C. Bende
342

EVALUATING EDUCATIONAL PROGRESS

by

Owen R. White

Introduction

It is the purpose of education to provide each child with those experiences which will foster the development of behavioral patterns and skills requisite to the highest possible degree of independent and productive action in the adult community. In short, it is the job of education to change the behavior of learners from, for example, non-readers to readers, or from destructive or disruptive influences to "socially adapted" individuals. In this basic concern, there is very little difference between the fields of education and clinical psychology. Specific targets of concern might vary, and certain intervention or instructional procedures might differ, but to the extent that both education and psychology are interested in behavioral change, their technologies should have certain basic approaches in common. This chapter has been prepared in hopes that an investigation of the latest educational assessment procedures might foster a type of cross-disciplinary interaction of interest and benefit to both the psychological and educational communities. Of course, in a chapter of such limited length, only the briefest of introductions into the procedures discussed can be undertaken. If, however, an interest in further investigation and discussion can be generated, this chapter will have accomplished its purpose.

This material represents a draft of a chapter which will appear in Behavioral Assessment: New Directions in Clinical Psychology, Cone, J.D. and Hawkins, R.P. (eds.), New York: Brunner-Mazel, in press.

<u>The Movement Cycle: Target of Assessment</u>

It is axiomatic in any behavioral approach that reliable and easily interpreted assessments must be based on the measurement of directly observable behavioral events. In the experimental analysis of behavior, where most studies are conducted under highly controlled and complex system-supported laboratory conditions, the precise specification of behavior is rarely a problem (e.g., Stiers, et al, 1974; Chrisman, 1974; Wright and Nevin, 1974; Flory, et al, 1974; Gonzalez, et al, 1974; Selekman and Meehan, 1974; Cohen, 1975, Baum, 1975, Rashotte, et al, 1975). With the application of behavioral technology to education, however, special guidelines were found necessary to assist the teacher in identifying appropriate behavioral targets which are easily measured and amenable to educational intervention. Lindsley (1964) was the first to formulate a set of just such guidelines with the teacher in mind. Since that time, those guidelines have been expanded, modified, and redefined several times (e.g., White and Liberty, 1974; White and Haring, in press) and have proven of tremendous help to teachers in selecting appropriate behavioral targets. Briefly, those guidelines are:

(1) <u>The behavior must involve some directly observable movement</u>. This insures that the target is, in fact, behavior in the original sense of the term (i.e., any transposition of the organism or part thereof through space, White, 1971); and avoids the problems of working with phenomena which represent the absence of behavior (e.g., "sitting still").

(2) <u>The behavior must have a definite cycle, with a clearly defined beginning and end.</u> Being able to precisely define the beginning and end of a behavioral event greatly facilitates the collection of precise and reliable data. This criterion, for example, forces the teacher to define the difference between "disruptive episodes" and "disruptions."

(3) <u>The behavior must be repeatable in easily identified instances of uniform importance.</u> The behavior must be repeatable, of course, if it is to be changed. The rationale for uniform importance of each repetition, however, is somewhat less obvious. Essentially, the logic goes like this: If each repetition of a behavior is of uniform importance, then a simple count of its occurrence will suffice for behavioral assessment. If, on the other hand, the importance of each repetition changes (e.g., as a function of its duration, latency, or some other behavioral dimension), then a simple behavioral count will <u>not</u> suffice to define the progress of the child. Given that simple behavioral counts are the easiest form of data to collect in an educational situation, it also follows that assessments based on that form of data will be more reliable and take less of the teacher's time away from instructional activities. In order to pinpoint behaviors which are of equal educational importance with each repetition, it is necessary to "calibrate" the movements of children (White and Liberty, 1974; White and Haring, in press). For

example, instead of counting answers written in response to addition problems (which might vary in length), the teacher will count the number of digits in each answer -- thereby calibrating problem length, at least in terms of the actual physical requirements for responding to each problem.

In reflection of the first two criterion for acceptance as a behavioral target, any behavior which meets all of the requirements specified above is called a movement cycle. Experience has shown that well pinpointed and calibrated movement cycles serve as excellent foci for instructional activities (concentrating the efforts of the teacher on behaviors which are, in fact, amenable to instructional intervention) and lead to more precise and easily interpreted assessments.

Assessing the Movement Cycle: Behavioral Counts

As mentioned earlier, one of the prime reasons for the specification of instructional targets in terms of movement cycles is the facilitation of assessment. If each instance of the behavioral phenomena is, by definition, of equal educational importance, then assessments can be based on simple behavioral counts. If a child reads more words today (under equivalent conditions) than he read the day before, then some assumption of progress is justified. Behavioral counts might be rather restricted (e.g., counts of correct and error words only), or more descriptive (e.g., breaking "errors" down into counts of omissions, substitutions, mispronounciations, insertions, and repetitions), depending upon the information needs of the teacher. In either case, however, the results of assessment are immediately interpretable and directly related to the behavioral phenomena in question. By specifying how often a behavior should occur during each assessment, and then keeping track of the actual behavior counts from day to day, the progress of the child, and effects of various intervention strategies are easily documented.

Other forms of data, e.g., durations or latencies, have also proven of some utility in monitoring the development of certain behaviors, but have been avoided in education for a very simple reason: they are more difficult to collect. In order for a teacher to collect duration data, for example, each and every instance of the behavior must be timed (usually with a stopwatch). If the behavior has been defined in a manner which allows simple behavior counts to suffice, on the other hand, it is usually possible to provide the child with a number of opportunities to move (e.g., a math-fact sheet) and then tally the results at some later time. If the movement is not dependent upon opportunities provided by the teacher (e.g., "out of seat" behavior), the advantages of a behavior count are even greater. It is likely, for example, that a teacher will notice that a child is out of his seat sometime during the out of seat episode -- thus prompting a tally of the behavior to be made. It is less likely that a teacher will notice exactly when the behavior was initiated and terminated -- obviating the possibility of reliable duration statements. That is not to say that duration statements would not be of some value (see, for example, Walker, 1968), only that those data would have a lower probability of being accurate, and therefore, could mislead a teacher in educational decisions.

Research to date has indicated that assessments based on simple behavioral counts tend to be quite reliable. In one case, for example, teachers involved in classes with up to 27 pupils were able to monitor behaviors on several students at once with 84 to 100% reliability when compared with trained professional observers (Hall, 1971). Self-recording and peer-recording studies have also demonstrated the potential advantages of this data type (e.g., Broden, et al, 1971; Risley and Hart, 1968). The implication for clinical psychology, of course, is that the client and/or another untrained observer (spouse or parent) might be expected to collect reasonably accurate behavioral data on the performance of the client outside of the counseling situation. If the required data were something more complex than simple behavioral counts, however, the data might be more misleading than helpful.

Adjustments in Counts

Simple behavioral counts only lend themselves to valid comparison when collected under equivalent conditions from one day to the next. In many situations, however, either the time allowed for assessment or the number of opportunities for the behavior to occur will change from day to day. To correct for these inconsistencies, and to provide a "universal base" against which the results of our observations might be interpreted by others, percentage or rate statements have often been employed to "adjust" behavioral counts.

Percentage statements are most commonly employed to correct for differences in opportunity for the behavior. If, for example, a child completes five out of ten items on one day and nine out of twenty items on the next, the behavior counts per se would lead one to believe that the child has improved. Since there were more opportunities for the behavior to occur on the second day, however, the results can be misleading. Percentage statements hold behavior counts relative to a hypothetical 100 opportunities, i.e., if the child had been given 100 opportunities for the behavior, how many movements might the child have made? For the first day, $(5 \times 100) \div 10 = 50\%$; and for the second day, $(9 \times 100) \div 20 = 45\%$. It is now apparent that the child actually performed a little more poorly on the second assessment, at least when we hold his performance relative to the total number of opportunities which were allowed for movement each day.

Rate statements are used to correct for differences in assessment times. If a child reads 125 words correctly during a two minute assessment on one day, and 35 words during a thirty second timing on the next day, has he improved? As with percentages, each count is adjusted to reflect the expected behavior count for a standard base -- in this case, a standard assessment time. In most situations a time base of one minute is employed. To find the number of words read per minute of assessment, the behavior count is divided by the number of minutes over which the assessment was conducted. For the first day, 125 words ÷ 2 minutes equals 62.5 words per minute; and for the second day, 35 words ÷ 0.5 minutes (i.e., 30 seconds) equals 70 words per minute. It would appear that the child improved slightly.

The calculation of either percentages or rates allows comparions to be made between counts collected under different assessment conditions. It should be noted, however, that the adjustment is artificial, and that the differences in assessment conditions still exist. It could be, for example, that the child's average rate per minute is higher in a 30 second timing because he is less fatigued. Had the child been timed for a full two minutes on the second day, his actual performance might have turned out no better than his performance on the first day. Rate and percentage statements will not pinpoint why performances differ, they will only help to identify cases in which they do differ.

The Record Floor

To clarify the existence of differing assessment conditions and help in their meaningful analysis, many educators have adopted the procedure of calculating and reporting record floors (White and Liberty, 1976 ~~in press~~; White and Haring, 1974 ~~in press~~). The record floor, in any given assessment situation, is the mathmetically lowest non-zero performance value which can possibly be recorded. Since the lowest non-zero behavior count is one (assuming that only whole movement cycles are counted), the lowest non-zero percentage or rate which can be recorded will be that value based on a count of one. For the examples provided above, ten opportunities would yield a record floor of (1 x 100) ÷ 10 = 10%; twenty opportunities produces a record floor of (1 x 100) ÷ 20 = 5%; a two minute timing will produce a record floor of (1 ÷ 2 =) 0.5 movements per minute, and a thirty second timing has a record floor of (1 ÷ 0.5 =) 2 movements per minute.

The record floor makes two statements: it tells us the lowest limit of our measurement (i.e., we can only measure performances equal to or greater than the value of the record floor), and it defines the smallest amount of behavioral change which we will be able to assess accurately (i.e., unless the performance increases or decreases by a value equal to or greater than the value of the record floor, we cannot measure the change). By comparing apparent changes in performance against the value of our assessment record floors, therefore, we are able to decide whether all of that apparent change might be due only to the differences in mathematically possible values. Going back to the example concerning percentages which was presented earlier, we find that the difference between the two percentage statements was (50% - 45% =) 5%. But the record floor for the first day's assessment was (1 x 100) ÷ 10 = 10%, a value larger than the apparent change, so we must conclude that all of the observed change might be due only to differences in assessment procedures. The calculation and use of record floors helps immensely in distinguishing real performance differences from those differences which are only a function of the way in which we choose to collect, adjust, and present our behavior counts. The importance of record floors to the meaningful analysis of behavior change cannot be overestimated.

Choosing Between Percentages and Rates

Percentage statements have been used with far greater frequency than rate statements in traditional educational and psychological literature. That tendency appears to be reversing itself for many reasons, however.

Percentage statements are used primarily to explicate the proportion on one particular behavioral phenomena in relation to some larger set of possible phenomena. In education, the most common example would be a statement of percentage correct behavior. The relative accuracy of a student's performance is certainly important, but many educators are rapidly coming to the conclusion that <u>fluency</u> (how rapidly the child works) is equally important. For example, if a child reads with perfect accuracy but at a rate of only 25 words per minute, he is likely to be far less successful in school than a child who reads with only 95% accuracy at a rate of 125 words per minute. Also, many management or social problems are almost exclusively a problem of frequency. All children will get out of their seats from time to time, but it is the child who does so with a rate of 0.10 movements per minute (i.e., one movement every ten minutes) who will come to the attention of the teacher.

Secondly, correct and error percentages are ipsative (that is, the value of one determines the value of the other). If a child's percentage of correct movements increases, his percentage of error movements <u>must</u> decrease. In fact, however, the actual <u>number</u> of both types of behavior might increase or decrease on any given day (assuming that the total number of behaviors is not fixed). Since rate statements do not mix correct and error counts, they allow the analysis of either form of behavior independently of the other; something which can prove quite useful in determining where a child's real problems lie.

Third, percentage statements have a definite ceiling: a child cannot be more than 100% accurate or take advantage of more than 100% of all behavioral opportunities. Just because a child reaches the magic level of 100% does not mean that his performances can no longer change, however. The child can still improve in the fluency and ease with which he performs his movements --something which only raw behavior counts or rate statements can reflect.

Since accuracy and fluency are both important parameters of a child's performance, one might be tempted to collect and calculate both types of data. In fact, rate statements alone will usually suffice. Since rate statements still contain information about the correct and error movement cycle counts, one may combine them to find percentage statements whenever necessary. Percentage statements do not contain any information about the assessment time, however, so they cannot be used to calculate rates. To save time, therefore, it is suggested that one collect rated information, and convert to percentages only when statements of relative accuracy are desired.

Frequency of Assessment

Learning, or behavior change of any type, is rarely a "one-trial" phenomena. Generally, behavior changes take place continuously, over time. To be most reactive to the needs of a pupil or client, therefore, it is essential that our assessments of behavior are as nearly continuous as possible. In most educational situations that will mean the scheduling of short, daily probes for each of the child's academic skills. In the counseling situation, that might mean the structuring of special self-observations or assessments at selected points during each day. In either case, our ability to employ consistently appropriate assistance to a client or pupil is directly limited by the frequency with which we assess the behavioral phenomena of concern. Of course, frequent assessments will be of little or no value unless we know exactly how to interpret and use the results of those assessments. To begin, we must know the aim of our interventions.

Aims: Norms, Criteria, and Proficiencies

If we are to be consistently successful in altering the behavior of our pupils and clients, it follows that we must know the aim or goal of our work. In general, we will want our charges to become "proficient" in some task, to "master" certain skills, or to reach some predetermined "terminal level of performance." Just how levels or proficiency, mastery, or terminal performances are set, however, is a matter of some debate.

Until recently, norm-references were the most commonly employed aims. That is, we attempted to make each child as much like other children as possible. Norms are established by measuring the performances of a child's chronological peers and then taking the average of those measures as the performance-level to which we will strive to bring each child. At one time in the history of education, there were only enough resources to provide a limited number of children with complete educational services. Measuring each child in terms of his relative prowess made sense. Only the "fittest" were allowed to proceed up the educational ladder. Now legislative and judicial mandates have made it quite clear that *all* children will be provided whatever services are appropriate to meet their individual needs. Knowing whether one child is superior to another is no longer important in and of itself. It is more important to determine whether or not each child is ready for particular educational services, and if not, to identify those services and experiences which are appropriate.

In response to this shift in the purpose of assessment, a second type of referent has emerged: the *criterion*. Supposedly, a criterion-referenced assessment is one in which the child's performance is compared against that level of performance required to be "successful" in a task. Unlike norm-referenced assessment, all children could presumably fail to reach a criterion, or all children could pass. Although many criterion in use today are simply the result of "arm-chair revelation" (reflecting only what one or more persons "feel" is an acceptable performance) successful task completion *should* be determined as that which results in one or more of the following conditions:

Transfer

(1) a performance on one task in a sequential task hierarchy which insures a high probability of continued progress on subsequent tasks in the hierarchy;

Retention

(2) a performance which insures the maintenance of a skill over time, or successful transfer of the skill from one situation to another, or;

usefulness

(3) a performance which meets the requirements for acceptable skill demonstration in "real world" situations.

Criteria for Progress

Criteria designed to insure a pupil's continued progress through the curriculum are generally the most immediately important to the classroom teacher. What levels of accuracy and fluency should a child reach in two-term, single-digit addition problems, for example, before he is really prepared to tackle two-term, double-digit problems successfully? Moving on too soon could result in difficulties with later curriculum, and moving on too late will waste time, at the very least, and could potentially result in loss of pupil interest.

There are essentially two methods for the determination of intracurricular criteria. First, all children could be moved from one level of the curriculum to the next at some predetermined time. This is a common programming tactic in any event. After the children have been working at the next level of the curriculum for some time, their performances are examined to see which of them have maintained or improved their rate or accuracy in the new material, and which have done more poorly than in the preceeding step. By re-examining their performances at the time of the program change, the minimum acceptable levels for advancement can then be deduced. This type of discriminate analysis (i.e., discriminating between successful and unsuccessful pupils on the basis of the performance they achieved in earlier parts of the curriculum) can either be performed at a simple paper and pencil level (e.g., Liberty, White, and McGuigan, 1975), or (where a great deal of data are concerned) with the aid of sophisticated computer analysis techniques.

Alternatively, all children are advanced to the next step in the curriculum whenever they reach that level of performance which one <u>believes</u> to be an appropriate criterion. The success and failure of all students in the next step of the curriculum are then examined. If most students <u>are</u> successful in each succeeding step, then it is assumed that the present criterion for advancement is appropriate. If a number of children fail, on the other hand, or get progressively worse, then the criterion for earlier performances is increased (to better prepare future pupils for the material which lies ahead). This method for the determination of progress criteria should really only be employed after some reasonable estimate of an appropriate criterion has been established, perhaps through the method described above.

<u>Criteria for Maintenance and Generalization</u>

The procedures for determining those criteria which insure maintenance or generalization of a skill are essentially the same as those for determining criteria for progressing through a curriculum. Frequently, however, maintenance and generalization can only be achieved if "overlearning" occurs. For example, if one wishes a child to maintain a reading rate of at least 100 words per minute, experience indicates that initial instruction will probably have to bring the child to a level of fluency in excess of that eventual goal (say, 125 words per minute).

Criteria for Successful Application in the "Real World"

Eventually, we want our pupils or clients to apply skills in their everyday lives. It is often difficult, if not impossible, however, to conduct follow-up studies to see whether application has occurred. As an alternative to the methods described above, therefore, we are often forced to seek out persons who already possess a skill at an obviously proficient level or to deduce what the real world working requirements will be for a skill, and use that information for the formulation of our performance criteria. For example, the criterion for oral reading might be set as the rate with which newscasters read prepared scripts on the television. Newscasters must read quickly and accurately, but must not read so rapidly that intonation and inflection suffer. In short, newscasters must be highly proficient oral readers. By setting their performance level (which, by the way, is surprisingly consistent from one newscaster to another) as our ultimate criterion, we are not likely to go wrong. A slightly different approach might be taken with silent reading. If the child with whom we are are working is likely to go on to college, we might seek out estimates of the amount of reading required for college students and the time they typically spend studying, and with that information, deduce the required reading rate. If the child is only likely to reach the level of sheltered workshop employment, then a list of required "survival" words (men, women, poison, stop, go, etc.) might be compiled by actually watching such people and noting the situations which they encounter.

Regardless of the approach we take in setting our criteria, it is likely to be a long and demanding task. Furthermore, as the requirements of our society change, recalibration of criteria will be necessary at regular intervals. Nevertheless, criteria are well worth the effort. Criterion-referenced assessments offer a pupil or client a chance for meaningful advancement and eventual success. Norm-referenced assessments only guarantee that a certain proportion of the children will look like relative failures, because their performances fall below the norm.

Criteria in Clinical Psychology

The establishment of meaningful criteria in clincial psychology will be much more difficult than in education. Problems tend to be more individual in nature (so there are fewer clients with similar problems to use as a group for discriminate analyses), and the "curriculum" through which a client must pass is far less well defined than that in the typical classroom. Even so, the potential advantages of empirically derived criterion remain, so the attempt should be made.

Progress criteria in clinical psychology should be established to indicate those performances which the client must display before the nature of therapy or counseling progresses from one phase to the next. When is a client ready for a group session? When can sessions be reduced from daily or weekly meetings? When can parents begin reducing the amount of artificial consequation (e.g., tokens for free time, or candy) they use in the home? At first, the criterion selected will be more the product of guesswork than empirical evidence; but if the attempt is made to quantify the basis of program decisions, the cummulative results of those data over time should begin to point out some consistencies. Of course, if several clinicians share their data, the process can be accelerated considerably.

Maintenance and generalization criteria can be tested by having the client or a member of his family keep records of the frequency of key behaviors outside of the counseling or therapy setting. How well must the client be made to behave in the special setting before the effects of that work begin to appear outside of that setting? Even if not all clients can be relied upon to produce accurate records, the results of work with those who can should eventually begin to provide some guidelines for work with those who cannot.

Application criterion might be estimated through the observation of persons deemed to be socially or psychologically adapted. What, for example, is the rate of negative statements with people who are not considered to be overly negative? In a way, the establishment of criteria based on the performances of other people can be construed as "norms." It must be noted, however, that only people who are considered to be successful are included in the sample, so the average performance does not necessarily reflect the "normal" performance of people in general.

Progress Evaluation

Knowing where a pupil or client should end up, either before moving on to the next step in our interactions with them, or before terminating services altogether, is only part of the answer to the full utilization of daily progress assessments. In order to be truly reactive to the individual needs of each person, we must be able to identify on a daily basis whether the pupil or client is progressing at an acceptable rate toward the criterion. Most people assume that a simple graph or chart of the daily assessments will suffice to meet this need. If the rates of the pupil or client are going up when they should go up, then all is fine. We leave the program alone. If the rates are going in the wrong direction, the program is changed. It is *not* as simple as all that! Most people will not progress evenly from one day to the next. Little patterns of ups and downs are likely to emerge which can obscure any overall, general pattern of growth or progress. Even if progress is continuously in the appropriate direction, there still remains the question of whether the progress is adequate to reach the goal *in time* (i.e., within the time available). Recently, an excitingly simple and effective method for daily progress assessment has been developed for use in education.

Minimum 'Celeration.[1] Given that we have decided the minimum

---

[1]The term *'celeration* derives from the terms *acceleration* and *deceleration*, the two different types of changes which we might want to achieve in the rate of a behavior. *Minimum 'celeration*, then, would be the least acceptable rate of change for any given behavior.

fluency and accuracy which we desire of a behavior, the client or pupil's present level of performance, and the time which is available to reach our aim, it should be possible to draw a line on a chart which describes how rapidly the performances must increase or decrease each day, on the average, to reach that aim. Then, by simply noting whether the client or pupil's performance meets or exceeds that expectancy each day, we can tell at a glance if the program must be modified or changed to avoid a potential failure. Specifically, the procedures for using a minimum 'celeration line are as follows:

(1) Draw an aim-star on the performance chart which represents the performance level you wish to achieve and the day by which that aim should be met.[2]

(2) Assess the performance of the pupil or client for three successive days (or sessions, whichever is most appropriate).

(3) Determine the pupil's or client's start-mark by finding the intersection of the mid-day (or session) and mid-rate for those first three assessments. Entry performance might be estimated with only a single assessment, but generally it is

---

[2] It makes some difference which type of chart is employed. In general, a chart which provides space for each calendar day will be best, since it takes into account both the time when we can work with the pupil or client and the time during which we cannot (but during which progress might still occur); and a chart should be selected which also makes the progress of the subject appear as linear as possible (i.e., the growth pattern should look like a straight line). . Semilog charts or log/log charts are usually more satisfactory in this last respect, since most human performance changes are usually proportionally, rather than absolutely, equivalent to the size of previous changes occurring at higher or lower rates.

better to use the median of several assessments to account for initial adaptation factors and other sources of error variance. The intersection of the mid-day and mid-rate is used so that the estimate of the person's entry behavior can be located in terms of both time and initial level of performance.

(4) Draw a line between the person's start mark and the aim-star. That line will represent how rapidly the performance must change over time, on the average, in order to reach the aim.

(5) Continue to work with and assess the behavior at regular intervals. Record the result of each assessment on the appropriate day line of the chart.

(6) Whenever the results of three successive assessments fall below the line, the probability that the person will reach the aim in time must be assumed to be unacceptably low, so the program must be changed in some way (to more frequent sessions, a different type of instruction or counseling, etc.). Some people have employed the criterion of two successive failures to reach the minimum 'celeration line; and in cases where it is extremely important that the subject have every chance of success, one might even change the program whenever any one of the assessments failed to reach the criterion. In general, experience has shown that a criterion of two or three days below the minimum 'celeration line will catch most programming errors and avoid most unnecessary program changes.

(7) If and when a change becomes necessary in a program, there are several options for determining a new minimum 'celeration line (Liberty, 1972; White and Liberty, 1976; White and Haring, ~~in~~ 1976). ~~press)~~ It is recommended, however, that the following procedures be employed:

(a) Establish a new start mark by drawing a line from the mid-rate of the last three assessments to the day on which the new program will begin.

(b) Re-evaluate the aim-star. If possible, leave it as it is. If there is some doubt that the subject will be able to overcome the original program problem prior to the date originally set for the aim, however, the aim-star can be adjusted back in time to allow more time.[3]

(c) Draw a new minimum 'celeration line from the new start-mark to the (original or adjusted) aim-star.

Continue the new program, the regular assessments and charting, and evaluate the subject's progress in accord with the rules specified above. The procedures for using minimum 'celeration are illustrated in Figure 1.

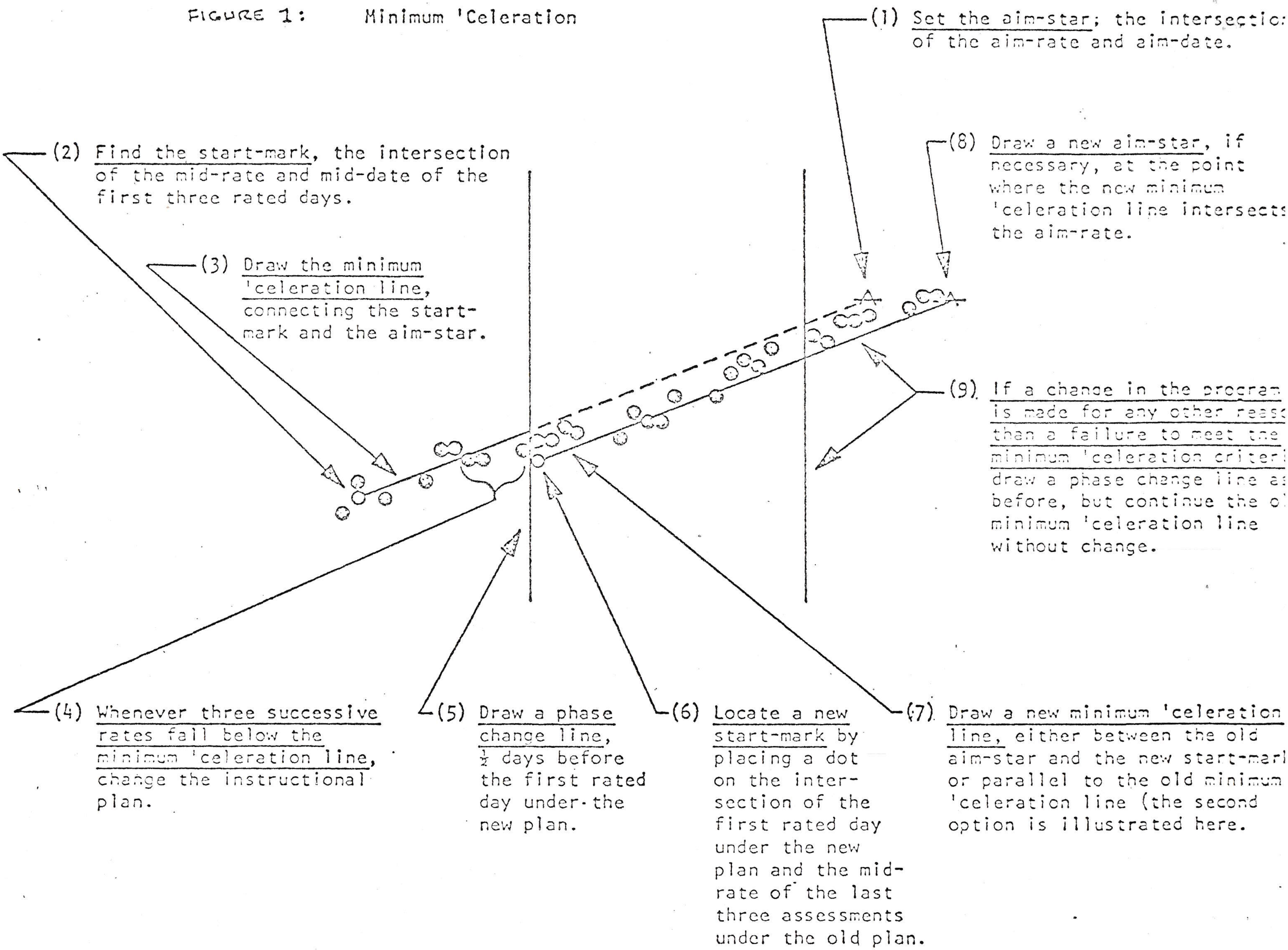
---

[3] In a surprising number of cases, the information gained from the first program failure is sufficient to prompt the development of a new program which is effective in helping a child to reach his aim within the time originally allowed (i.e., necessitating a rate of progress higher than originally expected). Since most immediate objectives are scheduled for completion within a period of one or two months, however, there is usually sufficient time left in the school year to readjust the aim-date if that is felt to be more practical than the demand for even greater rates of progress.

Insert Figure 1 About Here

The general procedures for employing minimum 'celeration techniques have been well documented, including certain cautions about a "blind" adherence to those rules (Liberty, 1972; White and Liberty, 1974; White and Haring, 1974a ~~in press~~). More importantly, evidence exists that the implementation of those rules for program evaluation can result in dramatically improved pupil progress and aim attainment. In a study involving 15 special education teachers and 74 learning disabled children, Bohannon (1975) found that when teachers employed simple, daily assessments and minimum 'celeration program change rules, they spent less time with pupils and yet were able to promote much higher rates of progress than when they tried to work without the daily data or rules. Starting with reading deficits of between one and three years, all of the children who were working under minimum 'celeration rules achieved a level of reading performance equal to or greater than the twenty-fifth percentile of their normal peers within twenty-eight days, while only two members of the contrast group achived similar progress. Moreover, teachers in the regular classes (who were unaware of which children were receiving which treatments) were able to identify members of the experimental group as having achieved better gains in <u>all</u> subject areas (presumably, as a result of their increased fluency in reading), and much better "affective" development in general.

FIGURE 1: Minimum 'Celeration



Although no specific research has been conducted on the application of daily data decision rules in clinical psychology, the implications are no less exciting. Since the rules are simple to follow and direct, it would be possible, for example, to have a client or parent chart the progress of a behavior and use that information to identify precisely when additional counseling or therapy is needed. In this way, clients could assume more of the responsibility for their own treatment, a desirable state of affairs in any event, and reduce the work load of the psychologist at the same time. Successes with procedures involving client collected data have already been reported (e.g., Patterson and Gullion, 1968; Deibert and Harmon, 1970; Knox, 1971; Stuart, 1971). It can only be assumed that rules which assist clients and managers in the more efficient and effective use of those data will be of tremendous benefit.

The Description of Progress

The minimum 'celeration line does not describe how the pupil or client is *actually* progressing; it only defines how he *must* progress in order to reach a level of proficient or acceptable performance within the time available. That will be all the information one requires when making daily programming decisions. It may be useful, however, to describe the actual progress of the subject for purposes of reporting or prediction. Three methods have been developed to do just that. The first two can be accomplished without special tools or extensive training. The third method will produce more reliable and predictively valid results, but usually requires the use of a computer. All three methods are based on the properties of the median, since that statistic tends to reduce the influence of the dramatic and non-random shocks to which all human learning curves are subject (e.g., days on which the subject is sick or is under the influence of some drug or emotional experience). Studies have indicated that mean-based descriptions of progress (e.g., regression) overemphasize the performance of the subject on unusual days, and the predictive validity of those descriptions suffers accordingly (White, 1972a, 1972b; Koenig, 1972).

The Quarter-Intersect Line of Progress

The quarter-intersect line of progress is a simple description of how rapidly the median performance of the pupil or client is changing over time (Koenig, 1972). After dividing the data to be summarized into two equal halves, one simply finds the intersection of the mid-rate and mid-date in the first half (counting only those days on which assessments were conducted), and the intersection of the mid-rate and mid-date in the second half. A line is then drawn through those intersections to describe how rapidly the rates of the subject are increasing or decreasing over time.[4]

---

[4]References here are made to "rates" simply for the sake of convenience. While it is true that most of the research concerning the predictive validities of lines of progess has been conducted on rated information, there is no reason to believe that these procedures cannot be applied with success to other forms of data as well (e.g., duration or latency times). If an ipsative datum is used, however (e.g., percentages), one must take care to account for any mathematical limits imposed by the measurement scale (e.g., a ceiling of 100%).

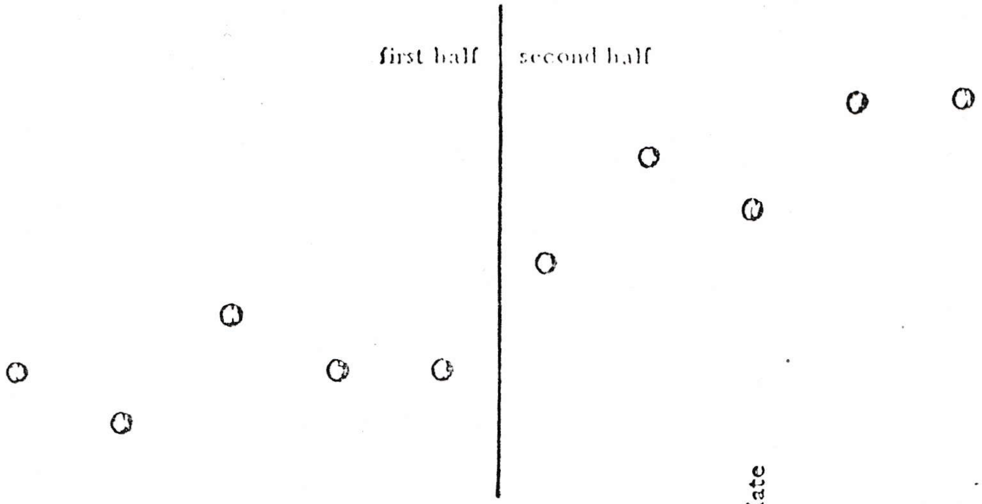
The Split-Middle Line of Progress

The quarter-intersect line of progress will represent a reasonable description of the subject's overall rate of progress. The line may, however, be a little high or low on the chart for an accurate description of the subject's average level of performance on any given day. The split-middle line of progress corrects for this problem by moving the quarter-intersect line up or down (keeping it parallel to its original slope) until 50 percent of the data fall on or above the line and 50 percent of the data fall on or below the line (i.e., until the properties of a median are achieved). By correcting for an imbalance of data above and below the line of progress, the split-middle will be somewhat more accurate in predictions and a more reasonable description of the child's performance. It should be noted, however, that this final adjustment might take more time than all preceeding steps combined, and so is usually reserved only for those cases in which the most precise description is required.[5] The procedures for finding the quarter-intersect and split-middle lines of progress are illustrated in Figure 2.

Insert Figure 2 about here

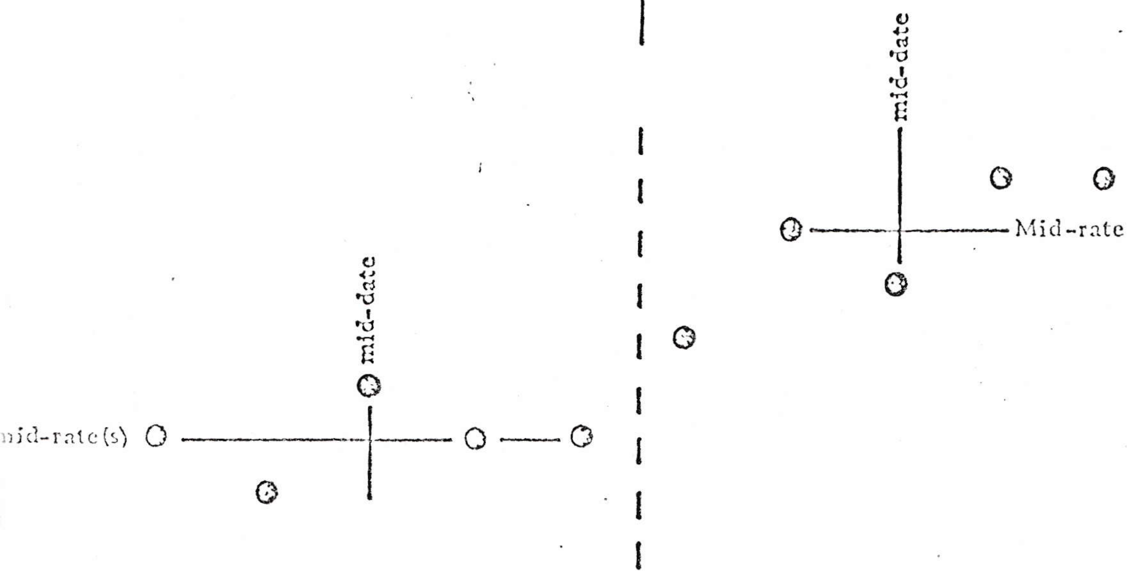
----

[5] I do not mean to imply that the time required to find either of these lines is very great. With practice, one should be able to find the quarter-intersect line for 20 data points in about ten seconds and the split-middle line in 20 or 30 seconds.



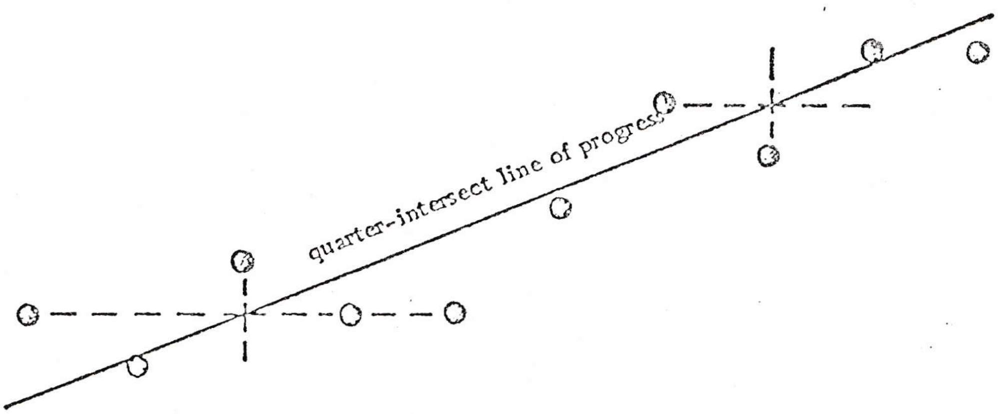
Step One

Divide the data to be summarized into two equal parts. If there are an even number of data points, the dividing line will fall halfway between two of the rates, as is the case here. If there are an odd number of data points, the dividing line will fall on one of the data points.



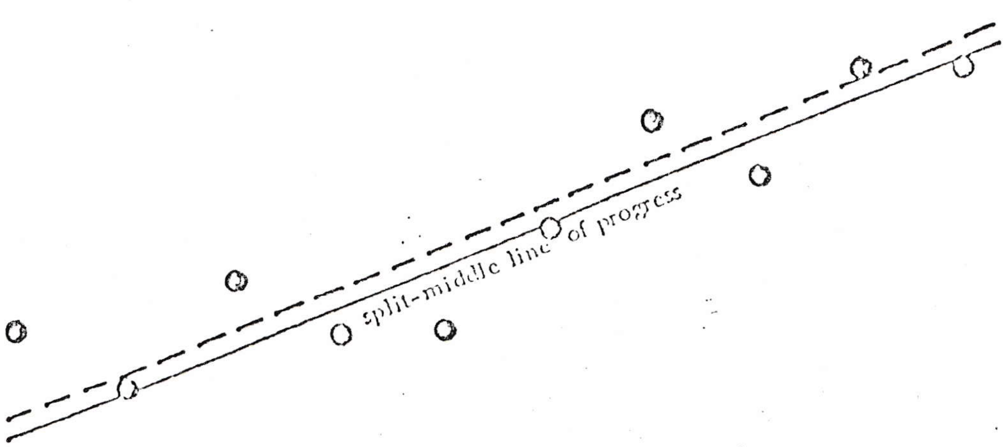
Step Two

Find the intersections of the mid-rate and mid-date for each half. In this case, since there are five data points in each half, the mid rates will be the third data point counting up or down the chart, and the mid-dates will be the third data point counting left or right on the chart. Several data points can all fall on the mid-rate, as illustrated in the first half of this example.



Step Three

Draw a line through the data which passes through both of the intersections found in step two, above. If you stop at this step in the process, you will have found the quarter-intersect line of progress. To find the split-middle line, complete step four, below.



Step Four

Count the number of data points which fall above and below the line drawn in step three, above. There should be the same number of data points falling on and above the line as there are falling on and below the line. If not, move the line up or down (keeping it parallel to the original line) until a balance is achieved.

The Median Slope Line of Progress

The quarter-intersect and split-middle lines of progress describe how a subject's median performance is changing over time. They do not, however, yield a line which actually displays the properties of the median with respect to all the data (i.e., those lines do not necessarily minimize the sum of the unsigned deviations of all data about it). Finding a line which does possess the properties of a dynamic median will result in a more valid and predictively useful description (White, 1971a, 1971b, 197~~5~~6). Unfortunately, to find such a line requires a great deal more time and effort on the part of the analyst. Two methods are available:[6]

(1) All lines between all possible pairs of data [that is, $\frac{1}{2}n(n - 1)$ lines] are generated, and the sums of the unsigned deviations about each line determined. The line(s) which minimize the sum is then selected as the median slope line of progress.

(2) Alternatively, a set of decision rules can be used to find the median slope line of progress without generating all possible lines. Based on certain geometric relationships which must exist between the possible lines of progress in a data set, these rules can reduce the time required to find a solution. The rules are quite difficult to program, however, so the development of appropriate software is usually undertaken only

---

[6]There are multivariate equivalents to the procedures which are described here that have been used in the field of econometrics for some years called the L_1 criterion (Sposito and Smith, 1974). Discussion of these multivariate techniques has been eliminated from this paper, however, since they have little application in clinical practice.

when large data sets are to be analyzed in great number (White, 1972b).

On occasion, more than one line of progress will satisfy the mathematical requirements for the median slope. In such situations, it must be assumed that the subject's performance characteristics are not clearly defined by the available data and that additional data should be collected. If that is not possible, then it is recommended that the most conservative of the alternative lines be selected for use in further analyses.

The Prediction of Progress

Having described a subject's rate of progress under a set of program conditions, one may wish to predict the level of future performances (assuming, of course, that all program conditions remain unchanged). Such predictions are useful in making prognostic statements, and for a detailed analysis of program change effects (to be discussed later).

To accomplish a prediction, one simply lays a straight-edge along the line of progress and draws an extension of it out to the date in question. Mathematical predictions may also be made, but rarely increase the accuracy of prediction to a degree which warrants the extra work in a clinical setting. Of course, predictions should never be allowed to exceed the limits of measurement (e.g., go above 100% when using percentage data, or below the record floor with either percentage or rate data).

The predictive validity of the quarter-intersect, split-middle, and median slope lines of progress has been investigated on several large data sets (White, 1972a, 1972b; Koenig, 1972). In all cases, semilog transformations of the data were employed, and only rate data have been analyzed in sufficient number to allow meaningful conclusions to be drawn. Empirically derived estimates of successful prediction (over varying periods of time, given varying amounts of data to use in formulating the prediction) are provided in Table 1 for the split-middle and median-slope methods of deriving a line of progress.[7] As can be seen in that table, levels of predictive validity acceptable for research purposes are not really reached until nine or eleven data points are available to draw the line of progress. Reasonable confidence might be placed in predictions based on as few as seven data points, however, for most clinical purposes. Also, note that although the median-slope procedure is consistently superior to the split-middle procedure, practical differences become quite small when nine or eleven data points are used.

---

[7]All predictions studied were conducted within phases of actual classroom programs, i.e., over periods during which program variables were held constant. In each case, the first few data points in each phase were used to formulate a prediction of the pupil's performance later in the same phase. Criteria for successful prediction were based on "envelopes" drawn about the line of progress in the predictor data set. One envelope was drawn to contain 50% of the data, another to contain 75%, and a third to contain 100% of the data used to formulate the line of progress. These envelopes were then projected into the predicted data set (along with, and parallel to, the line of progress). If the distribution of predicted performances within those projected envelopes equalled or exceeded 50%, 75%, and 100% respectively (i.e., attained the same distribution about the predicted line of progress as the original predictor data set), then prediction was considered sufficiently precise for experimental purposes. Criteria for success were also tested in which 25% of all predicted data were allowed to exceed expected deviations from the projected line of progress. Predictions meeting these criteria were judged acceptable for most clinical applications where somewhat less accurate predictions might still be considered functional.

Insert Table 1 About Here

It would seem reasonable that certain behavior types or performances with certain characteristics would be more predictable than others. Preliminary studies concerning behavior type (with behaviors ranging from simple pointing responses to complex reading behavior), performance variability, and initial rate of progress have failed to identify any differences in predictive validity, however. Thus far, it would seem that all general performance types are equally predictable, albeit, within a considerable range of predictability which applies to all performances (White, 1971c, 1972a). Despite these encouraging results, one must bear in mind that the prediction of future events cannot be mathematically or theoretically justified in the same sense that the appropriate application of an F-test can be justified. There is always some unknown probability that new variables or continued exposure to old variables will affect the performance of the subject in some unexpected manner and invalidate any prediction. It is suggested, therefore, that an attempt be made to empirically validate predictions whenever possible, and that a table of successful prediction probabilities be constructed which matches the behaviors and situations which you will encounter. Also, there will be occasions when a single line of progress will not serve as an adequate descriptor of a subject's progress during any given phase of a program. Perhaps the line intersects a record floor or ceiling, or perhaps the subject's reaction to treatment variables changes with increases or decreases in performance. By selecting the appropriate

TABLE 1: Percent of successful predictions using the split-middle and median-slope lines of progress

Number of days used to draw the line of progress		Number of days over which prediction is extended: 2	4	6	8	10	12	14	16	18	20
5	Split-Middle	62-64%	54-60%	48-56%	41-54%	41-48%	34-44%	31-41%	26-37%	24-32%	24-30%
	Median-Slope	68-71%	66-70%	63-69%	61-67%	59-66%	57-64%	54-63%	53-62%	52-61%	50-59%
7	Split-Middle	74-87%	71-84%	70-82%	64-78%	60-74%	60-70%	56-66%	51-63%	50-67%	50-67%
	Median-Slope	78-90%	76-88%	75-86%	73-85%	72-83%	70-81%	69-80%	67-78%	66-77%	65-75%
9	Split-Middle	93-96%	91-95%	83-93%	80-91%	77-89%	71-86%	70-83%	70-81%	70-79%	70-78%
	Median-Slope	88-96%	88-96%	87-96%	87-96%	87-96%	86-96%	86-97%	85-97%	85-97%	85-97%
11	Split-Middle	96-99%	94-99%	91-98%	87-92%	84-92%	82-92%	81-92%	81-92%	81-92%	81-92%
	Median-Slope	97-99%	96-99%	96-98%	95-98%	94-98%	94-97%	93-97%	93-96%	92-96%	92-96%

NOTE: The low percentage in each case is the actual percentage of predictions in which the distribution of predicted data about the line of progress was initially the same as the distribution of data about the line of progress in the predictor data set. The high percentage in each case is the percentage of predictions in which 25% of the predicted data exceeded the limits of the distribution defined by the predictor data set. All percentages are based on a sample size of 1,150 predictions of actual classroom data. All percentages are rounded to the nearest whole value.

chart, one which produces a visually linear pattern of growth,[8] many problems in description and prediction can be avoided; but under certain conditions, it will be necessary to divide a single phase of a program into two or more parts with a separate line of progress for each (White and Haring, in press). Of course, in such a case, only the last line in a phase would be used for actual predictions.

The Analysis of Program Change Effects

Although the line of progress and predictions based upon that line will be useful in estimating prognoses for program success or failure, the primary use of lines and predictions is in the analysis of program change effects. Data collected on a single individual over time represent a sequentially dependent time series. That is, each data point can be expected to influence, to some extent, the value of succeeding data points. As such, these data cannot be treated in the usual manner (e.g., with simple F or t tests) when it comes to estimating the magnitude and significance of changes in performance attributable to program alteration (e.g., see Glass, Wilson & Gottman, 1975). Although several valid statistical treatments have been devised to correct for sequential dependency (e.g., Bartlett, 1935; Anderson, 1942; Box, Jenkins, and Bacon, 2967; Coutie, 1962; Gottman, McFall, and Barnett, 1969; and Glass, Willson, and Gottman, 1975), none of these traditional procedures yields a description of performance changes which can assist the practitioner

[8] Usually a semilog or log/log chart is best (White and Liberty, in press).

in interpreting the meaningfulness of results. By describing the growth which occurred within each phase of a program with a line of progress, however, and then extending each line into the next treatment phase, an excellent and easily interpreted picture of performance trends and changes can be achieved. Figure 3 illustrates the use of lines of progress for a between phase analysis.

Insert Figure 3 About Here

A heavy vertical line is drawn on the chart to indicate the day on which the program was changed. A line of progress (in this case, the split-middle line) is drawn through the data in each condition to describe daily growth and progress. Dotted lines are drawn as extensions of each line of progress to indicate where we might predict the subject would have gone if conditions were left unchanged. The analysis of changes then proceeds as follows:

(1) The immediate impact of the new program is determined by examining where the first line of progress ends and the next line begins. This "step" change has often been equated with a Hawthorne or novelty effect, but may represent a true and permanent change in the subject's behavior.

(2) The change in rate of learning or behavior change is determined by comparing the slope of the first line of progress with the slope of

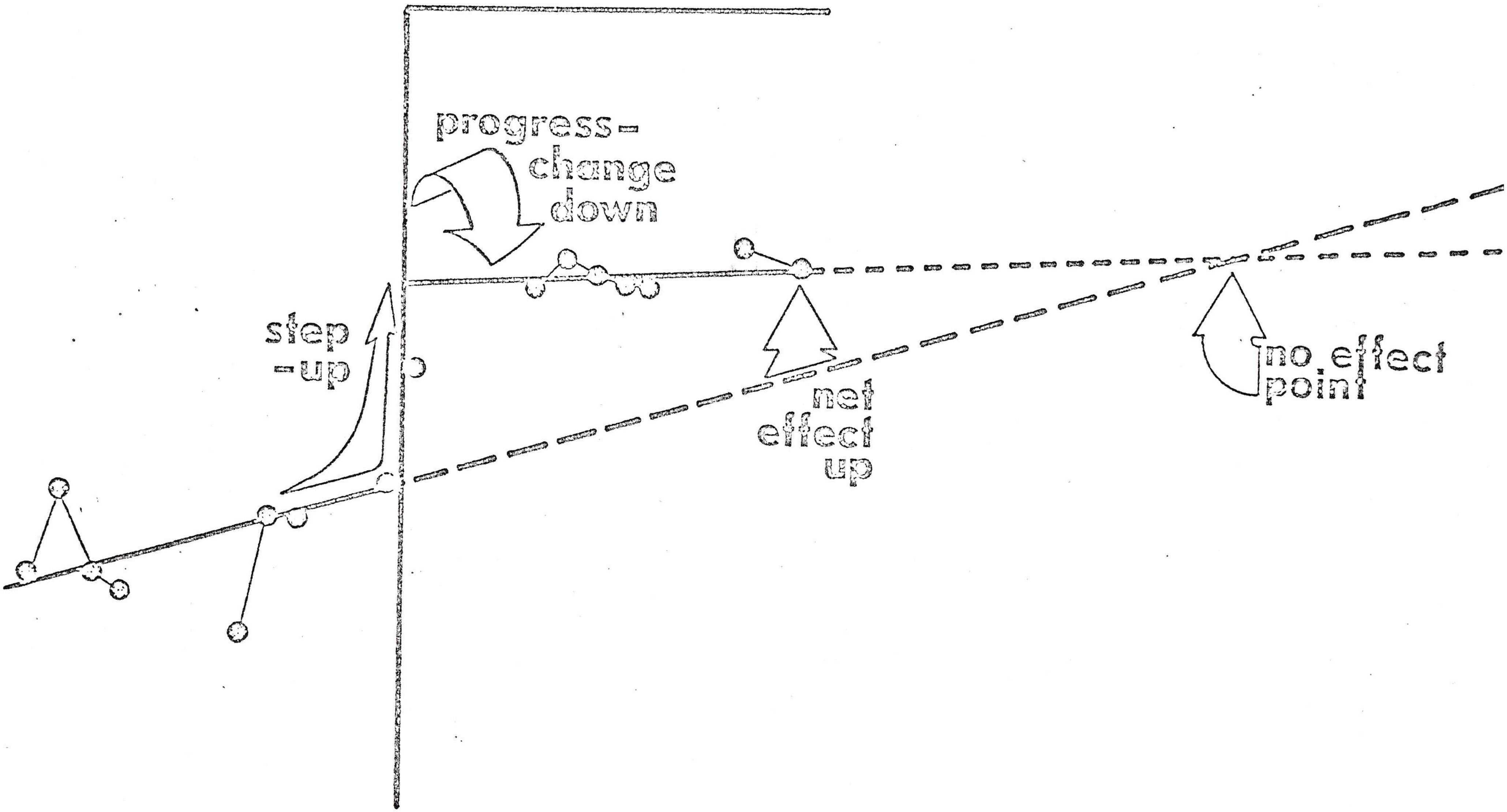


FIGURE 3: BETWEEN PHASE ANALYSIS

the second line of progress. In the example provided, the subject has changed from an upward trend to virtually no trend at all. A progress change will continue to affect performances for as long as the condition which produced it is left in effect.

(3) The net effect of the program change is determined by comparing the value of the new line of progress and the value of the predicted line of progress (i.e., the prediction based on the line of progress in the first phase) at the time the program is terminated or changed once more. In the example provided, the net effect is "up," since the child's actual level of performance exceeded that which we would have predicted on the basis of his old line of progress. Note, however, that if the second treatment had not been terminated when it was, we might expect that the child would eventually be performing below the predicted line of progress. The progress change down would eventually cancel out the effects of the step change up (note where the two prediction lines cross). If step and progress changes are in opposite directions, therefore, a program change cannot be considered either all good or all bad. It will depend on the location of the no-effect point (the place where the step and progress changes cancel out), and whether we terminate or change the program again before or after that point.

An analysis of lines of progress and their projections enables a detailed analysis to be made of changes within and between program phases that would not otherwise be possible. More detailed discussions of such analyses are beyond the scope of this manuscript, but may be found elsewhere (e.g., White, 1972a, 1972b, 1974; White and Haring, in press). Procedures for the application of inferential statistics to the question of the significance of step, progress change, and overall net effect of changes have also been developed (White, 1972b), but if significance of change statements are desired, it is suggested that the more powerful traditional time-series statistics be employed (e.g., Glass, Willson, and Gottman, 1975). Usually, significance statements will be of secondary concern to the clinical practitioner or teacher. The simple description of changes in the precise manner described above will be sufficient for the interpretation of program value, identifying which programs might be expected to produce which types of changes, for developing an expectancy table of initial effects (so an initial effect detrimental to the overall aim will not necessarily result in program alteration if an appropriate progress change can be expected to follow), and for determining how long different programs must be left in effect to achieve or avoid an overall combined effects of step and progress changes.

In Conclusion

The technology of data collection and analysis in education is advancing rapidly. Educational practitioners are, for the first time, beginning to realize the true potential of systematic, data-based instruction. I have only been able to scratch the surface of the available technology in this manuscript. But if I have been successful in pointing out some of the procedures and practices which might be of value in clinical psychology, then I have achieved my purpose. Availing oneself of the referenced literature will fill in the details needed to actually begin the implementation of an advanced data-based program.

References

Anderson, R. L. Distribution of the serial correlation coefficient. Annals of Mathematical Statistics, 1942, 13, 1-13.

Bartlett, M. S. Some aspects of time correlation problems in regard to tests of significance. Journal of the Royal Statistical Society, Series B, 1935, 24, 297-343.

Baum, W. M. Time allocation in human vigilance. Journal of the Experimental Analysis of Behavior, 1975, 23(1), 45-54.

Bohannon, R. Direct and daily measurement procedures in the identification and treatment of reading behaviors of children in special education. Doctoral dissertation, College of Education, University of Washington, 1975.

Box, G. E., Jenkins, G. H., & Bacon, D. W. Models for forecasting seasonal and non-seasonal time-series. In B. Harris (Ed.), Spectral analysis of time-series. New York: Wiley, 1967.

Broden, M., Hall, R., & Mitts, B. The effects of self-recording on the classroom behavior of two eighth-grade students. Journal of Applied Behavior Analysis, 1971, 4, 191-199.

Christman, R. L. A digital sequential generator for implementing reinforcement schedules. Journal of the Experimental Analysis of Behavior, 1974, 22(3), 577-580.

Cohen, I. L. The reinforcement value of schedule-induced drinking. Journal of the Experimental Analysis of Behavior, 1975, 23(1), 37-44.

Coutie, G. A. Commentary on paper by Box and Jenkins. Journal of the Royal Statistical Society, Series B, 1962, 24, 345-346.

Diebert, A. N., & Harmon, A. J. New tools for changing behavior (2nd. ed.). Champaign, Ill.: Research Press Company, 1971.

Flory, R. K., & Lickfett, G. G. Effects of lick-contingent timeout on schedule-induced polydipsia. Journal of the Experimental Analysis of Behavior, 1974, 21(1), 45-55.

Glass, G. V., Willson, V. L., & Gottman, J. M. Design and analysis of time-series experiments. Boulder, Colorado: Colorado Associated Press, 1975.

Gonzalez, F. A., & Waller, M. B. Handwriting as an operant. Journal of the Experimental Analysis of Behavior, 1974, 21(1), 165-176.

Gottman, J. M., McFall, R. M., & Barnett, J. T. Design and analysis of research using time series, Psychological Bulletin, 1969, 72(4), 299-306.

Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., Davis, F., & Porcia, E. The teacher as an observer and experimenter in the modification of talking out behavior. Journal of Applied Behavior Analysis, 1971, 141-150.

Knox, D. Marriage happiness: A behavioral approach to counseling. Champaign, Ill.: Research Press Company, 1971.

Koenig, C. H. Charting the future course of behavior. Unpublished doctoral dissertation, Department of Education, University of Kansas, 1972.

Liberty, K. A. Data decision rules. University of Oregon, Regional Resource Center for Handicapped Children, Working Paper No. 20, 1972.

Liberty, K. A., White, O. R., & McGuigan, C. A discriminate analysis procedure for determining proficient performance levels. A working paper available from the Experimental Education Unit, Child Development and Mental Retardation Center, University of Washington, Seattle, Washington, 1975.

Lindsley, O. R. Direct measurement and prosthesis of retarded behavior. Journal of Education, 1966, 9, 27-36.

Patterson, G. R., & Gullion, M. E. Living with children: New methods for parents and teachers (revised edition). Champaign, Ill.: Research Press Company, 1971.

Rashotte, M. E., Katz, H. N., Griffin, R. W., & Wright, A. C. Vocalizations of White Carneaux pigeons during experiments on schedule-induced aggression. Journal of the Experimental Analysis of Behavior, 1975, 23(2), 285-294.

Risley, R. R., & Hart, B. Developing correspondence between the verbal and non-verbal behavior of preschool children. Journal of Applied Behavior Analysis, 1968, 1, 267-281.

Selekman, W., & Meehan, E. An objective technique for recording shock-induced aggression in unrestrained pairs of rats. Journal of the Experimental Analysis of Behavior, 1974, 21(1), 177-182.

Sposito, V. A., & Smith, W. A sufficient condition for L_1 estimation. Unpublished manuscript, Iowa State University, 1974.

Stiers, M., & Silverberg, A. Lever-contact responses in rats: Automaintenance with and without a negative response-reinforcer dependency. Journal of the Experimental Analysis of Behavior, 1974, 22(3), 497-506.

Stuart, R. Trick or treatment. Champaign, Ill.: Research Press, 1971.

Walker, H., & Buckley, N. K. The use of positive reinforcement in conditioning attending behavior. Journal of Applied Behavior Analysis, 1968, 1(3), 245-250.

White, O. R. Glossary of behavioral terminology, Champaign, Ill.: Research Press, 1971a.

White, O. R. A brief outline of the "corrected median slope" method of trend estimation. University of Oregon, Regional Center for Handicapped Children, Working Paper No. 2, 1971b.

White, O. R. A pragmatic approach to the description of progress in the single case. Unpublished doctoral dissertation, Department of Special Education, University of Oregon, 1971c.

White, O. R. The prediction of human performances in the single case: An examination of four techniques. University of Oregon, Regional Resource Center for Handicapped Children, Department of Special Education, Working Paper No. 15, 1972a.

White, O. R. A manual for the calculation and use of the median slope -- A method for progress estimation and prediction in the single case. University of Oregon, Regional Resource Center for Handicapped Children, Working Paper No. 16, 1972b.

White, O. R. The "split-middle": A "quickie" method of trend estimation (third revision). Working paper available through the EEU/CDMRC, WJ-10, University of Washington, 1974.

White, O. R., & Haring, N. G. Exceptional teaching: A multimedia training package. ~~for exceptional children (Multimedia program)~~. Columbus, Ohio: Charles E. Merrill, ~~in press~~ 1976.

White, O. R., & Liberty, K. A. Behavioral assessment and precise educational measurement, in N. G. Haring & R. Schiefelbusch (Eds.), Teaching special children. New York: McGraw-Hill, ~~in press.~~ 1976.

White, ~~O. R., & Liberty, K. A. Practical classroom measurement:~~ Precision ~~teaching. In N. G. Haring & R. Schiefelbusch (Eds.),~~ Teaching ~~special children. New York: McGraw-Hill, in press.~~

incorporated in

Wright, A. A., & Nevin, J. A. Signal detection methods for measurement of utility in animals. Journal of the Experimental Analysis of Behavior, 1974, 21(2), 373-380.

Rosenthal, E. J. Evaluation history. In S. B. Anderson, S. Ball, R. T. Murphy, & Associates (Eds.), Encyclopedia of educational evaluation. San Francisco: Jossey-Bass Publishers, 1975.

Sax, G. The use of standardized tests in evaluation. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Corp., 1974.

Scriven, M. Maximizing the power of causal investigations: The Modus Operandi method. In G. V. Glass (Ed.), Evaluation Studies Review Annual (Vol. 1). Beverly Hills, California: Sage Publications, 1976.

Scriven, M. Evaluation perspectives and procedures. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Corp., 1974.

Stuening, E. L., & Guttentag, M. (Eds.). Handbook of evaluation research. Beverly Hills, California: Sage Publications, 1975.

Twain, D. Developing and implementing a research strategy. In E. L. Stuening & M. Guttentag (Eds.), Handbook of evaluation research. Beverly Hills, California: Sage Publications, 1975.

Walberg, H. J. (Ed.). Evaluating educational performance: A sourcebook of methods, instruments and examples. Berkeley, California: McCutchan Publishing Corp., 1974.

Walberg, H. J. Evaluating educational performance. In H. J. Walberg (Ed.), Evaluating educational performance: A sourcebook of methods, instruments and examples. Berkeley, California: McCutchan Publishing Corp., 1974.

Weiss, C. H. Evaluation research in the political context. In E. L. Stuening & M. Guttentag (Eds.), Handbook of evaluation research. Beverly Hills, California: Sage Publications, 1975.

White, O. R. Behaviorism in special education: An arena for debate. In R. D. Kneedler & S. G. Tarver (Eds.), Changing perspectives in special education, Columbus, Ohio: Charles E. Merrill Co., in press.

White, O. R., & Haring, N. G. Exceptional teaching: A multimedia training package. Columbus, Ohio: Charles E. Merrill Co., 1975.

White, O. R. The Prevalence and Costs for Educating Handicapped Children in Region X. Paper prepared for the Northwest Regional Resource Center for Handicapped Children, University of Oregon, Eugene, Oregon, 1973.

Wolf, R. M. Data analysis and reporting considerations in evaluation. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing Corp., 1974.