Problems in Rating Disturbed Behavior*  

HAROLD C. BATCHELDER AND BENJAMIN C. HAMMETT  

Virginia Treatment Center for Children, Richmond, 23219  

Introduction  
A maiden lady or shy bachelor may risk savings on an ocean cruise—envisioning romance under the stars, happy laughter, bubbling champagne and even finding a handsome and wealthy life partner. They may end up at home port broke, and with memories of mal de mer, finding only that they have spent much money to go nowhere and get nothing.  

The researcher runs a similar risk; he may have his vision and undertake an untold amount of work only to find he has not gone far. Even more frustrating, he may find upon the completion of his project that someone else has arrived before him.  

Yet, as with the romantic adventurers there is the need for the risk; there is always that chance that the desired goal may be realized and it may be better than even the wildest flights of scholarly fancy would permit. Of course, the adventurers have a distinct advantage over the researchers: if none of their friends or relatives accompany them, when they return home they can always lie. If they lie often enough and well enough about their experiences, in due time, they might even believe themselves. Fantasies are better than nothing on a cold winter’s evening.  

This past year Dr. Hammett and I have embarked on a research voyage and, if I may be permitted to use the analogy of a cruise once again, we have been on the slowest of freighters. The Ancient Mariner is our spiritual brother. In speaking of the problems of research, one of the problems is time! There are no short-cuts. There has been, for some time, an increasing awareness of the need to look at the causes as well as methods of treating emotional disturbance in children. Because of the addition of new staff members we have been able to undertake this extensive project without having to diminish our services to the children and their families under our care.  

The 1956 General Assembly of Virginia which brought the Center into existence recognized the need for research and included this as one of our major functions. Since our first patients arrived, eight years ago, the climate for research has been present and fostered. Over the years of operation, staff members have presented papers at local, regional and national meetings, as well as making substantial contributions to the professional literature.  

The project we are discussing here—producing a research form for rating disturbed behavior—proves that there is such a thing as evolution and like man, with all his frailties, has great promise for both the present and the future. Our first attempt at constructing this form was as sophisticated as pithecanthropus erectus and as manageable as the Frankenstein monster. In the planning stage it was a wonder to behold. When in our naive enthusiasm we tried it out in the field, we were wonders to behold as we attempted with assured professionalism to complete a six page, three part, 199 item questionnaire. It was unmanageable, frustrating and created such violent emotional reactions in even the calmest of our staff that it was consigned rapidly to an unmarked grave. The greatest success we had in constructing this monstrosity was that we managed to omit the very items which would measure emotionally disturbed behavior in children. Sadder and wiser we returned to planning, while the Center world waited to see what we would inflict upon them next.  

The initial step was to define the areas that we wanted to research. Basic to our problem was one large Commonwealth and one small Treatment Center. It is physically impossible to accept every child for whom application is made. Which children should we accept and which ones do best in our program? Throughout the years we have learned to identify the types of children we can help, fairly well; yet our procedure needs to be refined. We want to know: what are the common characteristics of the children? Why do some do well while others do not? Why do some of the children we think are going to do well

fail, while others about whom we have had grave reservations do very well? Is improvement predictable? These are only a few of the questions we are asking. Even if we were not asking questions there still is a point at which facts, figures and details must be recorded and analyzed in a systematic manner in order to assess programs, plan for the future, and identify both needs and accomplishments. To this end we have developed a behavioral questionnaire which in the recently completed pilot studies evidences great promise, not only for us but hopefully for similar institutions.

An intensive review was made of a number of questionnaires to determine what we felt should be the essential characteristics incorporated into a form specifically designed to measure the behavior of emotionally disturbed children. The criticism of R. N. Dreger (1964) "... That most behavior scales are second-order or third-order judgement measures, rather than true behavior measures" came to have special significance for us. We wanted to know how the parents of the children saw them actually behaving. We wanted a first hand report, not a second or third-order judgement. Further we realized that any questionnaire to be rated by both the lettered and the unlettered parent must contain simple, unambiguous and specific language. Instead of our asking whether the child "verbalized fantasies," we ask if he has "imaginary playmates." That we have succeeded in making our form understandable is demonstrated by the fact that the majority of parents are able to complete it in its entirety. Usually, the questions omitted are those having to do with specific matters, such as dating practices. Parents of the younger children usually will write in, "... He is too young to date" although they could check "Always" when the question asks, "Does he avoid dating."

Brevity and a consistent format were two other essential characteristics. Some of the questionnaires reviewed had as many as 295 items with many subsections. Others shifted in format from a five point rating scale, to multiple choice, to sentence completion and to True and False. A group of us tried out some of these forms on our willing families and found them confusing and time consuming. The untrained rater would not be able to complete them easily with any degree of accuracy. I do believe the same people who revised the recent Federal Income Tax Form must have learned their trade from some of these questionnaires. Even the simplest and most innocuous of forms can be a threat to some. Remember the furor caused by the 1970 Census Form? Parents of emotionally disturbed children are already under severe strain and we did not wish to threaten their already damaged parent-egos. We did not want them to feel the road to treatment was full of barriers and hazards.

Now, what is a short questionnaire and what is a long one? Of the forms reviewed the average number of questions per form was 100. The fewer items included, the fewer the behavior areas adequately tapped, so we settled on 105 items. In order to offset this ample number of items we avoided overcrowding, which seems to be a common fault of many questionnaires. The blocks in which the parents check their answers are large and, surprisingly, the form looks easy to complete. It is not formidable.

There is nothing more aggravating, when completing a form, than to have something like the following: "If you answered question No. 18 on page 2 'Yes' do not answer questions 27–39 but move on to question 40–69 omitting questions 48–53." We studiously avoided any such exercise in direction following and all of our questions are seemingly non-contingent items. I say, "seemingly" because the questions are interrelated in order to measure a number of areas of behavior. However, for the rater, how he answered one question has no bearing on how he answers those that follow.

The professional can answer inferential questions. The untrained will have great difficulty in doing this and so all our items have to do with present, observable behavior; for instance, "wets bed," "physically attacks adults," "sets fires" or "will not leave own room." Soggy sheets, black and blue shins, blazing wastebaskets and a pubescent Peter-the-Hermit are observable. No inferences are needed. All of this behavior is rated on a five-point scale ranging from "constantly" to "never" and we always deal with the child's present behavior. By using only present behavior we can measure behavioral changes over a specified period of time. The multiple-point rating scale as opposed to the two-point scale of the True—False variety makes it possible to measure changes in frequency of any given behavior over a period of time. It also permits a refined measure of agreement among raters. High agreement among raters is assured also by the use of present behavior. Historical data which is subject to the vagaries of memory can be better obtained in social histories. It is not infrequent that parents, who have a number of children, will become confused in trying to recall which one did what.

The parents are asked to complete the questionnaire at the time of their initial contact with the Center. They complete it again just prior to the Interpretive Interview. By this time, they have brought their youngster in for a diagnostic evaluation. We wonder whether the act of applying and
the diagnostic evaluation in any way effect either the child's behavior or how the parents see their child. The diagnostic evaluation can often be underestimated as to its therapeutic value. It is perhaps more frequent than less that the social worker, when he asks the parent: "Now tell me what Johnny does well," is answered with a surprised look or a comment such as, "Well, I really can't think of anything" or "He colors well" or "He can be sweet." It is not that the parent is callous or ignores his child; the problem is that he is daily faced with so many negative experiences with the youngster that he loses sight of his assets.

If the child is accepted for treatment the questionnaire is completed by the parents, the child's therapist, the parents' therapist, his teacher, and jointly by a nurse and child care technician every three months of the youngster's stay in residence. Upon discharge, the parents and teacher complete the form at intervals of six months, one year and two years.

It has been noted that the form seems heavily loaded on the side of pathology. A basic reason for the development of the questionnaire was to devise a reliable method to help in the selection of candidates suitable for treatment in our setting—not to measure behavioral characteristics of normal children. It will also be used to guide us in selecting therapeutic, and milieu goals; it and will be used to help in studying the effectiveness of the outcome of our therapy programs.

The item selection was based on a variety of commonly stated dimensions and frequently appearing factors of emotionally disturbed behavior in children.

Procedure and Results

The dimensions which guided our selection of items were chosen from several lists of behavioral terms compiled by three members of our Research Committee. A psychiatrist, psychologist and social worker, each laboring independently, listed a broad variety of general terms or categories that they felt described the disturbed behavior of children at the Center. Such categories included, for example: hostility-aggression, depression—self-destruction, withdrawal, sexual disturbance, and learning disturbance. Then, using the lists of categories as a guide, each worker created items of behavior he felt had been characteristic of children at the Center. We will refer to their suggested categories and sub-items of behavior as our original behavior scales. Thus, for example, the original behavior scale of the withdrawal category contains such items as: "daydreams," "avoids people," "hides when visitors come," and "will not talk to children outside of family." The questionnaire has been utilized in two pilot studies related to the problem of selection of children suitable for short term residential treatment.

Factor Analysis

The first pilot study we discuss was directed at finding out whether we were using appropriate labels to group the 105 items in the questionnaire. It was necessary that the items be grouped into a manageable number of categories. In order to accomplish this, the Department of Biometry of our Health Sciences Center performed a computer analysis of the answers of 159 parents to our questionnaire.* The program requested the computer to select the groups of items having something in common. The computer grouped the items in 18 different ways. The items of one series were found to be mathematically related to each other with hostility-aggression as the clearly predominant content. Items of another collection were found to be related in a second distinct way. The common feature of this second group was depression—verbal self-attack. The computer told us that, for example, the following items had something in common: "threatens to kill self," "threatens to injure self," "talks of how worthless he is," "talks about wanting to die," "overly critical of self," and "is picked on by other children." The third group of related items very definitely had withdrawal as a common characteristic. To summarize, the computer told us that certain items were related to each other in various groups. We examined the content of each group of related items and labeled the group according to what the items seemed to have in common.

The computer method used here is referred to as factor analysis. In our case, factor analysis is a mathematical endeavor to determine the number of different ways that our behavioral items can be grouped or classified. It is as though we have a large basket of blocks of different shapes and colors. By examining the characteristics of a particular group of blocks which the computer gathers for us, we might find that the blocks within it are similar in some respect. Thus, for example, they might all be some shade of green. Another grouping might have something different in common, such as the characteristic of roundness. Admittedly, this analogy may not be mathematically sophisticated, but we hope it serves to illustrate the point that the computer identifies the items that group together. We wish to emphasize that in spite of the computer's talents a trained clinician is necessary,

* We are indebted to John Howell and George Cobb of the Department of Biometry for developing the program and performing the analysis.
with his experience and knowledge of personality and disturbed behavior, to discover and appropriately label the essential nature of each item group.

Following the completion of this factor analysis through which groups of items with a common feature are derived, the next step for us will be to construct scales from each group of items. A child’s ranking on any given scale is equal to the sum total of the ratings which his parent gave him on the individual component items. Thus far we have firm basis for constructing scales of hostility—aggression, depression—verbal self-attack and withdrawal. We have distributed to each of five therapists a series of lists of items comprising each of the 18 factor-analytic groupings. For each group, the items are written out in a list one above the other. The items written out higher on the list are those which the computer has indicated contribute more to the particular item group. The therapists will consider this numerical contribution of the items to their groups, in addition to the content of the items, in giving a name to the whole group.

Once the various scales have been named, items will be dropped which may no longer mathematically belong on these scales as shown by a repeat factor analysis with a new set of parents. New scales may be added based upon a factor analysis of questionnaires filled out by Center staff. These scales would apply when the questionnaire is filled out by staff rather than parents.

After the scales have been refined, they will lend themselves as measures of behavior in various studies. The scale scores will aid in providing objective measures to be used in the assessment of the effectiveness of the therapy program for each individual child. This will be accomplished by noting the changes in the child’s scores on the scales every three months, relative to his therapeutic goals. For example, one would hope that the level of the withdrawal score of a shy, depressed child would drop and that his hostility-aggression score would rise in relationship to a decreasing score on depression—verbal self-attack. These scores will provide an objective, reliable way to assess progress in therapy. They will also serve as a guide in planning a child’s treatment program. For example, a number of children with high aggression scores might be placed in separate classrooms.

Once objective behavior change measures have been obtained, using our scales, they can be used in studies concerned with prediction of therapy progress. It should be valuable to study the scores representing therapy change in relationship to scores on the tests that were previously given to the parents on the day of the diagnostic evaluation. For example, we give the parents the Minnesota Multiphasic Personality Inventory (MMPI). It may prove valuable in study of the shapes and elevations of the parent MMPI profiles as they relate to the subsequent change or lack of change in their children’s questionnaire scale scores. Eventually, we would hope to use this MMPI information and other parent test information in objectively predicting with some validity the child’s therapy progress in a short-term residential setting.

Regression Analysis

The second pilot study was also related to the problem of screening of candidates for admission to the Center. This study was concerned with whether the computer could find individual items on our questionnaire which would prove particularly useful in choosing children suitable for short-term residential treatment. The computer’s job was to find a small, manageable group of 20 items which would best select treatment candidates. Its next job was to specify the best combination of these items. These 20 items should, when put in a simple formula, best separate children accepted for treatment from those not accepted because they are too disturbed. This classification would be accomplished solely on the basis of questionnaires filled out by the parents on initial contact with the Center.

The formula which George Cobb derived for us using the Department of Biometry computer worked well in separating the cases upon which the formula was based. The separation of 64 cases into the accept and too-disturbed groups would have occurred by chance alone only one time in 100. The formula was then checked for accuracy using a new sample of children. It blindly classified 20 out of 24 children correctly as to whether they had been accepted for treatment or referred elsewhere because they were too disturbed. We concluded that this type of analysis, referred to as regression analysis, may prove quite helpful to us. Our regression formula should become increasingly more accurate as it is revised by the addition of more cases as its basis.

We now plan to apply a similar regression analysis program in developing formulae which predict actual therapy progress rather than merely the judgements of the Screening Committee and evaluation teams as to expected therapy progress. To this end, we plan to develop formulae which discriminate groups of children whose questionnaire score changes reflect actual therapy progress from those children whose scores suggest minimal progress or increasing disturbance. These formulae will have the advantage of predicting therapy change ahead of time in specific areas, such as aggression, withdrawal, and depression, by merely using the questionnaires initially filled out by the parents before the child enters the program. Little staff time will
be required, since the parent does most of the rating, leaving only the scoring to trained personnel.

The feature which the above two studies have in common is the emphasis on predicting therapy progress. The first study will use the parents’ MMPI scores to predict therapy changes as measured by our questionnaire’s factor analytically derived scales. The second study uses a regression formula, incorporating individual questionnaire items answered by the parents upon application to the Center, to predict therapy change. It is intended that both the MMPI scores and regression formulae scores will be used only as additional aids and will not replace any of the careful thought and evaluation of each individual case being screened. The above mathematical techniques are designed to provide objectives and valid supplements to the present information which the Screening Committee and diagnostic teams have at their disposal in selecting appropriate candidates for short-term residential treatment.

Reference