DOCUMENT RESUME

ED 181 063 TM 009 972

AUTHOR Forsyth, Robert A.
TITLE Describing What Johnny Can Do. Iowa Testing Programs Occasional Papers, Number 17.
INSTITUTION Iowa Testing Programs, Iowa City.
PUB DATE Mar 76
NOTE 32p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; *Behavioral Objectives; *Educational Objectives; Elementary Education; Evaluation Methods; *Mastery Learning; Process Education; *Student Development; *Student Evaluation; Student Records; Teacher Responsibility

ABSTRACT

Most educational objectives can be classified as either mastery or developmental. For mastery objectives, teachers can develop satisfactory assessment instruments relatively easily and can describe student achievement as "can do" or "can't do." For developmental objectives, there are a number of evaluation methods, and student progress varies in attaining these objectives. An adequate description of achievement of these developmental objectives requires a set of reference points to identify stages of development. Teachers develop a sense of these criteria by considering the instructional materials and the performance of various groups of students. Standardized achievement tests may help the teacher assess students' levels of development. Describing student's performance in an educational program requires an assessment of what the student "can do" and an educated estimate of the student's progress in meeting important developmental objectives. (Author/MH)

IOWA TESTING PROGRAMS OCCASIONAL PAPERS

Number 17 - March 1976

Describing What Johnny Can Do

Robert A. Forsyth

Introduction[1]

Practically all educators, both defenders and critics of current practices, agree that evaluation is essential to effective education. Most educators also agree that the evaluation of student achievement is the central element in any effective evaluation program undertaken in the school. Silberman writes (1970, p. 138):

> ...evaluation [of student achievement] is an important and indeed intrinsic part of education--essential if teachers are to judge the effectiveness of their teaching, and if students are to judge what they know and what they are having trouble learning.

Ideally, an evaluation of student achievement will describe what a student (say, Johnny) can and cannot do. For example, after completing our evaluation of Johnny's achievements, we would like to make statements such as these:

(1) Johnny can recall and state the essential details of material he has read.

(2) Johnny can make inferences and draw logical conclusions concerning material he has read.

(3) Johnny can apply the Pythagorean theorem to situations he has not previously studied in class.

(4) Johnny can swim 4 laps across the pool.

(5) Johnny can evaluate the adequacy of experimental data for answering specific questions.

(6) Johnny can use proper subject-verb agreement in writing sentences.

---

[1]The author is indebted to W. E. Coffman, L. S. Feldt, H. D. Hoover, and D. R. Whitney for valuable criticisms of earlier drafts of this paper.

Throughout this paper, statements such as the above will be called "Can Do" statements. Such statements are easily made for many educational objectives. However, for some objectives, of equal or greater importance, such statements are difficult (if not impossible) to formulate. The primary purposes of this paper are (1) to identify the conditions under which it is reasonable to make "Can Do" statements and the conditions under which it is unreasonable to make such statements and (2) to examine methods that are used to describe academic achievement when it is unreasonable to make "Can Do" statements.

The Need for "Can Do" Statements

Before beginning our discussion of when the use of "Can Do" statements is appropriate, we should establish why such statements are important in the first place. "Can Do" statements aid the teacher in carrying out two important educational functions. First, they serve as guides to next steps in a sequence of educational experiences that will be of maximum benefit to the student. Secondly, they provide a basis for very explicit reports to the student and his parents of what the student has accomplished.

Of course, good teachers have always been concerned with identifying specifically what a student can do. They have used a variety of procedures to obtain the necessary data for these evaluations. Sometimes formal tests of the paper-and-pencil type have been used (e.g., the end-of-chapter tests given in some textbooks). In other instances, less formal procedures have been

used (e.g., homework assignments, classroom observations). Regardless of how the data were gathered, the basic purpose has been the same: to plan appropriate ~~educational~~ programs for students and to make accurate reports of progress.

If the current trend towards individualization of instruction continues, teachers will have an even greater need for "Can Do" statements (Hambleton, 1974). The three basic components of any individualization program are (1) instructional objectives; (2) instructional sequences; and (3) evaluation activities. Some school systems are buying and installing entire individualized programs (e.g., IPI; PLAN[2]); others are using commercially available sets of objectives and exercises (e.g., Houghton Mifflin's Individual Pupil Monitoring System in Reading, 1974). Still other systems are developing their own packages of objectives, materials, and evaluation devices. There is little doubt that the amount of time for formal evaluation activities in such programs is greater than in the more traditional ones and that there is a greater need for describing student achievement in "Can Do" terms.

A very simplified model of individualization includes the following components:

1. Give pretest for a given unit of instruction to determine if instruction is needed.
2. Provide instruction, if needed.
3. Give posttest to decide if the unit objectives have been achieved.

---

[2]Individually Prescribed Instruction (IPI), Glaser, 1970. Program for Learning in Accordance with Needs (PLAN), Flanagan, 1971.

Formalized systems of instruction using these three components force teachers to make statements describing what Johnny can and cannot do. The teacher must know what Johnny can do in order to assign him to the proper instructional units.

As noted previously, good teachers have always been concerned with knowing what each student can and cannot do. The major change created by the increased emphasis on individualization is the formalization of the evaluation system, including the development of specific instructional objectives.

As schools develop more formalized systems of evaluation, they will also probably develop more formalized reporting systems; systems that will convey to students and parents specific information about what the student can and cannot do. Obviously, the traditional letter grade does not convey information about what a student has learned as well as a list of specific statements; and, in fact, over the last few years the reporting of letter grades has decreased somewhat, particularly at the elementary level (NEA, 1970). Instead of letter grades many school systems are reporting information to the parents that attempts to show what their child can and cannot do. These "new" reporting procedures vary greatly from system to system and for different educational levels within the same system. For example, Time magazine (February 18, 1974, p. 59) reports that kindergarten through third grade students in the Dallas public schools receive "an 8 1/2 in. by 14 in. numbered sheet that looks more like a page

from a company audit than a report card." In addition, parents are "supplied a 32-page booklet called Your Child Starts School and a 28-page manual with the remarkable title Terminal Behavioral Objectives for Continuous Progression Modules in Early Childhood Education." Parents, then, can use the numbered sheet and the manual to make "Can Do" statements about their child. In contrast to this very elaborate system, other schools use relatively simple checklists. Figure 1 (page 6) represents a checklist that might be used in the language arts area.[3] Regardless of the complexity of the reporting system, the goal is the same: to indicate with some degree of specificity what Johnny can and cannot do.

Perhaps, the most ambitious effort to report to parents (and others) what students can and cannot do is occurring in the state of Oregon. Starting with the 1978 graduating class, all graduates of Oregon high schools will have to demonstrate "the competencies to function effectively on the job, as a citizen, as a learner, as a consumer, as an individual, and as a family member" before a graduation certificate is awarded (Parnell, 1974, p. 205). The specific competencies that all graduates must acquire are derived from a set of global program goals, and for each competency specific performance indicators are to be developed by local school districts. These performance indicators will then be utilized to certify whether or not Johnny has acquired the competencies for certification. The following example will help clarify this process.

---

[3] In this form "Has Acquired" would be considered equivalent to "Can Do".

PROGRESS REPORT TO PARENTS 19__-19__

Pupil_____ Language Arts Teacher_____

Basic Skills - Blank indicates skill not covered or evaluated

	Has Acquired*				Not Acquired			
1. Ability to infer the main idea of a selection								
2. Ability to identify the main parts of a story								
3. Ability to draw comparisons								
4. Ability to find specific details and imagery								
5. Ability to use vocabulary skills								
6. Ability to write proper sentences, paragraphs								
7. Ability to punctuate								
8. Spelling skills								
9. Oral reading								
10. Library skills								

*The four columns represent four marking periods.

Figure 1 Example of a Simple Reporting Form for Language Arts (Taken from Gronlund, Improving Marking and Reporting in Classroom Instruction, 1974, p. 34.)

One of the program goals in Social Responsibility education is (Graduation Requirements Task Force, 1973, p. 1):

> *Students will be able to accept responsibilities in social, economic, and political affairs.*

There are 19 competencies or course goals derived from this program goal. Two examples are (pp. 1 and 2):

(1) *Find acceptable solutions to social conflicts in everyday life.*

(2) *Predict the effect of a widespread similarity in consumer activities upon total economy.*

Two possible performance indicators given for course goal (1) are (pp. 1 and 2):

(a) *Given descriptions of everyday conflict situations, the student will observe and identify the sources of conflict.*

(b) *Given descriptions of conflict situations, the student will identify and evaluate alternative solutions.*

If the student has been certified as having performed "adequately" on a variety of such performance indicators, a graduation certificate will be issued. In other words, when Johnny graduates from an Oregon high school, we will know that he can perform certain, specified tasks.

Some Problems

Attempting to describe what Johnny can and cannot do is an important part of any educational program. While this process is conceptually simple, the actual implementation of it can be relatively complex. Before considering some of the complexities

that characterize such an approach, let us consider two kinds of situations where it is relatively easy to make "Can Do" (or "Can't Do") statements.

First, it is easy to make "Can Do" statements when our instructional objectives are so narrowly defined that basically only one evaluation exercise is possible for each objective. The following "Can Do" statements are derived from such narrowly defined objectives:

(1) Johnny can identify the letter A.

(2) Johnny can add 6 and 3.

(3) Johnny can describe the potentially harmful effects of bleach on the human body.[4]

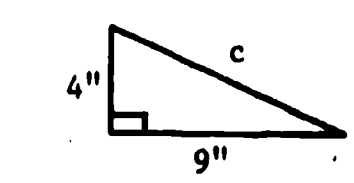
A second type of objective that allows us to make "Can Do" statements easily is one that permits an indefinitely large number

---

[4] Note that many ostensibly broader objectives are actually of this type. For example, consider the objective: Students should be able to describe the adverse effects of the 40 most commonly used chemicals on the human body. It is possible to consider this statement as representing 40 discrete objectives. If our evaluation procedures require students to describe the adverse effects of all 40 chemicals, then our interpretation of Johnny's performance would probably recognize these 40 objectives. For example, we might say that Johnny can describe the adverse effects of chemicals A, B, C, D, etc., but not the adverse effects of chemicals X, Y, and Z. On the other hand, if our evaluation procedures involve the selection of only a sample of the chemicals (say, 20), then, obviously, we would not be able to make "Can Do" type statements for all 40 chemicals. If the second type of evaluation procedure is used, then Johnny's performance would probably be reported in terms of the percent of items he answered correctly. Of course, a percent-correct score could be provided for the first type of evaluation procedure also. These percent-correct scores represent "degrees of mastery" of the objective. Sometimes, such scores are used to make instructional decisions. For example, we might say that if Johnny can describe the adverse effects of 90% of the 20 selected chemicals, he can move on to the next objective; i.e., we will, in effect, say that Johnny "Can Do" the objective.

of evaluation exercises for a specific instructional objective; however, the responses to these exercises are so highly correlated that we would be willing to make "Can Do" or "Can't Do" statements on the basis of just one or two evaluation exercises. By highly correlated we mean that if a student gets one of the exercises correct he will "very likely" get the other exercises correct also. Likewise, if he misses one item, there is a high probability that he will miss other items in the same set. For example, consider the following objective: Given the lengths of the two sides (a and b) of a right triangle, the student can use the Pythagorean theorem to find the length of the hypotenuse (c). Clearly, there is an unlimited number of evaluation exercises that could be used to measure Johnny's achievement of this objective. Two possible exercises are shown in Figure 2 below. If we are

Find the length of the hypotenuse (c) for the two right triangles shown below.



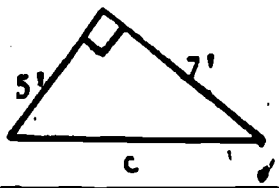


Figure 2 Exercises Used to Decide if Johnny Can Find the Length of the Hypotenuse of a Right Triangle Given the Lengths of the Sides.

interested only in Johnny's ability to substitute in the formula correctly (i.e., we are not concerned about Johnny's ability to do any of the necessary computation), then we would probably be willing to make a "Can Do" or a "Can't Do" statement for this objective on the basis of Johnny's responses to just the two exercises shown in Figure 2. Our willingness to make such statements on the basis of only two exercises is based on the assumption that Johnny's responses to

other exercises from the set of all possible exercises would be very similar to his responses to these two exercises. All such statements are, of course, tied to a particular moment in time. It is possible that Johnny may forget how to find the length of the hypotenuse. It should also be noted that if multiple choice exercises are used, more than two items may be needed because the probability that a student will get two answers correct merely by guessing may be too high.

The majority of courses (or units of instruction) contain a large core of objectives similar to those implied above. The evaluation exercises for such objectives are not difficult to construct. Therefore, for a large number of objectives in most courses it will be relatively easy to describe Johnny's performance in "Can Do" and "Can't Do" terms.

In addition to the type of objectives identified above, most courses also have objectives for which it is relatively difficult to describe Johnny's performance as either "Can Do" or "Can't Do". Consider the following instructional objectives (Gronlund, 1973, p. 17):[5]

Students should be able to

(1) Apply concepts and principles to new situations.

(2) Demonstrate mathematical reasoning ability.

(3) Write creative stories.

(4) Demonstrate critical thinking skills.

(5) Use a scientific approach in solving problems.

---

[5]The first seven objectives are given in Gronlund. The eighth objective is a very common objective found in most courses involving reading.

(6) Perform skillfully on a musical instrument.

(7) Evaluate the adequacy of a given experiment.

(8) Make valid inferences and draw logical conclusions concerning reading materials.

Most educators would agree that objectives similar to these are important objectives. Furthermore, similar objectives appear at all levels of instruction [e.g., consider objectives (1), (3), (4), and (8)]. For objectives of this type it is difficult, if not impossible, to describe Johnny's performance in "Can Do" or "Can't Do" terms, unless certain very restrictive conditions are imposed.

To illustrate some of the problems involved when we attempt to make "Can Do" statements for such objectives, consider the eighth objective above: Make valid inferences and draw logical conclusions concerning material read. Superficially, making "Can Do" statements related to this objective may seem like a simple task. After all, Johnny either can make inferences and draw logical conclusions or he cannot. However, there are two fundamental reasons why it is difficult to describe Johnny's performance on this objective in "Can Do" or "Can't Do" terms. The first is related to the global nature of such objectives and the second to the developmental nature of them. Each of these reasons is discussed in some detail below.

It is fairly obvious that the inference objective, as stated, represents a broad domain. For example, no limitations on the type of reading material is given. While this objective may be

appropriate at a number of different educational levels, we certainly don't expect third grade students, in general, to be able to make inferences and draw logical conclusions on the basis of reading material intended for high school students. However, even if we put certain reasonable restrictions on the reading material,[6] it is still possible to conceptualize the possibility of a very large number of assessment exercises for this objective. The fact that a large number of items exists does not necessarily restrict our ability to make "Can Do" statements. Previously, we indicated that if the responses to items were highly correlated then "Can Do" statements could easily be made. However, for objectives such as the inference objective, responses may not be highly correlated. That is, if Johnny can draw one valid inference, it does not follow that he will be able to draw all valid inferences. As an illustration of this problem, consider the following questions taken from the Iowa Tests of Educational Development:[7]

1. From his manner and formal training, what opinion might people have formed of John Marshall? (28%)

2. What do the last two sentences suggest about Patasonian's acceptance of U.S. aid? (44%)

3. Suppose an uninsured and unemployed motorist damaged someone's car. Which speaker offers a plan that would allow the injured party to collect benefits? (64%)

---

[6]At the third grade level, we might, for example, require the reading material to be "similar" to that contained in the textbook; at the high school level we might require the reading material to be newspaper stories.

[7]Iowa Testing Programs. Iowa Tests of Educational Development. Ability to Interpret Reading Material in the Social Studies, Form Y-5, University of Iowa, 1970.

Each of these items is associated with a particular reading passage and all three items require the student to reach a conclusion on the basis of the information in the passage. The percents in parentheses after each question show the percent of 10th grade students from a statewide sample in Iowa who answered the item correctly. The varying percents associated with the three questions above provide evidence that getting one item correct does not guarantee that a second item measuring the same objective (where the objective is defined in fairly broad terms) will be answered correctly. Furthermore, though this is not discernible from the data given above, all 28% who got the first item correct did not get the second or third items correct.

One partial solution to this problem (i.e., lack of correlation between responses of two exercises for the same objective) is achieved by defining the objective more narrowly. For example, we might specify the following as part of the inference objective:

(1) The number of words allowed in the reading material.

(2) The maximum sentence length.

(3) The maximum and minimum number of sentences.

(4) The minimum percent of words that must appear on some specified word list (e.g., EDL Research and Information Bulletin #5 A Revised Core Vocabulary, 2, Taylor, et al., 1969).

(5) The specific content area used for the reading passage.

(6) The nature of the test questions.

Now, our "Can Do" statement might be as follows:

> Johnny can make inferences and draw logical conclusions when given reading material that has no more than 200 words (at least 85% of the words are from the EDL list), that has between 10 and 20 sentences (with a maximum sentence length of 20 words), that is based on the life of John Marshall, and that requires him to answer multiple choice questions using four alternatives.

The limit of such specificity is reached when the evaluation exercise itself defines the objective. In these instances, our "Can Do" statements become so specific to the situation that they are of little use to teachers, students, and parents.

The second fundamental reason why it is difficult to make "Can Do" statements for this type of objective is related to the developmental nature of these objectives. Gronlund states (1973, p. 17):

> ...students cannot be expected to fully achieve such objectives. Even the simplest of these ... is a matter of degree and can be continuously developed throughout life. All we can reasonably do for a particular course or unit of instruction is to identify a sample of specific learning outcomes that represent degrees of progress toward the objectives.

We shall illustrate the concepts embodied in the above quote with an example. Consider the first statement in Figure 1. Stated as an objective, this becomes:

> Students should be able to infer the main idea of a selection.

To assess Johnny's performance on this objective, we might give a paper-and-pencil test to Johnny and the other students. This test might involve reading selections from a variety of content areas

and would presumably be at the "appropriate grade level"[8] for the students. We will assume that these results will be used to check either the "Has Acquired" (Can Do) or the "Not Acquired" (Can't Do) column on the report form shown in Figure 1. However, in our efforts to make these "Can Do" and "Can't Do" statements on the basis of the test results, we will immediately face the question: How many correct exercises are needed before we should say that Johnny can infer the main idea of a selection? Logically, we might contend that Johnny "Can Do" this objective only if he answers all items correctly. If this criterion was used, it is highly probable that in many classrooms the label "Can Do" would NOT be given to any student. Moreover, if such a test were given it would become immediately apparent that the ability to infer the main idea of a selection is not an "all or none" ability. Everyone, except possibly the severly mentally retarded, seems to have "some" of this ability. If we choose to report information in a "Can Do" fashion, how do we decide who "Can Do"? What criteria do we use to select that arbitrary point which discriminates between those who "Can Do" and those who "Can't Do"? Surely, when we put the check mark under "Has Acquired" on Johnny's report, we do not mean that Johnny can do this for all possible selections regardless of the type of material (e.g., Shakespeare, comic books, newspapers, magazines). What really lies behind the check is the contention

---

[8]Of course, determining "appropriate grade levels" may create problems. These problems will be discussed in the next section.

that Johnny has acquired "enough" of the skill at this moment in time to satisfy the teacher's criterion of mastery (perhaps, 80% correct on a set of exercises).[9] Likewise, when we check "Not Acquired" we only mean that Johnny does not have "enough" of the skill.

As a final illustration of the problems encountered when we attempt to make "Can Do" statements for certain objectives, consider the previously cited goal (objective) for students of Oregon high schools:

> *Find acceptable solutions to social conflicts in everyday life.*[10]

Note that the number of potential evaluation exercises for this objective is numerous. Moreover, it is not difficult to believe that Johnny might be able to find "acceptable" solutions in some conflict situations and not others. Oregon students will exhibit varying amounts of the "ability to find acceptable solutions to social conflicts in everyday life." The achievement of this objective is not an "all or none" matter. Thus, some arbitrary level of achievement will have to be selected to decide when the "Can Do" criterion has been satisfied. The validity of this arbitrary level could be established by showing that students above

---

[9] It is interesting to speculate on the parents' reaction to Johnny's report if he gets a check under "Has Acquired" for the first marking period and a check under "Not Acquired" for the fourth period. Should they conclude that Johnny has acquired this ability, then lost it?

[10] At this time we will not be concerned with defining what is meant by "acceptable". Obviously, in the implementation of the Oregon program such definitions become an important part of the process. Nor, are we concerned about the unrealistic nature of the objective.

this level "function effectively" as a member of society (the ultimate goal), and those below it do not. Of course, defining "function effectively" may be as difficult as determining the arbitrary cutting score for certification.

Another perspective on the problems involved in describing performance in "Can Do" or "Can't Do" terms is provided by Gronlund (1973). He distinguishes between two different types of learning: (1) mastery learning and (2) developmental learning.

> While learning at the mastery level is concerned primarily with simple knowledge outcomes (e.g., knowledge of terms) and basic skills (e.g., computation, grammar) learning at the developmental level is concerned with complex types of achievement. That is, achievement that goes beyond the simple remembering of learned material or the repetition of previously learned skills (p. 17).

In general, the types of objectives discussed previously fit nicely into these two categories. Those objectives for which it is easy to make "Can Do" statements occur at the mastery level. Those objectives for which it is difficult to make "Can Do" statements occur at the developmental level.

The factors which make it difficult to make "Can Do" statements for objectives at the developmental level are summarized rather nicely by Gronlund in the following statement (1973, p. 16):[11]

> The difficulty of [making "Can Do" statements] at the developmental level (i.e., learning beyond the minimum essentials) results to a large extent from the fact that ... the learning outcomes are complex (e.g., understanding, thinking skills), the domain of learning tasks is

---

[11] We have taken the liberty of substituting the phrase, "making 'Can Do' statements," for Gronlund's original phrase, "using criterion-referenced testing." This substitution has not altered the meaning of the quotation.

> virtually unlimited, ... the instructional objectives represent goals to work toward rather than goals to be fully achieved, for here the emphasis is on the continuous development of understanding and skill. Each student is encouraged to strive for the maximum level of achievement and excellence of which he is capable, rather than the mastery of some predetermined set of minimum essentials ...

In summary, for a large number of very important objectives in any curriculum area, it is impossible to make meaningful "Can Do" statements about Johnny. For these objectives the domain of assessment tasks is extremely large and the responses to such tasks are not highly correlated. Johnny may be able to perform "adequately" on some of the tasks and not on others. Furthermore, Johnny is not expected to fully accomplish these objectives at any particular moment in time. The learning of these objectives is usually viewed as a developmental process. Thus, it is reasonable to conceive of varying degrees of achievement of the objective, and the purpose of our educational program is to help students obtain "more" of the objective. When educators do make "Can Do" statements related to this type of objective, it should be clear that they mean only that a student has reached an arbitrary, predetermined level of achievement. Such statements may have some merit. These levels may have been identified primarily to enable teachers to make instructional decisions (or, in the Oregon situation, a certification decision). For example, if Johnny gets 80% correct on a test measuring a given objective, he

can go on to the next objective.[12] However, "Can Do" statements for objectives for which the student will show continued growth are somewhat deceptive. There will always be a degree of artificiality about such statements, since students will exhibit varying degrees of the ability and an arbitrary point on the measurement scale must be selected.[13]

---

[12]Teachers may also find it convenient to establish levels of achievement for objectives which are not of the developmental type. Consider, for example, the objective given in footnote 4: Students should be able to describe the adverse effects of the 40 most commonly used chemicals on the human body. As noted previously, this objective could be considered as 40 discrete objectives.

In footnote 4, we described two different evaluation procedures that could be used for measuring student achievement on this objective. For either procedure, we noted that a percent-correct score might be given and that instructional decisions might be based on that score. Percent-correct scores are most meaningful when there is a limited number of evaluation exercises that can be asked and the teacher can either ask all the exercises or select (preferably at random) a sample of the exercises. In such situations percent-correct scores represent estimates of the "degree of mastery" of the objective. Note that percent-correct scores from a set of evaluation exercises for the developmental objectives we have been discussing would not have much interpretive value, since the domain of such exercises is not well defined. Even when the domain is well defined, the identification of that arbitrary level of achievement which must be met before the student is labeled a "master" (i.e., can do the objective) involves problems which are somewhat similar to those discussed in relation to setting performance standards for developmental objectives.

[13]It may be possible for certain objectives to show that such a cutting point is reasonable. For example, it may be true that students above a certain score point experience success in some other endeavor and those below this point experience failure in this endeavor. However, situations with such "nice" cutting points are extremely rare. Note, however, that in the Oregon situation it seems imperative that these "nice" cutting points be identified.

Given that such "Can Do" statements are not entirely satisfactory, what kinds of statements can we make to provide a more adequate description of Johnny's capabilities? Is it reasonable to attempt to provide a description in terms of Johnny's location on a scale of development? In the next section we examine these questions.

Describing Johnny's Level of Development

Teachers frequently make statements that seem to convey an assessment of Johnny's level of development. For example, it would not be unusual to hear statements similar to the following: "Johnny is reading at the third grade level." It seems obvious that such a statement is intended to provide a description of where Johnny is in his reading development. On what basis was this statement made? How should we interpret this statement?

Before one can interpret the statement, "Johnny is reading at the third grade level," two basic questions must be answered: (1) What is "reading"? and (2) What is "third grade level"? Assume that by reading we merely mean that when given a set of written materials, Johnny can recognize the words and say them aloud. Also, assume that if Johnny can read materials intended for third grade students with less than five errors (e.g., one type of error would be the omission of a word) for every 100 words, then we will say that Johnny can read at the third grade level.

This procedure appears simple enough; however, there are problems. First, we have assumed a very restricted definition of reading. Our definition does not include any component related to the comprehension of the material read. Some people might not agree with our definition. Johnny may be able to recognize words but not "understand" what those words mean when put together into sentences and paragraphs. Nonetheless, given our definition, the meaning of the statement about Johnny's reading level is partially clarified.

A second problem concerns the definition of third grade materials. How do we know that the reading materials are at a third grade level? One very common method for deciding on the reading level of materials is to accept the level designated by the publisher. Thus, if a school is using reading series WYZ, and if, by some objective criteria, Johnny can "read" the material in the third grade book, but not materials in the fourth grade book, we will say Johnny is "reading" at the third grade level. However, the criteria that the publisher uses to decide what is third grade material are usually not clearly defined. In fact, if a different reading series is used, Johnny's reading level may change.

Another method that could be used to determine the reading level of materials involves actually having various groups of students read the materials. The students could be grouped by grade level. Then, the lowest grade in which 80% (or, some other agreed upon

percent) of all student can read the materials (by our previous definition) is the grade level of the materials. Thus, in effect, a scale of reading materials is developed which is based on the capabilities of several different groups of people. The values along this scale (e.g., third grade materials, fourth grade materials) have meaning only because various groups have provided normative data. In essence, we have provided reference points along a scale. Notice that if either of the above two procedures are used, then the statement, "Johnny can read at the third grade level," has meaning only because of the normative framework which has been established. In the first procedure the normative framework is the textbook series, and in the second procedure it is the various groups of students.

Regardless of how "third grade reading materials" is defined, the statement, "Johnny can read at third grade level," conveys little meaning unless samples of the reading materials are supplied or the domain of reading materials is well known. Thus, unlike the "Can Do" statements at the beginning of this paper [e.g., "Johnny can use the Pythagorean theorem to find the length of the hypotenuse given the length of the two sides."], this type of "Can Do" statement cannot be interpreted without additional information. Telling parents that Johnny can read at the third grade level seems to convey more information than it really does.

The preceding example serves to illustrate the need for reference points (i.e., a normative framework) when we are trying to

make statements about Johnny's level of development. Since many of our educational objectives are of a developmental nature, such reference points can be useful for charting Johnny's growth. The remaining part of this section examines how teachers probably establish these reference points.

Possibly the most common set of reference points used in the elementary and junior high schools is: below grade level; at grade level; and above grade level. The process used to establish these points essentially consists of two steps: (1) selection of a set of evaluation activities and (2) development of a set of performance standards. To illustrate this process, again consider the first skill in Figure 1: Ability to infer the main idea of a selection. For the evaluation activities the teacher might decide to use the tests accompanying the third grade reading textbook and the results from the class discussions of the reading selections in the text Thus, Johnny's development on this objective will be assessed by his performance on a written test and by his oral responses in the classroom.

The development of a set of performance standards which will enable the teacher to decide whether Johnny is below, at, or above grade level is not easy. Johnny's "score" (i.e., below, at, or above) probably will be obtained by comparing Johnny's performance with the third grade students the teacher has had. For example, if Johnny is similar to the "average" student, then Johnny's "score" will probably be "at grade level." If the textbook authors provide suggestions for interpreting the results of their tests, such

information may be used by the teacher to help define Johnny's "score" on this objective. Thus, in general, reference points for describing Johnny's development usually depend on a consideration of the instructional materials being used and the performance of vaguely defined groups of students.

If a teacher wishes to convey this type of information to parents, and if the form shown in Figure 1 is used, the teacher would probably check a box under "Has Acquired" when Johnny has been identified as at or above grade level. Likewise, if Johnny is below grade level, the box under "Not Acquired" would probably be checked. Of course, given our previous comments, such labeling creates a somewhat distorted picture of Johnny's achievement.

It may be informative to consider just what information such a procedure provides. Note first that this procedure differs only slightly from the traditional method of assigning letter grades. In fact, it can be contended that checking the "Has Acquired" box really means that Johnny has received an A, B, or C, while checking the "Not Acquired" box really means Johnny has received a D or F. In essence, a five point reporting scale has been collapsed into two points. The major difference between this method and the traditional method occurs not in how the "score" ("grade") is determined but rather in the number of "scores" supplied. For example, in the reporting form shown in Figure 1, we report 11 "scores"; whereas, previously Johnny may have received only one "score" in

language arts. In the language of educational testing, we have provided subtest scores instead of a composite score.

Secondly, note that, as with the reading level example above, the report that Johnny "Has Acquired" or "Not Acquired" certain skills can be interpreted only within the context of additional information. This additional information would be the evaluation materials, or facsimiles of such evaluation materials. With such materials parents would have a better understanding of Johnny's achievement in a given area.[14]

As indicated earlier, the above approach is probably the most widely used method of describing Johnny's development (i.e., identifying reference points) on a variety of objectives. In most schools, such a method is the only feasible procedure for decision-making in the day-to-day operation of the instructional program. It does, however, have several weaknesses which may limit its usefulness for describing Johnny's development to other interested persons (e.g., parents, other teachers). Foremost among these limitations is the extreme subjectivity of the description. How valid are the performance standards set by the teacher? How comprehensive are the materials used to evaluate the achievement? How representative are the students this teacher has had? This approach depends greatly on the instructional materials being used and the experiences of the teacher. Johnny's performance may be

---

[14]Some parent-teacher conferences involve an examination of such materials.

judged "third grade level" by one teacher and "fourth grade level" by another teacher in a different setting.[15]

These limitations may be less critical if Johnny is given a standardized achievement test on a fairly regular basis (e.g., once a year). A major purpose of many standardized achievement tests is to provide reference points for assessing pupils' educational development. Such tests supply developmental scales based on the responses of well-defined reference groups. For example, they may provide grade equivalent scores for different content areas (e.g., reading comprehension). Grade equivalent scales consist of reference points defined by the scores of students in various grades for specific populations (e.g., students in U.S., students in Iowa).[16] Thus, Johnny's grade equivalent score of 3.1 on a reading comprehension test means that Johnny's test performance is similar to the "average" score of third grade students (in some reference population) who took the test during the first month of school.

Well-constructed standardized tests can add a valuable dimension to the teacher's ability to describe Johnny's level of development. In essence, these tests furnish the teacher with a broader perspective of Johnny's abilities than do the sometimes small, and often non-representative, sample of students who are in his (her) classes. Such tests also supply a continuous scale of

---

[15] It is interesting to note that these same limitations have been raised with respect to the use of letter grades (Ebel, 1974).

[16] The basic assumption of such scales is that students in different grades are, on the "average", at different stages of development. This assumption does not seem unreasonable.

development for charting Johnny's growth over many years (e.g., grades 3-8). These developmental scales would be extremely difficult to develop at the local level. In addition, standardized tests usually focus on important developmental objectives within the various curriculum areas[17] and presumably measure the student's ability to transfer the ideas developed in the classroom.

As we have previously noted, for developmental objectives the interpretation of a score (e.g., "Has Acquired") is improved if the evaluation materials (or facsimiles) are provided; this is also true for scores from standardized tests. The fact that Johnny made a grade equivalent score of 3.1 on a reading comprehension test provides some information about Johnny's level of development. The interpretation of this score is enhanced if we know the kinds of questions that were asked and how many Johnny answered correctly.

In summary, the results from standardized achievement tests provide an additional source of data for assessing Johnny's development--data that supplements the other information available to the teacher. The major advantages of such tests are the developmental scales and the external "look" they provide.[18]

---

[17] Obviously, such tests cannot sample all the important objectives within a curriculum area.

[18] These brief remarks on standardized achievement tests are intended only to point out the potential that such tests have for providing reference points that may help describe a student's development in various areas. Whether or not a particular achievement test can validly be used for this purpose, must, of course, be considered by the educators using the test. Also, it is important to emphasize that standardized tests do not provide standards for development. They merely provide a type of normative information that is not otherwise readily available to teachers.

A Summary Statement

The majority of educational programs (courses, units of instruction) contain objectives that can be classified as either mastery or developmental. For mastery type objectives it is a relatively easy task for the individual teacher to develop satisfactory measuring instruments and to describe Johnny's achievements in "Can Do" or "Can't Do" terms. However, for developmental objectives, the number of possible evaluation exercises is large and at any particular time Johnny may be able to give correct answers to some exercises and not to others. Basically, developmental objectives represent "goals to work toward rather than goals to be fully achieved." (Gronlund, 1973, p. 16).

An adequate description of Johnny's achievements for these developmental objectives requires a set of reference points which identify stages of development. Such reference points are not easily derived. The typical teacher develops a sense of these points after considering the instructional materials and the performance of various groups of students with which the teacher is familiar. The results of standardized achievement tests may help the teacher assess Johnny's level of development. Many standardized achievement tests provide developmental scales (for fairly broad content areas) based on the performance of well-defined groups of students.

Developmental objectives represent important components of most educational programs. Describing Johnny's performance in a

given educational program requires not only a description of what Johnny "Can Do" but also our best estimates of Johnny's progress toward important developmental objectives.

References

Ebel, R. L. Shall We Get Rid of Grades? NCME Measurement in Education, 1974, 5, 1-5.

Flanagan, J. C. The PLAN system for Individualizing Education. NCME Measurement in Education, 1971, 2, 1-8.

Glaser, R. Adapting the elementary school curriculum to individual performance. In Procedings of the 1967 Invitational Conference on Testing Problems, Princeton, N. J.: Educational Testing Service, 1968.

Graduation Requirements Task Force. Oregon Graduation Requirements: Models and Guidelines for Social Responsibility Education. Oregon State Department of Education, 1973.

Gronlund, N. E. Preparing Criterion Referenced Tests for Classroom Instruction New York: The Macmillan Company, 1973.

Gronlund, N. E. Improving Marking and Reporting in Classroom Instruction New York: The Macmillan Company, 1974.

Hambleton, R. N. Testing and Decision-Making Procedures for Selected Individualized Instructional Programs. Review of Educational Research, 1974, 44, 371-400.

National Educational Association. Marking and Reporting Pupil Progress. Research Summary 1970 S-1 Washington D. C.: NEA Research Division, 1970.

Parnell, D. The Oregon Walkabout. Phi Delta Kappan, 1974, 56, 205-206.

Silberman, C. E. Crisis in the Classroom New York: Random House, 1970.

Taylor, S. E., Frackenpohl, H., and White C. E. A Revised Core Vocabulary: A Basic Vocabulary for Grades 1-8. An Advanced Vocabulary for Grades 9-13. Research and Information Bulletin No. 5 (Revised) New York: EDL/McGraw Hill, 1969.

Time. The Dallas Monster. February 18, 1974, 59.