beginning.

PRE-POST STRATEGY

The pre-post strategy is a common approach used by educational researchers. It is also used by many school systems to evaluate total special education programs. A pretest is given to students upon entering a program. At the end of the program, the same test (or a different form of it) is administered. The difference between the pretest and posttest is considered to be a measure of change.

This strategy is used in some instructional programs in a slightly different manner. A pretest is given to determine the skills and knowledge that the student brings to the instructional session. This information is then used to place the student at the proper level in the program. When the pre-post strategy is used this way, it can be valuable to the teacher. The pretest provides information about the developmental or instructional level of the child at the beginning of the program. The post measure provides a summative description of the level of the child at the conclusion of the program.

The pre-post design is an efficient one for program administrators, since this type of data is easy to analyze and summarize. Teachers who are content with before and after measures only will also find this strategy appealing. This approach does not require either periodic or constant measurement on the part of the teacher. Final reporting is simplified because pupil progress can be quickly and efficiently reduced to a few tables.

If the pre-post strategy is utilized, much data along the way is lost, of course. Using only the pre-post strategy does not allow for sensitive adjustment of the instructional program to accommodate the unique learning patterns and rates of the individual child. The only adjustment that can be made, if the pre-post strategy alone is used, is to start the program over at the beginning. The precision teacher or the teacher who operates under a behavior management model would find this alternative unacceptable because of the limited data available under such a strategy.

In this strategy, Type I and Type III measurement can be used. Persons using this strategy for research purposes usually use standardized tests, while teachers may use any of the instruments that fit in Type I and Type III measurement.

PROGRESS TEST STRATEGY

The progress testing is very similar to the pre-post strategy. The difference is primarily conceptual, but the contrast is significant enough to treat progress testing as a separate strategy.

Progress test strategy usually starts with a pretest to determine the correct placement

level. A mastery test (usually a criterion test) is administered after a unit of instruction is completed to determine whether or not the learner is ready to proceed to the next unit of instruction. Failure to reach criterion either recycles the student through the same instructional unit or branches him to a remedial unit.

Some books and work texts use a measurement at the end of a chapter or unit which they call a "progress test." As the term is used in this article, these chapter tests do not meet the requirements of a progress test, since no criteria for moving to the next level of the program exists. These chapter tests are used primarily to determine what the student learned; the results are seldom used to modify the program.

The difference between the pre-post strategy and the progress strategy is related primarily to the size of the unit of instruction and the need to reach a criterion level before moving to the next instructional unit. And, the pre-post strategy measures at the end, while the progress test strategy also measures behavior several times during the program.

The progress strategy allows for a great deal more sensitivity to individual student learning styles than does either the posttest only or pre-post strategy. Program adjustments are possible on a more frequent basis. Difficulties can be noted immediately and the recycling process is more efficient.

CONSTANT RECORDING STRATEGY

Recently, several calls for more precision in measurement have been reported (Lovitt 1975a,b, Lovitt & Hansen 1976, Haring & Schiefelbusch 1976, Haring & Gentry 1976, White & Haring 1976, White & Liberty 1976). Lovitt (1975a,b) states that only through direct and daily measurement is it possible to adequately monitor the progress of learning disabled children.

At the present time, few commercial materials incorporate the constant recording strategy. Those that do are designed primarily for the moderately and severely handicapped child. (Edmark 1975, Project MORE 1973). Lovitt and Hansen (1976) in this issue of the journal (p. 481) have published their preliminary attempts to modify a reading series to this strategy. Teachers desiring to use a constant recording strategy must modify the material themselves.

Only Type II and Type IV measurements are possible within this strategy. But only Type IV is immediately applicable. Most persons presently using a constant recording strategy also demand direct measurement (Lovitt 1975a, 1976). And, as discussed earlier, Type II measurement is nonfunctional at the present time.

Several different approaches to constant recording have been used. One of the most common approaches is the recording of percent correct (or incorrect). This information is then displayed on a graph. The most obvious drawback to percentage measures is that the number of opportunities to succeed is not reflected in the data. When comparing 50% correct on two opportunities and 50% correct on 100 opportunities, the results must be interpreted very differently. Also, the length of time taken to obtain the correct answers is not evident in percentage measures.

A second approach often used is a trials-to-criterion approach (or errors-to-criterion). In trials-to-criterion measurement the teacher records the number of attempts (trials) that a child makes before reaching a predetermined standard. Trials-to-criterion is often used in the acquisition of new behaviors such as sight words and basic number facts. This type of measure allows the teacher to see if a child is developing a learning strategy. That is, does the child require fewer trials in later sessions than he did in the earlier sessions?

White and Liberty (1976) have described several other approaches to constant recording strategies including duration of responding, response latency, and fluency. Duration of responding is a measure of the length of time that a child works on a given task. The task may be one such as eye contact, on-task behavior, or out-of-seat behavior. This measure is of some special value when prolonged responses are the desired outcome.

Response latency refers to the length of time that elapses from the stimulus to the response. Some responses need to be very automatic such as sight words and basic number facts. This type of measure is best utilized when the goal is

to reduce the length of time between the presentation and the response.

Fluency (rate) is a measure of the number of correct responses in a given unit of time. Fluency is usually used for words-per-minute read correctly or for such operations as math problems added per minute correctly. This measurement can be utilized with almost any academic behavior.

Some recently published materials (Edmark 1975, Project MORE 1973) have included another aspect of constant recording: recording of teacher behavior during the teaching process. In these programs, the number and types of teacher assistance may be recorded. The materials allow for recording the number of demonstrations and prompts given by the teacher as a means for measuring pupil progress. While this relatively new approach needs considerably more attention before definite conclusions can be drawn, teachers and publishers should definitely consider it.

LOOKING AT INSTRUCTIONAL MATERIALS

The following section will examine three currently available instructional materials which employ Type I, III, and IV measurement. As mentioned previously, Type II measurement is not really functional in terms of classroom application to instructional materials. In this examination process, the primary interest will be on the measurement types and the strategies by which information is obtained.

However, the editors of this column previously indicated that they hoped to acquaint readers with many materials and to point out noteworthy features — good and not so good. With that stipulation in mind, the following section will present basic information about the total program as well as emphasizing the various aspects of measurement.

TYPE I — INDIRECT AND NONCONTINUOUS

GOAL (Game Oriented Activities for Learning) *by Merle B. Karnes, 1972. Milton Bradley Company, Springfield, Mass. ($100).*

An example of indirect and noncontinuous measurement is provided by GOAL (Game Oriented Activities for Learning). The target population for which this program was designed is children, ages three through five, in day care, Head Start, and regular nursery schools and kindergartens. Program activities are structured to be presented to a group of five to eight children. The manual suggests that children be seated at a round table where visual and auditory distractions can be minimized. This program is based on the Illinois Test of Psycholinguistic Abilities (ITPA) instructional model.

Six subgoals of the program are noted in the manual:

(1) Develop a desire to learn
(2) Develop a positive self-concept and good emotional development
(3) Develop good social skills
(4) Enhance creative and problem-solving abilities
(5) Promote motor skills
(6) Encourage teachers and parents to focus on the language development of children.

The GOAL manual states that the primary goal of the lessons is the promotion of language processing skills. The six subgoals listed above are to be promoted. It is interesting that this author has chosen the term "promote." The word promote implies a moving onward or upward, but does not indicate a specified growth level. The six subgoals listed are very global and only minimally related to the actual program content.

The GOAL program utilizes ten different types of instructional stimuli. The core of the program is the 337 model lesson plan cards. These cards are divided into eleven language processing areas. The eight- by nine-inch cards may be pulled from the storage file to be used as a guide for each lesson. The major portion of the card contains the procedure for teaching the task. Following the procedure section, on some cards, is something called criterion activity. This is an attempt to provide the teacher with a procedure to see if the skill or concept has been attained. This section is not a criterion activity in the technical sense of the word, however. No information is supplied as to when this measure should be given or how

many responses the student should be required to make.

A wide variety of instructional materials are utilized in conjunction with the lesson cards. There are two sets of picture cards which are used to stimulate discussions, for sorting and classification activities and to stimulate manual expression. Other visual stimuli in the form of situation pictures, posters, and murals are also included. Three puppets (to be used as motivational tools), games, puzzles, and templates complete the program materials. An interesting feature of the program, one which is very beneficial to users, is the specific suggestions given for the use of other appropriate materials which can be purchased separately. This feature further extends the capability of an already comprehensive program.

This program as currently presented would be useful to the teacher who needs a resource file of language development activities. A wealth of activities with accompanying materials is included in this program. It is a practical and useful program for the teacher who operates with minimal structure and is more interested in the overall creative growth of her total class rather than in the development of specific skills and responses in an individual child.

For the teacher whose primary concern is the measurable growth of individual students, this program has severe limits. The first limitation is the nature of the program goals. The terminology used in these goals is so vague that they are unmeasureable. Though it might be possible to measure some individual lessons, overall child progress cannot be precisely measured at the end of the program.

Noncontinuous measurement is utilized in this program, creating a second difficulty for the data-based teacher. The only measurement is a checklist, which is used in a pre-post strategy to measure pupil progress. The manual states the evaluation process should be ongoing, in a regular cycle, but no suggestions for establishing such a cycle are given. One difficulty with the checklist system provided is that it allows for much subjectivity on the part of the observer. The teacher must make a judgment as to whether the student is functioning below average, average, or above average. Unless the user has a good understanding of the levels of expectation in all these areas, accurate recording of child functioning may be questionable.

A third factor which would make this program inappropriate for the data-based teacher is the indirect measurement approach utilized. In many instances the tasks on the checklists do not have exact correlates in the lesson cards. The lesson cards often provide such a wide range of activities relating to any particular item on the checklists that it would be difficult to determine which lesson items should be related to a given checklist item that needs improvement.

Any teacher who relies heavily on data to move children through materials would find it difficult to use this program. The actual content of this program is well done and of great value to those interested in utilizing the ITPA model to teach language skills. The teacher who wants to utilize this approach, as well as function with a data base, would have to make many modifications to make this program meet her needs. This kind of instructional material requires hard choices on the part of the teacher. She must either forfeit good content and materials, forget precise data collection, or spend hours modifying the materials to serve both needs.

TYPE III – DIRECT AND NONCONTINUOUS

DISTAR Language I *by Siegfried Engelmann, Jean Osborn, and Therese Engelmann, 1969. Science Research Associates, Chicago, Ill. (Teacher Kit, $170; Student Material [5 students] $23.75).*

The DISTAR Language Program is an example of direct and noncontinuous measurement. The program is intended to teach basic language concepts to children of most any age and racial or ethnic group who for some reason have not learned the concepts presented. This program focuses on the language of instruction in an attempt to assist children in learning complex concepts.

DISTAR Language is the appropriate selection when the goal is to teach specific language

responses. Users of this program must accept the philosophy that language deficient children, for whatever reason, must first learn the concepts necessary for logical thought. Acceptance of this program means concentrated effort on specific responses, with social language to come later.

The material consists of presentation books, a story book, a color book, and take-homes. The presentation books contain the content to be presented to the students. Instructions are concise and specific, a great strength of this program. Instructions, keyed by various colored ink, tell the teacher what to say, where to point, the expected student response, and often the kind of correction procedure to be used.

Some of the concepts presented include: polars (opposites), prepositions, pronouns, and multiple attributes. The manual contains a good explanation of each type of concept presented, with a general "Problems and Suggestions" section devoted to each. This information is then followed in the presentation book with specific instructions to the teacher. Overall, the manual is well designed and provides easy teacher access to a great deal of information.

The storybook is used to teach color, shape, and pattern. Work begins in this book after students are well into the program. This material is not presented as a part of the regular daily sequence of lessons, but is to be taught at another time during the day.

Take-homes provide extended practice in using the skills taught in a specific lesson. They are given at the end of a lesson as a reward for hard work and the mastery of skills taught in a particular lesson. The take-homes are packaged in large, tear-apart sheets.

The DISTAR language program is not designed to develop social language skills, but to develop specific language responses. The teacher who wants to develop language in free-flowing discussion groups, by drawing out each child, will find DISTAR highly mechanical. If teacher philosophy is such that language grows out of group discussions, DISTAR must be eliminated as a possible teaching program.

Neither should this program be considered by the teacher who sets a schedule, groups children, and then remains with those scheduled groups for a specified period of time before altering the schedule. The DISTAR teacher must have a flexible schedule to accommodate not only changing groups of children, but perhaps even the addition of another group. If the teacher finds herself operating without any assistance and under time constraints, the need to regroup children may result in management problems. First, the need for formation of another group may cause a great strain on teacher time. Secondly, if the teacher is not able to regroup, she faces a management problem within the group because of highly disparate learning levels.

Though no record forms for keeping data are provided, this program employs what is essentially a direct and noncontinuous type of measurement, based on total group performance as opposed to complete individual performance. The measurement is usually direct in that test items sample the exact concepts, require the same responses, and use some of the same pictures that have appeared in the lessons. In the second test, Book A-Part 1, competency level in identity and action statements, students have had the opportunity to answer all questions asked on the test pages. The stimulus picture for the test is a picture of a boy asleep in a bed. Students have already seen this stimulus in previous lessons. Therefore, this is a direct measure of previous learning.

Measurement is noncontinuous in that oral mastery tests appear only at scheduled intervals. And, it is on the basis of this scheduled interval testing that students are moved through the program. The short tests are given individually to each child within the group. Each test specifies the criteria for either moving ahead or for recycling through a group of presentations. The manual notes that, after each round of testing, it may be necessary to regroup the children. Criteria to move the group on through the program appears at the beginning of each test. The decision to move the group forward or to recycle is based on total group performance. Though general movement through the program is made on the basis of group performance, the manual provides information on how to manage the individual student who falls behind.

The precision teacher will find the measurement procedures in this program unacceptable for two primary reasons. First, she will not be satisfied with the noncontinuous measurement. Primary measures of acquisition are taken only at scheduled intervals determined by the program. Therefore, data-based program changes can occur only at program specified intervals. A second difficulty arises because students are cycled through the program on the basis of total group functioning rather than by measures of individual responses.

A teacher with children who are making smooth and fairly adequate growth (but who still need a comprehensive program) will be attracted to the simple measurement procedures used in this program. While the measurement procedure provides a very direct assessment of pupil growth, very little teacher time is required. This combination of directness and efficiency is attractive to many behaviorally-oriented teachers.

TYPE IV – DIRECT AND CONTINUOUS MEASUREMENT

Edmark Reading Program, *1975. Edmark Associates, Bellevue, Wash. ($396).*

The Edmark Reading Program, which is a beginning reading and language development program, is an example of an instructional material which utilizes direct and continuous measurement within a program test strategy. This reading program is designed for students with a wide variety of handicaps. The manual, which is located at the front of the student record book, notes that this program has been used with preschool through adult learners.

An extremely important aspect to be considered in matching materials to learner capability is the skills the child must possess before he enters the learning program. Many currently existing programs do not specifically indicate the necessary prerequisite skills. Lack of attention to necessary prerequisites to entry may result in a loss of valuable instructional time.

There are only three stated prerequisites for learner entry into the Edmark Program. First, the student must be able to indicate with some part of his body a correct choice from a multiple choice array. The student must also be able to say back, or in the case of the deaf, sign back, the words said by the teacher. Adequate receptive language to follow trainer cues constitutes the third prerequisite.

It should be noted that the following description of the Edmark program will be based on current packaging. Users of the earlier version will find the description at variance in some places. Program content and sequence have not been changed, however.

This program consists of four types of lessons which are used to teach the 150-word basic type vocabulary, as well as comprehension of the words. Words are first taught in the Word Recognition Lesson, which can be completed in two to five minutes. Each word recognition lesson is presented in exactly the same manner. Students move on to the next lesson if the error count is four or less. Comprehension activities begin when the student has mastered five words.

When ten words have been taught, a pre-posttest is administered. This is really a test of retention over time of the ten words already taught. It also provides information about student knowledge of words to be presented. The pretest does not serve as a device to eliminate lessons, however. The manual states that all lessons should be used, for purposes of reinforcement and review. If a word is missed on a pre-posttest, the student is cycled back to repeat the necessary lesson. Detailed procedures for correcting errors on word recognition lessons appear in the teacher's instructions.

Learning to follow directions begins early in the Edmark Reading Program. Direction book lessons begin after five words have been mastered. These lessons give the student the opportunity to practice reading skills, as well as teaching comprehension and how to follow directions. The direction book lesson requires the manipulation of small illustration cards in response to printed words and phrases in the direction books. A real strength of the direction book lessons is that they represent a total departure from the procedure used in the word recognition lesson. Direction book lessons are repeated until a zero error level is achieved. The

manual contains comprehensive instructions on cuing, correction procedures, and scoring.

The picture-phrase matching lessons provide functional reading experience. The student must choose a word, phrase, or sentence to match the appropriate illustration of an object, situation, or event. As for the other types of lessons, directions for corrections and scoring procedures are well defined.

The fourth type of lesson is the storybook lesson. It provides a practice situation of the words taught in a regular reading format, and for those students who may not be quite sure, it convinces them they are really reading!

Teachers searching for a very structured and individualized approach to the acquisition of reading skills will find the Edmark Program highly desirable. This material requires one-to-one instruction. This program will also appeal to teachers who utilize reinforcement procedures as a part of their total teaching style.

Some teachers may find this program to be more mechanical than they like. If the teacher views reading as a global process, the Edmark program will seem to be just word calling, even though many concepts are presented to the students and sentences are read. This is not an appropriate choice for the teacher who views reading as an integrated part of the total language arts system.

One particularly strong feature of the Edmark program is the total packaging. Packaging becomes especially critical when several pieces are necessary for a lesson. The procedure used to package the direction book and illustration cards is an excellent example of publisher response to teacher needs and the teaching process. Here is a program with numerous components that have been packaged to make each part easily and readily accessible to the teacher.

Now let us examine specifically the measurement system this program employs. First, it can be seen that measurement is continuous. Each opportunity for student response requires a corresponding response on the part of the data recorder. The scoring sheets are set up to allow three attempts for each response. The suggested procedure for indicating student performance is as follows:

✓ — Correct answer
X — Incorrect answer
O — No response

This manual also suggests an alternate recording procedure:

A — Successful completion, no cue
B — Verbal prompt required
C — Demonstration required
D — Physical assistance required

This recording procedure infers that teacher behavior is an important factor in student performance.

The use of the A, B, C, D, recording system may be slightly confusing to the data recorder. From personal experience in using this recording system and in training others to use it, we suggest that a slight modification might be helpful. Following is a modified system:

I — Independent completion
V — Verbal Prompt
D — Demonstration
PP — Partial prompt (touching to initiate response)
FPP — Full Physical Prompt

Use of these letters provides some cues to the data recorder and seems to increase the reliability factor in the data recording process.

The measurement in the Edmark Reading Program is direct because it is the actual response to the stimulus that is recorded. From this direct and continuous data, session-by-session decisions about movement through the program are made. It is this combination of direct and continuous data that allows the teacher to intervene at the earliest possible moment to change the learning program to make it more appropriate for the learner.

This program also utilizes direct and non-continuous measurement through the strategy of pre/post tests. The measurement is direct because the behavior sampled is of the exact behavior taught. It is noncontinuous because the tests are scheduled after every ten words. It appears that the direct and noncontinuous measurement is not used as a primary measure of acquisition. The acquisition measure is provided on a daily basis by the direct and continuous measurement process.

The measurement and data collection procedures of Edmark will be especially appealing to the precision teacher and to the teacher oriented toward behavior management. Some teachers may be disappointed that rate measures are not included as a part of the measurement system. They could be added with only slight modification of the current system. Even teachers who are minimally oriented toward data collection procedures should find this system easy to manage.

A teacher who believes that teaching is primarily a process of interacting with the pupil and the material will not be attracted to Edmark measurement procedures. The collection and recording of pupil and teacher behavior may be oppressive to such a teacher. For some teachers, spontaneity is more important than the strict attention to data collection.

SUMMARY

This article presented a two-dimensional model of educational measurement. From this model, four types of measurement were derived. Several strategies for applying three of the four types of measurement to instructional programs were discussed.

The potential consumer of any major instructional program must consider carefully the data collection procedures the material employs. The more direct and the more continuous the data system, the more precise the teaching-learning process becomes. Direct and continuous data make possible early teacher intervention in the instructional program. As frequency and directness of measurement decreases, so does teacher control of pupil learning behavior. The learning disabled child is behind. Therefore, time is always at a premium. The use of more precise measurement procedures helps extend both pupil learning and teaching time. — *McCollum Elementary School, 10090 San Jacinto, NE, Albuquerque, N.M. 87112.*

REFERENCES

Edmark Reading Program. Bellevue, Wash.: Edmark Associates, 1975.

Engelmann, S., Osborn, J., and Engelmann, T.: DISTAR: Language I. Chicago: Science Research Associates, 1969.

Hall, R.V.: Part 1: Behavior modification: The measurement of behavior. In R.V. Hall: Managing Behavior. Lawrence, Kan.: H&H Enterprises, 1971.

Haring, N.G., and Gentry, N.D.: Direct and individualized instructional procedures. In N.G. Haring and R.L. Schiefelbusch (Eds.): Teaching Special Children. New York: McGraw-Hill, 1976.

Haring, N.G., and Schiefelbusch, R.L.: Teaching Special Children. Chicago: McGraw-Hill, 1976.

Karnes, M.B.: GOAL: Language Development. Springfield, Mass.: Milton Bradley, 1973.

Lovitt, T.C.: Applied behavior analysis and learning disabilities: Part I. J. Learning Disabil., 1975, 8, 432-443. (a)

Lovitt, T.C.: Applied behavior analysis and learning disabilities: Part II. J. Learning Disabil., 1975, 8, 504-518. (b)

Lovitt, T.C.: Applied behavior analysis techniques and curriculum research: Implications for instruction. In N. Haring and R.L. Schiefelbusch (Eds.): Teaching Special Children. New York: McGraw-Hill, 1976.

Lovitt, T.C., and Hansen, C.L.: The use of contingent skipping and drilling to improve oral reading and comprehension. J. Learning Disabil., 1976, 7, 481-487.

Project MORE. Lawrence, Kan.: Univ. Kansas, 1973.

White, O.R., and Haring, N.G.: Exceptional Teaching: A multimedia training package. Columbus, O.: Charles E. Merrill, 1976.

White, O.R., and Liberty, K.A.: Behavioral assessment and precise educational measurement. In N.G. Haring and R.L. Schiefelbusch (Eds.): Teaching Special Children, Chicago: McGraw-Hill, 1976.