

A Behavior Rating Scale for the Preschool Child¹

LENORE BEHAR² AND SAMUEL STRINGFIELD

Child Development Institute
University of North Carolina at Chapel Hill

The increase in the number of children attending preschools or day care centers has provided an opportunity for early detection of emotional problems. However, few screening instruments have been standardized for children in this age group. A modification of Rutter's Children's Behaviour Questionnaire was standardized on a sample of 496 normal and 102 disturbed preschool children. The modified questionnaire (renamed the Preschool Behavior Questionnaire) was found to possess criterion validity and high interrater and test-retest reliabilities. Three factors were extracted and were labeled Hostile-Aggressive, Anxious-Fearful, and Hyperactive-Distractible. The data indicated that the Preschool Behavior Questionnaire may be valuable in either clinical or research settings for the first step in early detection of emotional problems.

With many preschool-aged children now attending nursery schools, day care centers, and kindergartens, there exists an increasing opportunity to screen children earlier for developing behavior problems. However, it is frequently difficult to identify children with early signs and symptoms of emotional problems in this age group, for few standardized instruments exist to assess the social emotional functioning of young children or to differentiate normal behavior from deviant behavior in this age group.

Many checklists, rating scales, and other screening instruments have been developed for school-aged children (Digman, 1963; Peterson, 1961; Rutter, 1967; Schaefer, Droppleman, & Kalverboer, 1965; Spivack & Swift, 1966; Stott & Sykes, 1956; Walker, 1967; Werry & Quay, 1969), but few such in-

struments exist for the preschool-aged child (Eisenberg, Landowne, Wilmer, & Imber, 1962; Kohn & Rosman, 1972; Levine, Freeman, & Lewis, 1969; D. H. Stott, 1960). These instruments have been developed for a variety of purposes; therefore some are useful in the study of normal personality development, while others do, in fact, represent screening tools to differentiate normal behavior from disturbed behavior. However, it is difficult to select from those instruments in the latter category one that is applicable to the preschool child and that also has been standardized on both a normal and disturbed population and, in addition, is brief enough to be used as a screening tool by a teacher.

Thus, the development of the present instrument represents an attempt to carry assessment procedures for the preschool child one step further and provide a tool to be used by preschool teachers and child care workers to identify children who show symptoms that may suggest emotional disturbance. The criteria used in the development of the scale were that it (a) show validity in discriminating between normal and deviant populations; (b) show both interrater and test-retest reliability; (c) be standardized across a wide range of preschoolers, age 3 to 6, male or female,

¹ This research was supported by Grant 04-000628-01-0 from the National Institute of Mental Health and by Grant 5501RR05406 from the U.S. Public Health Service, National Institute of Health. The authors would like to thank Michael Rutter for permission to expand and modify the Children's Behaviour Questionnaire. The authors are grateful to the staffs of the 24 centers who participated in this project.

² Requests for reprints or the Preschool Behavior Questionnaire and manual should be sent to Lenore Behar, Biological Sciences Research Center, Child Development Institute, University of North Carolina, Chapel Hill, North Carolina 27514.

black or white; and (d) be brief, so that a teacher or other rater could conveniently use the instrument.

The present scale represents a modification of the Children's Behaviour Questionnaire, a 26-item behavior checklist previously standardized in England on elementary school children (Rutter, 1967). On Rutter's scale for each behavior described, the rater checks either "does not apply," "applies sometimes," or "frequently applies." The 3-point scaling system was retained in the present scale.

After consultation with the teachers in six local preschools for normal children and 4 teachers in a local therapeutic preschool for emotionally disturbed children and after reviewing the items on several similar scales, 10 new items were added to Rutter's checklist and the wording of several of Rutter's items was changed. The additions and changes were not made to reflect any particular theoretical positions but were chosen because the authors in consultation with 10 experienced preschool teachers felt that each new question tapped a problem behavior that occurred frequently enough to be considered separately for the preschool-aged child. To avoid confusion with Rutter's scale, the new scale was named the Preschool Behavior Questionnaire (PBQ).

Method

The standardization of the scale consisted of determining the reliability and validity of the scale on a stated population. Feeling that content or face validity was too subjective, the authors sought to determine the criterion-related validity of the PBQ; that is, how well the PBQ could discriminate between a normal and a deviant sample. The normal population was defined as children in any preschool that served the general public

and was not intended specifically for the care of the autistic, emotionally disturbed, retarded or otherwise special segments of the preschool population. The deviant population was defined as those children who had been previously diagnosed professionally as behavior disturbed, again excluding children who were primarily retarded, autistic, or other such cases.

Clearly there were problems inherent in such a criterion for differentiation. It seemed extremely probable that the normal group thus defined would contain many children who were in fact deviant but who had never been diagnosed as such; moreover, the deviant group might well contain a few relatively normal children who had been improperly diagnosed or some previously disturbed children now nearing the end of a successful treatment program. The additional problems presented by using a population of children in a treatment program as a comparison group involved the possibly biased reporting of the teacher-rater. Two assumptions were possible: (a) The teachers in these facilities knew that the children were disturbed and therefore would rate them as more disturbed, or (b) the teachers in these facilities were used to disturbed behavior and therefore would rate the children as less disturbed. In the first case, the children would be rated as more disturbed because of a halo effect; in the second case, the children would be rated as less disturbed because of an habituation effect. These two types of effects are the base of all rating procedures, and unfortunately there was no way to insure that either or both of these two types of influences were operating and influenced the results to an unknown degree.

Sample

The normal sample of 496 children was selected from 5 preschools in Durham, North Carolina, and 2 preschools in Portland, Oregon. An effort was made to select schools from diverse areas of the cities so that they represented socioeconomic groups ranging from lower- to upper-middle-class families. The samples as shown in Table 1 were drawn to be roughly comparable to the general population in terms of the number of white and black, male and female children. The emotionally disturbed sample was obtained from 15 preschools throughout the country involved in early intervention work with behavior-disturbed children. From the total sample of 124 subjects, 22 were eliminated, as their primary diagnosis was mental retardation.

Table 1: Breakdown on Subjects in the Preschool Behavior Questionnaire Validity Study

| Age (in years) | Subjects (598) ^a | | | | | | | | | | | |
|-------------------|-----------------------------|--------------|----------------|---------------|--------------|---------------|---------------|--------------|---------------|--------------|--------------|---|
| | Normal (496) | | | | | | Deviant (102) | | | | | |
| | Male (261) | | | Female (223) | | | Male (71) | | | Female (27) | | |
| White (207) | Black (50) | Other (3) | White (155) | Black (66) | Other (1) | White (47) | Black (14) | Other (0) | White (16) | Black (7) | Other (2) | |
| 3 (100) | 46 | 8 | 2 | 27 | 5 | 1 | 3 | 1 | — | 3 | 1 | — |
| 4 (158) | 55 | 7 | 1 | 56 | 13 | — | 10 | 5 | — | 3 | — | 1 |
| 5 (178) | 63 | 9 | — | 41 | 20 | — | 22 | 7 | — | 5 | 3 | 1 |
| 6 (65) | 22 | 4 | — | 14 | 4 | — | 12 | 1 | — | 5 | 3 | — |

Note. The figures in parentheses indicate the number of subjects included per category.

^a The sample of 598 included 122 children on which information on age, race, and/or sex was not made available.

dation, leaving a sample of 102 preschoolers whose primary diagnosis was emotional disturbance. Teachers of the children in both groups were asked to fill out the PBQ on each child in their classes.

Results

Each item on the PBQ was scored by counting "does not apply" as 0, "sometimes applies" as 1, and "frequently applies" as 2. The total score for each subject was obtained by addition of item scores. From the sample of 496 children in the normal group, the resulting mean was 9.12 with a standard deviation of 7.67. For the 102 children in the deviant group, the mean was 23.36 and the standard deviation, 7.30.

Validity

Each item and then the total scale score on the PBQ were analyzed using a chi-square test to determine whether the teacher's ratings of children significantly differentiated between normal and disturbed groups. Significance was defined as alpha level less than .01. As can be seen in column 5 of Table 2, 31 of the 35 items differentiated significantly between the two groups; moreover, the overall scale differentiated beyond the .0001 level of significance between the children in normal preschools and children previously diagnosed as disturbed with the disturbed children scoring higher. The PBQ appears to have demonstrated sufficient criterion validity.

The multiple regression technique was useful in determining, first, which items best differentiated between the deviant and normal groups and, second, what proportion of the total sample variance could be accounted for by the deviant-normal dimension. The results of the regression are summarized in the two right-hand columns of Table 2. In Table 2 simple regression is synonymous with zero-order correlation with the criterion being group membership. Multiple regression rank refers to the order in which the stepwise multiple regression included the variables; that is, which variable could account for the greatest amount of variance, given the variance accounted for by the variables previously included.

The item that best discriminated between

the normal and disturbed groups was "other speech difficulty." This supports Brown's (1960) finding that the most typical characteristic of disturbed preschoolers was the odd quality of their speech. Eisenberg et al. (1962) and Rutter (1967) have also found that items pertaining to language development discriminated well between deviant and normal populations. In like manner, identical or paraphrased versions of all the items on the PBQ with a regression ranking of 25 or above have been found useful in the identification of deviant children by Brown (1960), Digman (1963), Eisenberg et al. (1962), Kohn and Rosman (1972), Peterson, Quay, and Cameron (1959), Peterson (1961), or Rutter (1967). These items appear to be valid measures of deviant behavior in a variety of settings.

With all items included and with group membership (normal versus deviant) as the criterion variable, the PBQ obtained a total multiple regression of .734; thus 53.9% of the variance in the 36 items can be accounted for as group difference. Given the probable presence of a few relatively normal preschoolers in the disturbed groups and undoubtable presence of disturbed children in the normal population, it would seem that a multiple regression of .734 is acceptable.

Factor Analysis

Factor analysis has been frequently used in the development of psychological scales for children (for reviews see Kohn & Rosman, 1972; Peterson, 1961; Walker, 1967). Peterson found that when scales attempt to assess personality problems in older children and adolescents, two factors typically emerge. One factor, which he labeled Conduct Problems, includes such items as disobedience, disruptiveness, destructiveness, and uncooperativeness. This dimension closely resembles the like-named factor isolated by Himmelweit (1953) and Unsocialized Aggression as defined by Hewitt and Jenkins (1946). Peterson's second factor, Personality Problems, includes feelings of inferiority, lack of self-confidence, and social withdrawal. This second factor also resembles Himmelweit's Factor 2 and is much like the Overinhibited

Table 2: Means and Standard Deviations for Normal and Deviant Populations, Mean Group Differences, Simple Regression, and Multiple Regression Rankings of the Preschool Behavior Questionnaire Items and Total Score

| Item ^a | Normal | | Deviant | | M_{diff} (deviant - normal) | Simple regression | Multiple regression rank |
|-------------------------------------|----------|-----------|----------|-----------|----------------------------------|-------------------|--------------------------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| 1. Restless | .64 | .70 | 1.08 | .77 | .43*** | .22 | 16 |
| 2. Negativistic ^b | .41 | .63 | 1.12 | .68 | .70*** | .39 | 29 |
| 3. Squirmy | .51 | .65 | 1.09 | .82 | .58*** | .30 | 18 |
| 4. Destructive | .14 | .40 | .44 | .57 | .30*** | .25 | 30 |
| 5. Fights | .37 | .56 | .62 | .63 | .25** | .16 | 31 |
| 6. Disliked | .12 | .37 | .49 | .58 | .36*** | .32 | 12 |
| 7. Worries | .22 | .48 | .39 | .66 | .17* | .13 | 27 |
| 8. Solitary | .57 | .68 | 1.07 | .76 | .50*** | .26 | 20 |
| 9. Irritable | .27 | .52 | .78 | .77 | .51*** | .32 | 10 |
| 10. Unhappy | .22 | .46 | .68 | .68 | .45*** | .32 | 33 |
| 11. Twitches | .06 | .27 | .32 | .68 | .27*** | .26 | 11 |
| 12. Sucks thumb ^b | .20 | .50 | .31 | .66 | .11 | .08 | 35 |
| 13. Bites nails | .07 | .29 | .21 | .53 | .14* | .15 | 17 |
| 14. Often absent ^b | .08 | .32 | .12 | .40 | .03 | .03 | 34 |
| 15. Disobedient | .34 | .54 | .93 | .66 | .59*** | .37 | 4 |
| 16. Poor concentration | .50 | .62 | 1.31 | .68 | .82*** | .44 | 15 |
| 17. Fearful | .34 | .57 | .98 | .74 | .64*** | .37 | 36 |
| 18. Fussy | .19 | .44 | .50 | .71 | .31*** | .23 | 14 |
| 19. Lies | .12 | .37 | .31 | .54 | .20*** | .18 | 7 |
| 20. Steals ^b | .04 | .22 | .11 | .38 | .07 | .10 | 13 |
| 21. Soils self | .08 | .32 | .45 | .71 | .37*** | .32 | 5 |
| 22. Complains of aches ^b | .14 | .39 | .07 | .29 | -.07 | -.07 | 9 |
| 23. Tearful on arrival ^b | .15 | .40 | .31 | .66 | .16*** | .13 | 32 |
| 24. Stutters | .05 | .26 | .21 | .55 | .16*** | .18 | 19 |
| 25. Other speech difficulty | .16 | .51 | 1.18 | .91 | 1.01*** | .54 | 1 |
| 26. Bullies | .21 | .48 | .44 | .64 | .24*** | .14 | 28 |
| 27. Inattentive | .55 | .61 | 1.28 | .64 | .73*** | .41 | 23 |
| 28. Does not share | .38 | .56 | .92 | .67 | .54*** | .33 | 24 |
| 29. Cries easily | .33 | .54 | .79 | .78 | .46*** | .29 | 22 |
| 30. Blames others | .39 | .56 | .47 | .70 | .08* | .44 | 8 |
| 31. Gives up | .29 | .53 | 1.01 | .67 | .72*** | .44 | 6 |
| 32. Inconsiderate | .25 | .51 | .66 | .68 | .40*** | .27 | 26 |
| 33. Sexual problems | .03 | .16 | .08 | .34 | .05* | .10 | 21 |
| 34. Kicks, hits | .27 | .54 | .58 | .67 | .31*** | .20 | 25 |
| 35. Stares into space | .14 | .39 | .93 | .75 | .79*** | .53 | 3 |
| 36. Behavior problems | .26 | .50 | 1.13 | .69 | .86*** | .52 | 2 |
| Total | 9.12 | 7.67 | 23.36 | 7.30 | 14.24*** | | |

Note. Multiple $R = .74034$; $R^2 = .54910$.

^a Items are presented in abbreviated form here.

^b Omitted in the shortened version of the Preschool Behavior Questionnaire.

* $p \leq .01$.

** $p \leq .001$.

*** $p \leq .0001$.

Behavior dimension found by Hewitt and Jenkins.

Recently Kohn and Rosman (1972), using their Symptom Checklist on over 1,000 New York City preschoolers, found essentially the same 2 factors in their data. Schaefer et al. (1965) have argued for a 3-factor design with the relevant dimensions being Introversion versus Extroversion, Hyperkinetic-Distractible versus Hostile, and Adjustment versus Maladjustment, which they also

called Positive, Loving Behavior versus Hostility. Becker (1960), using a normal preschool sample and a 73-item semantic differential, found that 5 factors, Hostile Withdrawal versus Warm Extroversion, Relaxed versus Nervous Disposition, Submission versus Dominance, Lack of Aggression, and Conduct Problems, best summarized his data. Still others, using longer scales, have derived more than 5 factors. L. H. Stott (1962), for example,

found 14 significant factors in his analysis of a 220-item scale.

Digman (1963, 1965) presented a broad summary of many attempts to factor analyze data on childhood behavior and concluded that in the past most studies presented one of two solutions: simple (i.e., 2 or 3 factors) or complex (i.e., 8 or more factors). Digman further pointed out that there has been a parallel situation in the analysis of adult behavior; for example, Eysenck (1953) used 2 or 3 factors, and Cattell and Coan (1957) typically worked with models involving at least 12 factors. Digman (1963, 1965) attempted to resolve these different models by performing two studies that involved factor analyzing data on early primary students. He extracted all factors with eigenvalues greater than 1.00 in the two studies, obtaining 11 and 8 factors, respectively, and then factored the factors to arrive at second-order factors, obtaining 3 in each study. In both cases, Digman's first two second-order factors were Successful versus Unsuccessful Socialization and Extroversion versus Introversion, which he also called Freedom of Movement versus Restraint. The third factor in his first study he called Sex, noting that boys are more hostile and less compliant than girls. In his second study his third factor was called Anxiety.

Those who have obtained complex (eight or more) factor solutions have used what Digman chose to call the rule of 1.00. That is, regardless of the factor-analytic procedure they followed, they extracted every factor with an eigenvalue greater than or equal to 1.00. The rule of 1.00, also known as the little jiffy criterion, is not, however, a rule of factor analysis. Rather it is one of several frequently used criteria for determining the number of factors a researcher may extract from a data set. It is a method that when the data to be analyzed consist of a large number (typically 30 or more) of scales and items, leads to what Digman called a complex solution (i.e., eight or more factors). Interestingly, Digman violated the rule of 1.00 himself and in his 1965 article worked with eight factors for theoretical reasons when the rule dictated seven factors.

Those researchers who have obtained a

simple (two or three) factor solution have done so largely for theoretical reasons. Kohn and Rosman (1972) rotated nine factors but reported only two because "our interest was primarily in a two-factor model [p. 433]." Peterson (1961), using his own criteria plus his theoretical beliefs rotated two and then five factors, found the two-factor model to be more stable and used it. It would seem that frequently researchers let their theoretical perspectives overrule their statistical results.

In choosing to rotate nine factors, Kohn and Rosman explained that beyond this point the latent roots leveled out. This is a root number, root plot analysis and, given that this type of analysis frequently leads to a solution involving fewer and more stable factors than the rule of 1.00, in the present study the authors chose to use this type of factor analysis and further decided to analyze all of the factors that this system produced.

Using all of the subjects, both normal and deviant, from the above analysis, the data was factor analyzed using a principle-component analysis, and an examination of the root plot led to a three-factor solution. These three orthogonal factors were then varimax rotated (Kaiser, 1958). The three rotated factors of the PBQ accounted for 37.7% of the total variance of the scale. The next factor accounted for 4.4% of the total variance.

Each of the three major factors of the PBQ was unipolar. To illustrate what was measured by the factors, the items that showed the highest loadings on each factor are presented in Table 3.

Factor 1 appeared to measure a Hostile-Aggressive dimension. The items loading highest on Factor 1 indicated lack of consideration for others, irritability, and fighting with peers. Factor 1 appeared to be almost identical to Peterson's (1961) Conduct Problems and the inverse of his Lack of Aggression, similar to the Hostility pole of Schaefer et al.'s (1965) Factor 3 and Kohn and Rosman's (1972) Factor 2 on their Symptom Checklist.

Factor 2 included such items as fearful, unhappy, cries easily, and stares into space and seemed to be an Anxious-Fearful

Table 3: Preschool Behavior Questionnaire Items with the Highest Loadings on the Three Factors

| Factor and item | Factor | | |
|--|--------|------|-----|
| | 1 | 2 | 3 |
| Factor 1 | | | |
| Inconsiderate of others | .78 | .16 | .16 |
| Fights with other children | .77 | -.03 | .13 |
| Destroys own or others' belongings | .70 | .12 | .21 |
| Bullies other children | .71 | .00 | .05 |
| Kicks, bites, hits other children | .68 | .08 | .20 |
| Does not share toys | .65 | .24 | .21 |
| Blames others | .64 | -.03 | .07 |
| Factor 2 | | | |
| Tends to be fearful or afraid of new things or new situations | .06 | .66 | .14 |
| Appears miserable, unhappy, tearful, or distressed | .19 | .66 | .06 |
| Stares into space | .04 | .57 | .37 |
| Cries easily | .24 | .48 | .14 |
| Gives up easily | .16 | .47 | .40 |
| Factor 3 | | | |
| Inattentive | .19 | .24 | .80 |
| Has poor concentration or short attention span | .12 | .26 | .80 |
| Restless, runs about or jumps up and down, does not keep still | .36 | .02 | .69 |
| Squirming, fidgety child | .37 | .09 | .68 |

dimension. Again all of the items or their synonyms that loaded highly on this dimension also loaded highly on Peterson's Personality Problems dimension; moreover, Factor 2 was similar to Kohn and Rosman's Factor 1. Thus the first two dimensions of the PBQ strongly resembled both Peterson's and Kohn and Rosman's two dimensions. Peterson's statement as to the enormous generality of these two factors appears to have been well documented since he made it 12 years ago.

In addition, the PBQ contained a third significant factor that was characterized by poor attention span and restlessness. Factor 3 appeared to measure a Hyperactive-Distractible dimension. Interestingly, Peterson used several items nearly identical to those in Factor 3 in his research, and none of them loaded highly on either of his two dimensions; yet for theoretical reasons Peterson maintained a two-factor model. Factor 3 was quite similar to two of D. H. Stott's dimensions, Restlessness and Lack of Staying Power, and to the Hyperkinetic-Distractible pole of Schaefer et al.'s second factor.

To see to what extent the PBQ and its three factor scores yielded independent information, a set of correlations (Pearson's r) were computed using the following scoring system: Scores for each of the three subscales were computed by summing the items that loaded highest on those factors. The equations were Factor 1 (Hostile-Aggressive): sum of Items 4, 5, 6, 9, 15, 19, 28, 30, 32, and 34; Factor 2 (Anxious-Fearful): sum of Items 7, 10, 11, 17, 18, 25, 29, 31, and 35; Factor 3 (Hyperactive-Distractible): sum of Items 1, 3, 16, and 27.

It is indicated in Table 4 that while none of the factors were correlated as highly with any other factor as with the total PBQ score (from which each factor was a subset), all of the factors were moderately and positively correlated.

Because it was not statistically valid to test the validity of scales on the data from which they were derived, no analysis was made on normal versus deviant group scores on three factors. Such statistical analyses were performed during the replication-reliability study that follows, and there (see Table 6) the data for the two groups in this original study is listed with no statistical analysis.

Item Deletion

Having completed the major analyses on the standardization sample, the present authors concluded that the PBQ could be shortened without significantly affecting the validity of the instrument. To be included in the shortened version of the PBQ, an item had to (a) differentiate significantly on the chi-square test, which eliminated Items 12, 14, 20, and 22 and (b) either rank in the highest 25 on the stepwise multiple regression or have a factor loading higher than .55 on one of the three factors, which eliminated Items 2 and 23.

To test how much information was lost

Table 4: Correlation Amount, Preschool Behavior Questionnaire Total, and the Three Factors

| Factor | Total | Hostile | Anxious |
|-------------|-------|---------|---------|
| Hostile | .87* | | |
| Anxious | .77* | .38* | |
| Hyperactive | .74* | .50* | .45* |

* $p \leq .001$.

from the original 36 items by deleting the 6 items listed above, a canonical correlation between the total score on the long and short forms was computed; the result was a correlation coefficient of .990. In addition, the shortened form lost less than 1/2% of the discriminative power of the longer form on a multiple regression. The shortened form of the PBQ appears to lose very little information while cutting rating time by as much as one sixth.

With the deletion of the six items, a new set of totals was appropriate. The mean for the normal population became 8.007 with a standard deviation of 7.72 representing a decrease from 9.12 and 7.67, respectively. The mean for the deviant group became 21.324 with a standard deviation of 6.80 decreasing from 23.36 and 7.30. On the shortened PBQ a total of 33 preschoolers in normal settings, or 6.65% of the normal sample, scored above the mean for the deviant group; and 2 children previously diagnosed as deviant, or 1.97% of the deviant sample, scored below the mean of the normal sample. Given the problem stated in the introduction that the criterion groups would almost certainly contain a number of subjects who should have been in the opposite group, these percentages seemed adequately low.

Effects of Age, Race, Sex, and Group

In order to determine if there were any significant effects of age, race, sex, or group, an analysis of variance was computed using the shortened version of the PBQ. As can be seen in Table 5, males scored significantly higher than females ($M_{diff} = 2.15$ in the normal sample), and blacks scored significantly higher than whites ($M_{diff} = 1.80$). The normal versus deviant group differences have previously been discussed. Age as a variable was not significant. Levine et al. (1969), in looking at a normal sample of preschoolers, found that both sex and age were significant variables on their California Preschool Social Competency Scale. Levine did not use race as a variable.

It is the authors' belief that had the teachers been able to rate all of their children in comparison with all other preschool-aged groups on each item of the

Table 5: Analysis of Variance on the Preschool Behavior Questionnaire by Age, Race, Sex, and Group

| Source | df | MS | F |
|-----------------------|-----|-----------|-----------|
| Age (A) ^a | 1 | 144.228 | 3.274 |
| Race (B) ^b | 1 | 445.214 | 10.108* |
| Sex (C) | 1 | 943.086 | 21.411** |
| Group (D) | 1 | 10727.906 | 243.559** |
| A × B | 1 | 178.981 | 4.063 |
| A × C | 1 | .297 | .007 |
| A × D | 1 | 60.842 | 1.381 |
| B × C | 1 | 11.991 | .272 |
| B × D | 1 | 261.971 | 5.948 |
| C × D | 1 | 19.420 | .441 |
| A × B × C | 1 | 82.343 | 1.869 |
| A × B × D | 1 | 266.084 | 6.041 |
| A × C × D | 1 | 31.217 | .709 |
| B × C × D | 1 | 44.907 | 1.020 |
| A × B × C × D | 1 | 52.943 | 1.202 |
| Within | 460 | 44.046 | |
| Total | 475 | 13315.476 | |

^a In order for all cells to contain cases, it was necessary to collapse the variable age to age ≤ 4 and age ≥ 5.

^b Includes blacks and whites only.

* $p \leq .01$.

** $p \leq .001$.

PBQ as is done on the California Preschool Social Competency Scale, then most likely the children's scores on the PBQ would have diminished with age. However, while the authors were gathering the data, many teachers remarked that they could only rate the children relative to other children with whom they had been familiar, that is, children of the same age group. Thus, in considering an item such as "restless," the scores for 3-year-olds were similar to the scores for 5-year-olds. It would appear that this similarity in scores resulted not because the degree of restlessness was similar in these age groups but rather because an individual child's behavior was being rated as to its similarity to others in the same age group.

The sex and race differences that occurred on this scale were in agreement with frequently noted differences in the socialization process for different groups within our society. None of the interaction effects were significant.

Replication and Reliability Study

A second study was designed to replicate the previous results and to determine the interrater reliability of the PBQ and the three factors found in the previous research.

For the second study, 80 subjects for the

normal preschool sample were randomly drawn from three preschools not in the original sample. For the deviant sample, 9 new children were selected from a school for emotionally disturbed preschoolers. In each of the four schools, a teacher and an aide rated each child. All of the schools used in the second study were located in the area of Chapel Hill, North Carolina, and none had been used in the original study.

For the test of the validity of the three factor dimensions and the replication of the validity of the overall scale, the chi-square test was used and alpha was again set at .01. As can be seen in Table 6, the teacher's ratings of the children differentiated significantly between the normal and deviant groups on the PBQ and on each of the three subscales. The PBQ and the subscales all appeared to measure concepts or dimensions on which there were significant and rather large differences between normal and deviant preschoolers.

The mean scores for the two groups on each of the four scales were similar to those obtained in the previous study. The scales appeared to be stable across groups of raters.

On the average the aides did not differentiate as clearly between normal and deviant

behavior and showed a great deal more variance in their ratings of children. The only exceptions to this were the unusually well trained aides at the therapeutic preschool. Interestingly, in both normal and deviant settings, the aides rated the children an average of nearly 4 points higher on the PBQ than did the teachers.

Interrater reliability coefficients (Pearson's r) were computed for each of the seven classes, and a summary of these findings is presented in Table 7. The highest mean reliability rating was for the total score with the factor dimensions showing decreasing reliability with the decreasing size of the factors.

As previously stated, teachers and aides in seven classes completed the PBQ on all of their students whose ages fell between 3 and 6 years. Shortly after this set of data had been gathered, the authors discovered that whereas in six of the classrooms, the teachers and aides had been working together in their classes for the entire school year (nearly 20 weeks at the time), one teacher had been a new addition to her class only 5 weeks prior to filling out the PBQ. The authors, feeling that perhaps this was not sufficient time to get acquainted with a class, held the data on this class for separate

Table 6: Normal versus Deviant Groups: Means, Standard Deviations, and Mean Differences as Rated by Teachers and Aides

| Scale | Validity study | | | Reliability-replication study | | | | |
|---|----------------|----------|-----------|-------------------------------|----------|-----------|----------|-----------|
| | <i>n</i> | <i>M</i> | <i>SD</i> | <i>n</i> | <i>M</i> | <i>SD</i> | Aides | |
| | | | | | | | <i>M</i> | <i>SD</i> |
| Hostile | | | | | | | | |
| Deviant | 102 | 6.24 | 4.15 | 9 | 8.11 | 3.22 | 8.56 | 3.71 |
| Normal | 496 | 2.66 | 3.35 | 80 | 3.16 | 3.03 | 4.35 | 4.24 |
| Difference | | 3.58 | | | 4.95** | | 4.21** | |
| Anxious | | | | | | | | |
| Deviant | 102 | 6.73 | 3.61 | 9 | 6.22 | 3.49 | 5.89 | 3.26 |
| Normal | 496 | 1.96 | 2.34 | 80 | 2.42 | 2.40 | 2.51 | 3.05 |
| Difference | | 4.77 | | | 3.80** | | 3.38* | |
| Hyperactive | | | | | | | | |
| Deviant | 102 | 4.75 | 2.26 | 9 | 3.22 | 1.20 | 2.44 | 2.67 |
| Normal | 496 | 2.21 | 2.21 | 80 | 1.82 | 1.81 | 2.05 | 2.19 |
| Difference | | 2.54 | | | 1.40** | | .39 | |
| Preschool Behavior Questionnaire total | | | | | | | | |
| Deviant | 102 | 21.30 | 7.19 | 9 | 21.22 | 5.21 | 24.67 | 8.90 |
| Normal | 496 | 8.09 | 7.25 | 80 | 8.74 | 5.96 | 12.41 | 10.01 |
| Difference | | 13.21*** | | | 12.48*** | | 12.26* | |

* $p \leq .01$.

** $p \leq .001$.

*** $p \leq .0001$.

Table 7: Interrater Reliability of the Preschool Behavior Questionnaire

| Ratings ^a | Total | Hostile | Anxious | Hyperactive |
|----------------------|-------|---------|---------|-------------|
| Highest | .97 | .96 | .97 | .90 |
| Lowest | .67 | .61 | .42 | .30 |
| Mean | .84 | .81 | .71 | .67 |
| Class 7 | .42 | .29 | .55 | .30 |

^a Ratings are given for six of the seven classes of the sample; Class 7 is given separately because the teacher had worked with the children only five weeks.

analysis, and it is presented separately in Table 7 as Class 7.

Interrater (teacher-aide) reliability in Class 7 in which the teacher had worked with the students for five weeks was much lower than for the other classes. This finding suggests that a rater must be fairly familiar with any child before he can rate the child's behavior reliably on the PBQ. For the other classes the mean interrater reliability coefficients were in the moderate to high range on all of the scales.

A third study was initiated to measure the test-retest reliability of the PBQ. The sample for this study was drawn from a local preschool for normal children and from the therapeutic preschool that had been used in the replication-reliability study. The normal preschool had not been used in previous PBQ research.

A total of five teachers rated the children in their classes. In each case the second rating was three to four months after the first and was performed by the same teachers as the first rating. A summary of test-retest reliability findings analyzed using Pearson's r is presented in Table 8. All of the mean reliabilities were in the moderate to high range.

Discussion

In the search for screening instruments available to preschool teachers and day care workers, existing scales did not seem useful because either they were not designed for the preschool age range, they were inadequately standardized, or they were too lengthy. The PBQ was developed to serve as a short screening instrument for the identification of preschool children with behavior problems. The PBQ was standardized on a sample that included 496 children enrolled in normal

preschools and 102 children enrolled in a variety of special education programs designed to assist emotionally disturbed preschoolers across the country. Of the 36 items included in the scale, 32 items discriminated significantly between the two groups. The total score on the PBQ differentiated at the .0001 level of significance between the two groups, and 53% of the total variance in the scale could be accounted for by this between-groups difference. A factor analysis of the data revealed three significant unipolar factors, Hostile-Aggressive, Anxious-Fearful, and Hyperactive-Distractible. The similarity between these three factors and those found by other authors (Becker, 1960; Digman, 1963, 1965; Kohn & Rosman, 1972; Peterson, 1961; Schaefer et al., 1965) is noteworthy.

These factors have appeared repeatedly in studies within the limited age range of the preschool population. These factors have also appeared in very much the same form with the recorded problems of treatment cases (Hewitt & Jenkins, 1946; Himmelweit, 1953), and remarkably similar factors have appeared in the questionnaire behavior of delinquent boys (Peterson et al., 1959).

To provide replication data and reliability data, a second study involving preschoolers in normal settings and students in a therapeutic preschool was undertaken. Each child in the second study was rated by two raters, his teacher and the teacher's aide. A mean interrater reliability of .79 was derived for the overall scale and .76, .70, and .61 for Factors 1, 2, and 3, respectively. The PBQ total score and each of its three factor scores differentiated significantly between the two groups.

It would seem then that it has been demonstrated that the PBQ is a valid and reliable instrument and warrants use as a screening tool for teachers, day care work-

Table 8: Test-Retest Reliability of the Preschool Behavior Questionnaire

| Ratings | Total | Hostile | Anxious | Hyperactive |
|---------|-------|---------|---------|-------------|
| Highest | .98 | .99 | .90 | 1.00 |
| Lowest | .53 | .85 | -.15 | .86 |
| Mean | .87 | .93 | .60 | .94 |

ers, and others who have familiarity over time with the child. In addition, the PBQ offers interesting possibilities as a research tool to measure such variables as change as the result of intervention, or similarities or differences in subject groups.

REFERENCES

- Becker, W. C. The relationship of factors in parental ratings of self and each other to the behavior of kindergarten children as rated by mothers, fathers, and teachers. *Journal of Consulting Psychology*, 1960, **24**, 507-527.
- Brown, J. L. Prognosis from presenting symptoms of preschool children with atypical development. *American Journal of Orthopsychiatry*, 1960, **30**, 382-390.
- Cattell, R. B., & Coan, R. W. Child personality structure as revealed in teacher ratings. *Journal of Clinical Psychology*, 1957, **13**, 315-327.
- Digman, J. M. Principal dimensions of child personality as inferred from teachers' judgments. *Child Development*, 1963, **34**, 43-60.
- Digman, J. M. Child behavior ratings: Further evidence of a multiple factor model of child personality. *Educational and Psychological Measurement*, 1965, **25**, 787-799.
- Eisenberg, L., Landowne, J., Wilmer, M., & Iamber, D. The use of teacher ratings in a mental health study: A method for measuring the effectiveness of a therapeutic nursery program. *American Journal of Public Health*, 1962, **52**, 18-28.
- Eysenck, H. J. *The structure of human personality*. New York: Wiley, 1953.
- Hewitt, L. E., & Jenkins, R. L. *Fundamental patterns of maladjustment: The dynamics of their origin*. Springfield, Ill.: Green, 1946.
- Himmelfeit, H. T. A factorial study of children's behavior problems. In J. Eysenck (Ed.), *The structure of human personality*. New York: Wiley, 1953.
- Kaiser, H. F. The varimax criterion for analysis rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
- Kohn, M., & Rosman, L. A Social Competence Scale and Symptom Checklist for the preschool child: Factor dimensions, their cross-instrument generality, and longitudinal persistence. *Developmental Psychology*, 1972, **6**, 430-444.
- Levine, S., Freeman, F. F., & Lewis, M. *California Preschool Social Competency Scale*. Palo Alto, Calif.: Consulting Psychologists Press, 1969.
- Peterson, R. Behavior problems of middle childhood. *Journal of Consulting Psychology*, 1961, **25**, 205-209.
- Peterson, R., Quay, H. C., & Cameron, G. R. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. *Journal of Consulting Psychology*, 1959, **23**, 395-399.
- Rutter, M. A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 1967, **8**, 1-11.
- Schaefer, E. S., Droppleman, L. F., & Kalverboer, A. F. Development of a classroom behavior checklist and factor analysis of children's school behavior in the United States and the Netherlands. Paper presented at the meeting of the Society for Research on Child Development, Minneapolis, March 1965.
- Spivack, G., & Swift, M. S. The Devereux Elementary School Behavior Rating Scales: A study of the nature and organization of achievement related disturbed classroom behavior. *Journal of Special Education*, 1966, **1**, 71-90.
- Stott, D. H. A new delinquency prediction instrument using behavioral indications. *International Journal of Social Psychiatry*, 1960, **6**, 195-205.
- Stott, D. H., & Sykes, E. G. *Bristol social adjustment guides nos. 1 and 2. The child in school*. London: University of London Press, 1956.
- Stott, L. H. Personality at age four. *Child Development*, 1962, **33**, 287-311.
- Walker, H. M. Construction and validation of a behavior checklist for the identification of children with behavior problems. Unpublished doctoral dissertation, University of Oregon, 1967.
- Werry, J. S., & Quay, H. D. Observing the classroom behavior of elementary school children. *Exceptional Children*, 1969, **35**, 461-470.

(Received July 13, 1973)