1984

The Development and Validation of the Student Personality Assessment Form

Earl H. Baker
University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Baker, Earl H., "The Development and Validation of the Student Personality Assessment Form" (1984).
Open Access Dissertations. Paper 1036.
https://digitalcommons.uri.edu/oa_diss/1036

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

THE DEVELOPMENT AND VALIDATION OF THE
STUDENT PERSONALITY ASSESSMENT FORM

BY

Earl H. Baker

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
PSYCHOLOGY

University of Rhode Island

1984

Abstract

Recently, psychologists have become increasingly concerned with personality of preschool age and lower grade school age children. The most efficient method of measuring personality for this age child is through the use of ratings by others who are familiar with the child. The current project consists of four studies which describe the development and validation of an instrument designed to measure personality in children of kindergarten, first or second grade age. The first study describes the development of the Student Personality Assessment Form (SPAF) which was designed to measure seven traits identified in research with children or adults. Principal component analysis with Varimax rotation failed to reproduce the seven dimensions. Instead, three well defined components and a fourth less well defined component resulted. The SPAF was subsequently revised to reflect these results. Items were added to further define the fourth component. Administration of the revised SPAF resulted in replication of the three well defined components and two less well defined components. Subsequent validation studies were conducted on the three well defined dimensions which were tentatively named Compliance, Interpersonal Affect and Extroversion. The second study examined the relationship of the SPAF and Devereux Elementary School Behavior Rating Scale (DESB).

Results of zero order correlation analysis, multiple regression analysis and canonical variate analysis further defined, clarified and supported the SPAF traits. The third study used a multitrait-multimethod matrix to determine the relationship between questionnaires completed by teachers and parents. Results indicated low to moderate convergent validity similar to that found in similar research. Some aspects of discriminant validity were met and others were not. Implications of the low to moderate convergent validity coefficients and suggestions for improving discriminant validity are discussed. The fourth study focused on the relationship of the SPAF, DESB, and Meeting Street School Screening Test (MSSST) to objective and subjective measures of achievement. Results indicated the SPAF as the best predictor of objective measures and the DESB as the best predictor of subjective measures. Combining the scales increased the prediction of objective and subjective measures of achievement. The total project supports the validity and utility of the SPAF. Discussion of the total project involves theoretical implications, empirical implications, practical implications and limitations and future directions.

Acknowledgement

I wish to thank my wife, Ambia, for her continued support during the long process of meeting the requirements for the PhD degree. Without her support, none of this would have been possible. I also want to thank each of my children (Douglas, Caroline, James and Steven) for being understanding and quiet during many long weekends of study when they would rather have been enjoying my attention. Special thanks must also be given to my major professor, Wayne Velicer, who has been both a mentor and friend through this. His comments and suggestions have been invaluable. I also wish to thank the members of my committee, Janet Kulberg, John Stevenson, and Stewart Cohen, for their suggestions and support. A special thank you is also given to the parents and teachers of Coventry, R.I. who participated in the studies compiling this dissertation and to the members of the Coventry Special education staff for their suggestions.

Table of Contents

Acknowledgements.....	iv
List of Tables.....	ix
Section	
I. Introduction.....	11
II. Study One: Development of the Student Personality Assessment Form (SPAF).....	17
Method.....	17
Subjects.....	17
Instruments.....	18
Procedures.....	22
Results.....	23
Factorial Validity of the 43 item SPAF.....	36
Subjects.....	36
Procedure.....	36
Results.....	37
Discussion.....	46

III. Study Two: Relationship of the SPAF and the Devereux Elementary School Behavior Rating Scale (DESB).....	48
Method.....	48
Subjects.....	48
Instruments.....	49
Procedures.....	51
Results.....	52
Zero order correlation.....	52
Multiple regression analysis.....	56
Canonical variate analysis.....	61
Discussion.....	63
IV. Study Three: Relationship of the SPAF Completed by Teachers and Parents.....	69
Method.....	72
Subjects.....	72
Instruments.....	73
Procedures.....	73
Results.....	75
Factorial validity of the ParSPAF.....	75

Analysis of Convergent and	
Discriminant Validity.....	83
Convergent validity.....	83
Discriminant validity.....	85
Discussion.....	87
V. Study Four: Relationship of the SPAF and	
Other Scales to Academic Achievement.....	93
Method.....	97
Subjects.....	97
Instruments.....	98
Procedure.....	102
Results.....	103
Relationship of the SPAF, ParSPAF	
and MSSST.....	103
Relationship of Individual Predictor	
Sets to Achievement and Grades.....	105
Relationship of the Combined Predictor	
Sets to Achievement and Grades.....	108
Discussion.....	111
VI. Discussion and Implications.....	118

Theoretical Implications.....	123
Empirical Implications.....	128
Practical Implications.....	137
Limitations and Future Directions.....	142
References.....	146
Appendices.....	156
A. Limited Review of Representative Research.....	156
B. SPAF (90 item).....	183
C. Cover Sheet (90 item SPAF).....	189
D. SPAF (43 item).....	191
E. Cover Sheet (53 item SPAF).....	194
F. ParSPAF.....	198
G. Cover Sheet (ParSPAF).....	202
H. ParSPAF Instruction Sheet.....	206
I. ParSPAF follow-up letter.....	209
J. Bartlett's test for remaining eigen values.....	211
Bibliography.....	213

LIST OF TABLES

	<u>PAGE</u>
1. Varimax Rotated Component Patterns for the Ninety Item SPAF.....	24
2. Cross Classification of Items Between Components and Content Defined Scales of the Ninety Item SPAF.....	30
3. Items with Highest Loading Listed by Component of the Ninety Item SPAF.....	31
4. Varimax Rotated Component Pattern for the Forty-Three Item SPAF.....	39
5. Cross Classification of Items Between Components and Content Defined Scales of theForty-Three Item SPAF.....	42
6. Means, Standard Deviations, Alpha Coefficients and Interscale Correlations of the SPAF Scales..	45
7. Pairwise Correlation Coefficients of SPAF Scales and DESB Factors.....	53
8. Multiple Regression Coefficients and Order of Inclusion of Variables into the Stepwise Regression.....	57
9. Cancnical Variate Loadings.....	62
10. Varimax Rotated Component Pattern for the Forty-Three item ParSPAF.....	77

11.	Cross Classification of Items Between Components and Content Defined Scales of the ParSPAF.....	80
12.	Means, Standard Deviations, Alpha Coefficients and Interscale Correlations of the ParSPAF Scales.....	82
13.	Inter- and Intrascale Correlation Coefficients for the SPAF and ParSPAF.....	84
14.	Pairwise and Multiple Correlation Coefficients for SPAF and ParSPAF with MSSST Scaled Scores.....	104
15.	Multiple Regression Coefficients for Individual Predictor Sets with Achievement and Grades.....	106
16.	Multiple Regression Coefficients for Combinations of Predictor Sets with Achievement and Grades.....	110

INTRODUCTION

The study of personality has frequently focused on adults or older children. The major approach has been self-report forms. These are difficult to adapt for younger children, possibly explaining why personality research with pre-school and lower grade school age children has been limited. Recently psychologists have become increasingly concerned with personality for this age group and the amount of research has increased. This increase is partially the result of growing recognition that there are a number of potential uses for knowledge of younger children's personality styles in the school setting. Among the uses are: (1) an examination of the relationship of personality and cognitive development, (2) an examination of the relationship between school performance and personality styles, and (3) assisting teachers in the development of teaching strategies that take into account the relevance of various personality styles. Research on personality for pre-school and lower grade school children has relied on self-report as well as alternative methods such as: (1) interview of the subject (2) interview of persons knowledgeable of the subject, (3) projective techniques conducted with the subject, or (4) objective tests or questionnaires completed by persons knowledgeable of the subject. Examples of each of these approaches will serve to illustrate the advantages and disadvantages of each. The

focus of the present study is the latter method, because it is potentially the most efficient.

An example of the self report approach would be the scales developed by Cattell and his associates (Coan & Cattell, 1966; Porter & Cattell, 1959) such as the Children's Personality Questionnaire (CPQ) and Early School Personality Quiz (ESPQ). These instruments are completed by the children and demonstrate adequate test-retest reliability. However they are rather lengthy (60 to 100 minutes in two sessions for the ESPQ) and have little established external validity (Buros, 1978). A general difficulty with this approach is that the verbal skills limit the type of items that can be employed.

An example of an approach using direct interview of the subject is that conducted by Rutter and Graham (1968). They use a standardized interview format with nonclinic children and classified them as showing no psychiatric abnormality, some abnormality or definite psychiatric abnormality. Using this classification format resulted in good overall agreement between raters ($r=.84$). However, the agreement on specific items was much lower. Another part of that study looked at test-retest reliability with psychiatric patients and found the agreement on overall psychiatric state with four weeks between interviews to be .61. Again, agreement on specific items was much lower. These results indicate

that moderately reliable and valid classifications as to gross psychiatric condition can be made with standardized interview procedures. However, there is less agreement on clinical inferences which could be useful in determining diagnosis and treatment. This process is also inefficient, requiring a significant amount of professional time.

An example of a method using an indirect interview of persons knowledgeable of the subject was that developed by Garside, Birch, Scott, Chambers, Kolvin, Tweddle, & Barber (1976) for use with parents of young children. The technique used a semi-structured, open-ended interview conducted with the mother. The results were then converted to scores on scales which had been previously defined. The emphasis of the interview was on "how" the child interacted with his environment across situations to produce data on temperament. The interview was conducted two times at approximately a one month interval. Principal component analysis resulted in four main components across the two administrations. They were labeled: (1) withdrawal, (2) assertiveness, (3) high activity and (4) negative mood. Inter-rater reliability on these dimensions was good, ranging from .92 to .98. However, test-retest reliability was not as good. An interview conducted with the same parents approximately one month later showed that "about 50 percent of the 39 items reliabilities were above 0.8 and 90 percent above 0.06" (Garside et al, 1976, pg 38). This

procedure is also inefficient, again requiring a significant amount of professional time.

Projective techniques have been used for many years in the examination of personality with both adults and children. An example of this technique is the Thematic Apperception Test (TAT; Murray, 1943) which is widely used with both children and adults. With the TAT, the subject is asked to make up stories about pictures depicting various situations. These stories purportedly reveal needs or motives of the person being tested. Advantages of this technique are that the individual is allowed to reveal the underlying personality structures which reflect his organizational patterns and they are simple to administer. A major problem with this and other projective techniques is the limited reliability and validity of the method created by the difficulty interpreting the results (Achenbach, 1974). For example, a study by Cox and Sargent (1950) resulted in 11 of 15 normal subjects being classified as emotionally disturbed when their protocols were evaluated by experienced clinicians.

An alternative approach that avoids the inefficiency problems as well as the difficulty produced by the limited language development of young children is the utilization of a rating scale completed by an individual familiar with the subject. Carey (1970) developed a questionnaire to be

completed by parents which was designed to assess the temperamental styles of infants between the ages of four and eight months. The temperamental categories used were from the New York Longitudinal Study (Thomas, Chess & Birch, 1968). Test-retest reliability data were collected with an extremely small sample (N=3) and no reliability coefficients were given. Much work needs to be completed on this questionnaire before it is used extensively but it is also not applicable to the preschool and early school age child.

Another questionnaire was developed for children between the ages of three and seven years by McDevitt and Carey (1978). This instrument, the Behavioral Style Questionnaire (BSQ), is completed by parents. The authors report test-retest reliability for the complete instrument as being .89. The coefficients for the individual categories of temperament ranged from .67 to .94. Internal consistency coefficients ranged from .47 to .80. From this information, the BSQ appears to be adequately reliable, but is lacking internal consistency in some categories. Construct validity also needs to be determined.

The present study will focus on an item pool which was partially developed from the Teacher Temperament Questionnaire (TTQ). This instrument was based on the New York Longitudinal Study (NYLS; Thomas, Chess & Birch, 1968) which utilized three techniques of assessment. Data

collection was begun in 1956 and has been continued to the present. It has been gathered from interviews with parents and teachers, observations of the children and from objective testing. Qualitative analysis of the data resulted in the delineation of nine temperamental categories of reactivity or traits. The categories were: (1) Activity Level, the motor component present in a given child's functioning, (2) Rhythmicity (regularity), the predictability of a child, (3) Approach or Withdrawal, the nature of the initial response to a new stimulus, (4) Adaptability, the ease with which a child's responses are modified in the desired direction, (5) Threshold of Responsiveness, the level of stimulation necessary to evoke a response, (6) Intensity of Reaction, the energy level of a response, (7) Quality of Mood, the amount of pleasant behavior contrasted with unpleasant behavior, (8) Distractibility, how effective external stimuli are in interfering with ongoing behavior, and (9) Attention Span and Persistence, two related categories. Attention Span concerns the length of time a child stays with an activity. Persistence refers to the continuation of an activity in the face of obstacles to the maintenance of the activity. From these categories there were three temperamental constellations identified. There was the Easy Child, who showed regularity, positive approach to new stimuli, high adaptability and a mild or moderately intense mood which is usually positive. There was the Difficult Child, who showed

irregularity, negative withdrawal responses to new stimuli, non-adaptability or slow adaptability and intense mood expressions which were often negative. The third style of functioning was the Slow-to-warm-up Child. This type child showed negative responses of mild intensity to new stimuli with slow adaptability after repeated contact and was characterized by mild intensity of reaction, whether positive or negative and by less irregularity.

In conjunction with the study just described, the NYLS research group developed a questionnaire (Thomas & Chess, 1977) for use with children from three to seven years of age which can be completed by teachers in a very short time. The Teacher Temperament Questionnaire (TTQ) was designed to delineate eight of the nine categories of temperament delineated in the NYLS research. Rhythmicity is not identified because it requires knowledge of the child's functioning over a 24 hour period. The questionnaire was developed originally for parents by culling items of behavior from the NYLS interview protocols which seemed to be typical and unambiguous and did not overlap with other categories of temperament. These items were then given to twenty mothers with three to seven year old children and were then revised using comments and suggestions made by these mothers. The questionnaire was reduced to 132 questions using this procedure. The TTQ was developed by modifying these questions and making them applicable to the

school setting. Some items were dropped as totally inappropriate. Sixty questionnaires were then completed by a group of teachers. The results were reviewed and refined to determine whether the item should be included in the final form which was reduced to eight items for each of the eight traits.

Limited data exist concerning the reliability or validity of this instrument. In the development of the parent questionnaires, fifty mothers were asked to complete the questionnaire and were then interviewed using the NYLS interview format two weeks later. Another fifty mothers were interviewed and then completed the questionnaire two weeks later. The data generated were used to reduce the parent questionnaire from 132 items to 72 by keeping those items which showed significant correlation between the interview and the question. These data were not generated for the TTQ as there was no interview completed. The questions for the TTQ were reviewed by four judges who rated each item for the temperamental trait for which it should be scored. Items were deleted if the four judges did not agree as to which category the item was applicable. This provides minimal content validity. The only other reliability study reported for the TTQ was Hubert, Wachs, Peters-Martin, and Gandour (1982) which cited unpublished data and indicated that reliabilities for the categories ranged from .69 to .81. A recent concurrent validity study reported low-

moderate correlations ($r = .18$ to $.46$) between the categories of the TTQ and other instruments with similar categories (Billman & McDevitt, 1980).

Research by Baker and Velicer (1982) examined the reliability and construct validity of the TTQ. Construct validity was assessed by determining to what extent the eight theoretical scales could be reproduced empirically. The sample ($N=116$) was made up of kindergarten, first and second grade children in a southeastern Massachusetts town. Their teachers were asked to complete a TTQ for each child in their class. Four weeks later, they were asked to complete a second TTQ. A principal component analysis (PCA) with Varimax rotation was performed on both administrations. There were four stable components identified from each administration. They were labeled: (1) Compliance, which involved attentiveness, distractibility and persistence, (2) Extroversion, which involved shyness, comfort in novel situations and social interactions, (3) Interpersonal Affect, which involved competitiveness, cooperation with others and a tendency to be argumentative and (4) Environmental Sensitivity, which involved light and temperature sensitiveness and responsiveness to stories. Each of these components was composed of a mix of items from the theoretical scales of the NYLS (Thomas & Chess, 1977).

Test-retest reliability coefficients were computed for each of the four components. The results were: Component 1: .76, Component 2: .73, Component 3: .75, and Component 4: .86. Internal consistency was also determined, resulting in Alpha Coefficients of .97, .89, .88 and .45 for Components 1, 2, 3, and 4 respectively for the first administration and .97, .87, .85, and .76 for the second administration. The results of this study, then, did not support the theoretical scales of Thomas and Chess (1977) but did suggest that an alternative method of scoring could be used to produce scales with adequate reliability and validity.

Besides the analysis of the two administrations, Baker and Velicer (1982) presented a partial replication using a shortened version of the TTQ. This version included fifteen items from each of the previously described scales of Compliance, Interpersonal Affect and Extroversion. Five teachers rated 155 children as a part of a kindergarten screening. A principal component analysis with Varimax rotation resulted in four components which accounted for 65 percent of the variance. The first component was composed of eighteen items, fourteen of which were from the Compliance scale. The second component was made up of twelve items, eleven of which were from the Interpersonal Affect scale. The third component included nine items, all from the Extroversion scale. The fourth component was

composed of four items, three from the Extraversion scale and one from the Interpersonal Affect scale. As can be seen from these results, 34 of the 45 items were included on the same scale as the original study. Two of the remaining eleven items loaded on more than one component and can be excluded. Two items did not load on any component and four items were assigned to the fourth component. The other three items loaded on different components than those predicted from the original study.

It appears then, that the questionnaire developed by the NYLS project is capable of identifying stable personality traits. However, the research by Baker and Velicer (1982) indicated that the TTQ was delineating three clearly identified traits and a fourth which was less clearly identified rather than the eight hypothesized scales. The four components were made up of a mix of items drawn from the theoretical scales and demonstrated good test-retest reliability. They were labeled Compliance, Extroversion, Interpersonal Affect, and Environmental Sensitivity. This research indicated that the structure of child personality being identified by this instrument was complex and perhaps not limited to the four dimensions just described.

The present paper will describe several studies based on an extension of the Baker and Velicer (1982) results. The first study attempted to develop other trait measures that might be useful in the context of personality research. New areas of personality were selected for investigation on the basis of: (1) previous identification in adult scales, (2) previous identification in child scales and (3) practical utility as determined by personal experience and discussion with other professionals.

The development of a personality rating scale, however, does not end with the description and rationale for its dimensions. As Hubert et al. (1982) point out in their review of temperament instruments, there are several steps which must be taken to determine the psychometric properties of an instrument before it can be considered ready for use. Test-retest reliability of the instrument should be determined as well as internal consistency. There are also several types of validity which should be investigated such as content, construct and criterion-related validity.

Content validity is determined by the judgement of the person creating the instrument and other professionals. The development of the questions on the current scale from research efforts with adults and children could be considered as an indication of adequate content validity. The items were developed to measure those dimensions

described in research and were also reviewed by other professionals to consider the question of content validity.

The Baker and Velicer (1982) study addressed the factorial validity and reliability of four of the seven scales described in the first study. The other three scales have been based on research with adults or children supporting their content validity but the construct validity must be established for the present form. The first method of considering construct validity was the principal component analysis described in the first study which examined the factorial validity.

The second method of examining criterion related validity consisted of comparing the personality scale with an instrument designed to measure overt behaviors. This resulted in a comparison of personality and external behaviors and was thus a measure of construct validity. The instrument chosen was the Devereux Elementary School Behavior Rating Scale (DESB; Spivack & Swift, 1969). The DESB has been extensively researched and good psychometric properties have been established. It was chosen because it was specifically designed to identify behaviors which hamper learning in the school setting. Therefore, examination of the relationship between the behavioral and personality ratings gave an indication of the relationship of personality and academic progress.

The third method of examining construct validity was more complex. This method looked at the convergent and discriminant validity of the personality instrument when compared to a second personality instrument and method of gathering the data (Campbell & Fiske, 1967). The instruments used were the Student Personality Assessment Form (SPAF) developed and revised in the first study and the same scale revised to reflect the home situation and named the Parent Student Personality Assessment Form (ParSPAF). The ParSPAF was developed and used in this study since there were no other scales available which reflected dimensions similar to those of the SPAF.

Examination of the relationship of these instruments will determine how well the SPAF measures what it was designed for when compared to a second instrument measuring the same constructs.

Criterion-related validity was considered by comparing the instrument of concern with an external variable or instrument which was believed to be related to the trait in question. The major task of criterion related validity is to determine how an instrument relates to or predicts a criterion variable. This is in contrast to construct validity which examines how well an instrument supports its theoretical definition.

The method of examining criterion-related validity involved examination of the relationship between the personality instrument and various academic achievement criteria. This says very little about "what" the test measures but does give an indication of the practical value of the instrument. The academic criteria used were: (1) the Meeting Street School Screening Test (MSSST; Hainsworth & Sigueland, 1969) which is a cognitively oriented screening test, (2) grades in reading math and conduct given at the end of the year and (3) achievement test scores in reading and math as shown by the Comprehensive Test of Basic Skills (CTBS; McGraw-Hill, 1982). These measures were used since they reflect a child's functioning in different situations. The MSSST is a measure of how well a child does on cognitive tasks when interacting with a stranger on an individual basis. The grades reflect the teachers evaluation of a child and his academic progress over a long period. The CTBS reflects how a child functions in a group setting with little feedback as to how well he/she is doing and gives a measure of the child's academic skills at one point in time.

In summary then, the purpose of this project was the examination of the validity of a personality questionnaire modified from the Teacher Temperament Questionnaire (TTQ) (Thomas & Chess, 1977) and expanded to delineate other aspects of personality. The project was composed of four complementary studies. The first study examined the

factorial validity of an instrument deemed the Student Personality Assessment Form (SPAF). The second study examined the construct validity of the SPAF using the Devereux Elementary School Behavior Rating Scale (DESB; Spivack & Swift, 1969) as a criterion measure. The third study examined the construct validity of the SPAF via a multitrait-multimethod matrix design utilizing parent and teacher ratings (Campbell & Fiske, 1967). The fourth looked at the criterion-related validity of the personality scale with objective and subjective measures academic achievement.

Study One: Development of the
Student Personality Assessment Form
(SPAF)

The first study is a description of and rationale for the development of a teacher rating form to delineate academically relevant traits in kindergarten through second grade students. The scale, named the Student Personality Assessment Form (SPAF), was developed for the express purpose of delineating behavioral styles or traits and was designed to be quickly and easily completed by teachers or aides.

Method

Subjects

The subjects for this study were children in kindergarten, first or second grade in a centrally located Rhode Island school district. The sample was created by asking 25 teachers of the appropriate grade level to complete one questionnaire (appendix B) for each of ten randomly selected students. The random selection was accomplished by generating random numbers according to class size and selecting students by their placement on the

teacher's class list. The class lists were separated by sex, therefore the selection of subjects was alternated by sex. The resulting sample size was 120.

Instrument

The instrument developed and used in this study was an experimental version of the SPAF. It was composed of 90 items designed to delineate seven behavioral styles or traits: (1) Compliance, (2) Interpersonal Affect, (3) Extroversion, (4) Environmental Sensitivity, (5) Aggression, (6) Organization, and (7) Impulsivity. The first four traits were identified through prior research (Baker & Velicer, 1982). The latter three were created by altering items from other scales (Jackson, 1976; Comrey, 1970) or research (Digman, 1963; Kagan, 1963) to make them applicable to young children.

Baker and Velicer (1982) had tentatively matched three of the four components with those commonly found in research by Comrey (1970) and Jackson (1976). The Jackson and Comrey scales were chosen for comparison because they: (1) are recently developed, (2) exhibit good psychometric properties and (3) both claim to be a sample of the most widely identified personality traits.

The Jackson Personality Inventory (JPI; Jackson, 1976) measures sixteen traits, of which there appears to be a correspondence between "Compliance" and "Conformity" (JPI), "Extroversion" and "Social Participation" (JPI) and "Interpersonal Affect" and the scale by the same name on the JPI. The "Environmental Sensitivity" dimension does not appear to correspond to any of the JPI scales. The JPI "Organization" scale taps an area of concern to many teachers. This scale is characterized as a high score indicating a tendency to use time effectively, to complete work on schedule and to be able to avoid distractions. These skills appear to relate significantly to academic success. Therefore, an "Organization" scale was added to the SPAF.

The Comrey Personality Scale (CPS) also appears to have scales which correspond to the Baker and Velicer (1982) scales. "Social Conformity" (CPS) appears to correspond to some degree with the "Compliance" scale, "Extroversion" (CPS) seems to reflect similar behaviors to the "Extroversion" scale of Baker and Velicer, and the "Empathy Scale" (CPS) appears to correspond to the "Interpersonal Affect" scale. The "Orderliness" scale (CPS) also appears to be tapping an area which could effect academic functioning. A high score on this scale indicates a tendency to be concerned with neatness and to like living in a routine way versus being unsystematic and sloppy. This

scale appears to overlap somewhat with the "Organization" scale of the JPI and would also seem to be measuring a trait relevant to academic functioning. Therefore, some of the Orderliness items were altered and used on the "Organization" scale of the SPAF.

There have also been traits identified in research with children which are not similar to the Baker and Velicer (1982) traits, which could be useful in defining personality in children. Digman (1963), for example, using a teacher rating method, identified eleven factors using a sample of first and second graders. He utilized the list of trait statements from a study by Cattell and Coan (1957). The trait statements were not written in behavioral terms as were the statements in the TTQ which may tend to limit the predictive validity (See the review of the ESPQ in Buros, 1976). However, the research did delineate some personality traits which could be useful in research on child personality. The one factor which did not overlap with traits described earlier and would seem to have the most utility in the classroom is the Hostility factor. Items from the Hostility factor were altered to reflect behaviors rather than attitudes and the resulting scale was renamed "Aggression."

Another behavioral style which research indicates is relevant to school functioning is Impulsivity vs Reflectivity (Kagan, 1963; Mussen, Conger & Kagan, 1974). Impulsivity in children is characterized by a tendency to respond to a task without pausing to reflect on the quality of the answer. Children who respond quickly tend to make more errors than children who respond more slowly. The seventh scale, then, was constructed to reflect impulsivity.

Since the Compliance, Interpersonal Affect and Extroversion scales were shown to be adequately stable, they consisted of ten questions each. The questions were made up of those items which had the highest loading on the PCA completed by Baker and Velicer (1982). The Environmental Sensitivity scale was made up of items identified by Baker and Velicer as well as new questions and was 15 items in length. The Aggression, Organization and Impulsivity scales also consisted of fifteen questions. The questions are listed in Appendix B. These scales were chosen because of their potential relevance to academic functioning. Other scales were not chosen in the interest of brevity since an instrument will only be useful if teachers are willing to complete it.

The present project, then, was begun with the purpose of designing and analyzing a child personality scale which would reflect seven dimensions. The research by Baker and

Velicer (1982) defined four of the scales: (1) Compliance, (2) Extroversion, (3) Interpersonal Affect, and (4) Environmental Sensitivity. A fifth scale entitled Organization, was developed by altering items from research by Comrey (1970) and Jackson (1976). A sixth scale, Aggression, was developed by altering items from the Hostility trait defined by Digman (1963). The seventh scale was designed to reflect Impulsivity (Kagen, 1963; Mussen, Conger & Kagen, 1974).

Procedures

Each teacher was told that the research project was investigating the relationship between personality and school progress. A cover sheet (Appendix C) was attached to each questionnaire to provide instructions pertaining to completion. The questionnaires were given to the kindergarten, first and second grade teachers (n=25) of four of the six elementary schools in the district in April, 1981 and collected over the ensuing eight weeks as the teachers completed them. Participation in the project was voluntary and the teachers in two schools did not participate. Fourteen of the 25 teachers who volunteered completed 120 questionnaires. The number of questionnaires completed by each teacher ranged from five to ten. The median number of forms completed was ten. There were 25 questionnaires completed by three kindergarten teachers, 58 forms by six

first grade teachers and 37 by five 2nd grade teachers.

Results

A principal component analysis (PCA) was performed on the 90 by 90 matrix of interitem correlations. The Minimum Average Partial (MAP) method (Velicer, 1976) and Scree method (Cattell, 1966) were used to determine the number of components to retain. Recent research (Zwick & Velicer, 1982) has indicated that these are the two most accurate methods. The MAP procedure indicated the retention of nine components. The Scree method resulted in three natural breaks, indicating the retention of seven, six, or four components. In order to determine which of these solutions was the most appropriate, the Varimax rotated patterns for four, six, seven, and nine components were examined. Solutions for the larger number of components resulted in the retention of unique or trivial components, i.e., components identified by only a few items with low loadings or one or two items with high loadings. Therefore, the four component solution was selected for interpretation.

Table 1 presents the Varimax rotated component pattern when a .40 cutoff is used. Examination of the results permit an interpretation of the four dimensions being

Table 1

Varimax Rotated Component Patterns for the Ninety Item SPAF

Item	Component				Theoretical scale
	1	2	3	4	
1	-.75				Co
2				.70	Ex
3		-.59			IA
4					ES
5		.73			Ag
6	.68				Or
7	.66				Im
8	.78				Co
9			.48	-.56	Ex
10					IA
11			.50		ES
12		-.58			Ag
13					Or
14	-.64				Im
15	.54				Co
16	.42			.59	Ex
17		.73			IA
18			.63		ES
19	.48	.66			Ag
20			.58		Or

Table 1. Continued

Item	Component				Theoretical scale
	1	2	3	4	
21	.66	.49			Im
22	-.55				Co
23			.45		Ex
24		-.43	.55		IA
25			.47		ES
26	-.43	-.53			Ag
27	-.53				Or
28		-.42			Im
29	.74				Co
30			.44		Ex
31		.74			IA
32		-.57			ES
33		.59			Ag
34	-.61				Or
35	.59	.41			Im
36	-.55	-.44			Co
37				.41	Ex
38		.52			IA
39			.70		ES
40		-.54			Ag
41	.72				Or

Table 1. Continued

Item	Component				Theoretical scale
	1	2	3	4	
42	.79				Im
43					Co
44				-.49	Ex
45		.55			IA
46					ES
47		.74			Ag
48	-.76				Or
49	-.61	-.41			Im
50	-.73				Co
51	.42			.56	Ex
52		.67			IA
53			.66		ES
54		-.57			Ag
55					Or
56		.56		-.50	Im
57		-.66			Co
58					Ex
59		.79			IA
60					ES
61		.70			Ag
62	.67				Or

Table 1. Continued

Item	Component				Theoretical scale
	1	2	3	4	
63					Im
64			.52		Co
65				.66	Ex
66		-.48			IA
67					ES
68		-.60			Ag
69	.72				Or
70	.67				Im
71	.80				ES
72		-.64			Ag
73	.53				Or
74	.71				Im
75			.41		ES
76				.46	Ag
77	-.55				Or
78			.49		Im
79	-.48				ES
80		-.63			Ag
81	.65				Or
82	-.69				Im
83		.51			ES

Table 1. Continued

Item	Component				Theoretical scale
	1	2	3	4	
84				.51	Ag
85	.64				Or
86	-.57				Im
87					ES
88		.74			Ag
89					Or
90	-.57				Im

Note. Loadings $< .40$ are deleted.

Note. $\underline{n} = 120$

measured by the 90 item questionnaire. They are: (1) A dimension measuring a general orientation toward academic achievement characterized by Compliance, Organization and Impulsivity items, (2) an interpersonal skills dimension characterized by Interpersonal Affect and Aggression items, (3) a dimension which has not been clearly defined but is characterized by what is currently named Environmental Sensitivity items and (4) a dimension characterized by Extroversion items.

Table 2 presents the cross classification of items from the four component solution with the theoretical scales. The first component was defined mainly by Compliance (7 items), Impulsivity (11 items) and Organization (11 items). The second component was defined by Interpersonal Affect (8 items) and Aggression (13 items). The third component was not as clearly interpretable but was defined by Environmental Sensitivity (6 items). The fourth component was defined by Extroversion (7 items). The four component solution appears, from this analysis, to yield the most clearly defined as well as the most clearly interpretable components. It is evident, then, that the seven hypothesized scales were not reproduced with Varimax rotation.

Table 2

Cross Classification of Items Between Components and Content Defined
Scales of the Ninety Item SPAF

Theoretical scale	Component			
	1	2	3	4
Co	7	2	1	0
Ex	2	0	3	7
IA	0	8	1	0
ES	2	2	6	0
Im	11	5	1	1
Or	11	0	1	0
Ag	2	13	0	2
Total	35	30	13	10

However, Varimax rotation forces an orthogonal solution. A possible explanation for this failure to reproduce the seven hypothesized scales is that they are not independent. To test this hypothesis, the data were examined using Oblimin oblique rotation (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) with three different levels of correlation for seven, six and four component solutions. The levels of correlation were (1) nearly orthogonal ($\Delta = -4$), (2) a moderately correlated solution ($\Delta = -2$) and (3) a fairly correlated solution ($\Delta = 0.0$) (Nie, et al., 1975). Analysis of the Structure and Pattern matrices at each level of correlation for seven, six, and four component solutions showed less clearly defined and less meaningful patterns than was found with Varimax rotation. Examination of the four component solution with both the Varimax and Oblique rotations showed very similar results. Other rotations and component solutions did not improve upon these results. Therefore, the four component solution with Varimax rotation was chosen for further analysis since it was apparent that other rotations did not meaningfully alter that pattern and it was also apparent that the seven hypothesized scales could not be reproduced. Table 3 shows the five items from each component which had the highest loading.

Table 3

Items With Highest Loading Listed by Component of the Ninety Item SPAF

Item			
Loading	No.	Scale	Item
COMPONENT 1 - Compliance			
.79	42	Im	Child stays with a task only for a short length of time before switching to another.
-.78	8	Co	Child is easily distracted from work by noises or movements.
-.76	48	OR	Child finishes work before it is due.
-.75	1	Co	Child tends to return to interrupted activities.
.74	29	Co	Child starts but does not finish class work.
COMPONENT 2 - Interpersonal Affect			
.79	59	IA	Child becomes upset or angry when not allowed to engage in a desired activity.
.74	31	IA	Child becomes upset when another child wins a game.
.74	47	Ag	Child bullies other children.
.74	88	Ag	Child teases other children cruelly.
.73	17	IA	Child argues with other children.

Table 3. Continued

 COMPONENT 3 - Environmental Sensitivity

.70	39	ES	Child is very aware of changes in the level of light in the room (brighter or dimmer).
.66	53	ES	Child notices new or unusual clothes worn by others.
.63	18	ES	Child is aware of or comments about the class being hot or cold.
.58	20	OR	Child likes to follow the same routine daily.
-.57	54	Ag	Child is able to play without forcing others to do as he/she wishes.

COMPONENT 4 - Extroversion

.70	2	Ex	Child is shy with unfamiliar adults.
.66	65	Ex	Child is shy with children who are unfamiliar.
.59	16	Ex	Child prefers to watch new activities rather than participate.
-.56	9	Ex	Child tells stories of personal experiences with enthusiasm and excitement.
.56	51	Ex	Child adjusts slowly to new situations.

These results indicated that: (1) the Compliance and Interpersonal Affect dimensions delineated in the research by Baker and Velicer (1982) were expanded and further defined by items from the hypothetical scales of Organization, Impulsivity and Aggression, (2) the Extroversion Scale could be delineated by itself, and (3) the Environmental Sensitivity scale was still not clearly defined. Considering these results, the SPAF was revised and shortened to 43 items. The revised version of the SPAF was designed to measure four scales. The Compliance scale consisted of eight original Compliance items with four items from the Organization and Impulsivity scales added for a total of twelve items. The Interpersonal Affect scale consisted of eight items from the original scale with four items from the Aggression scale for a total of twelve. The Extroversion scale was made up of seven original Extroversion items and three Extroversion items from prior research for a total of ten. The fourth scale was a revised version of the Environmental Sensitivity scale.

Since items which had been added to the experimental (90 item) SPAF to define the Environmental Sensitivity scale had failed to do so, four new items were added to the SPAF in an attempt to determine what the fourth scale was measuring. Re-examination of the five items which loaded on the Environmental Sensitivity scale identified by Baker and Velicer (1982) suggested that verbal skills could be what

was being measured. Therefore, four new items, designed to measure verbal facility, were added. The fourth scale was then re-named "Verbal Facility."

The 43 item version of the SPAF then, consisted of items from the Compliance, Interpersonal Affect and Extroversion scales plus additional items from other scales added to each of these and a fourth scale using items from the Environmental Sensitivity scale plus new items and renamed Verbal Facility. Before this version was used, however, the construct validity was verified by principal component analysis (PCA).

Factorial validity of the 43 item SPAF

Subjects.

The subjects for this part of the study were kindergarten, first and second grade children in the same centrally located Rhode Island school district. The sample was selected by asking each teacher to complete a questionnaire for each of ten randomly chosen students. The students were chosen by generating random numbers according to class size and selecting students according to their placement on the class list. The class lists were separated by sex, therefore, the selection of subjects was alternated by sex. The kindergarten teachers were asked to complete ten questionnaires for students in their morning and afternoon classes.

Procedures.

Each teacher was told that the research project was investigating the relationship between personality and school progress. A cover sheet (Appendix E) was attached to each questionnaire to provide instructions for completion. The forms were given to the kindergarten , first and second

grade teachers (n=30) of five of the six elementary schools in the district in April, 1982 and collected over the following weeks until the end of the school year. Participation was voluntary and the teachers in one school did not participate. Seventeen of the 30 teachers who indicated interest completed 185 questionnaires. Two kindergarten teachers, who were interested in earlier research results, volunteered to complete questionnaires for all of the children in their classes. The number of forms completed by each teacher ranged from one to 46. The median number of forms completed was ten. There were 122 forms completed by six kindergarten teachers, 45 forms by six first grade teachers and 18 forms by three second grade teachers.

Results.

A principal component analysis (PCA) was performed on the 43 by 43 interitem correlation matrix. The PCA was computed to determine whether the scales which were identified with the 90 item SPAF would be reproduced and to see if the newly defined Verbal Facility scale could be delineated. The Minimum Average Partial (MAP) (Velicer, 1976) and Scree (Cattell, 1966) methods were used to determine the number of components to retain. The MAP method resulted in seven components being retained whereas

the Scree method indicated that four or five should be retained. The seven component solution resulted in the retention of unique or trivial components. This solution, therefore, was considered an over extraction. The four component solution eliminated a well defined component and was considered an underextraction.

The retention of five components, as indicated by the Scree method, resulted in a pattern similar to that found with the seven component solution but with a better component definition. Table 4 presents the Varimax rotated component pattern. Table 5 presents the Cross classification of items with the four theoretical scales. The first component consisted of eleven items from the Interpersonal Affect scale and one each from the Compliance and Verbal Facility scale. The second component consisted of nine items from the Extroversion scale and one each from the Interpersonal Affect and Verbal Facility scales. The third component consisted of four items from the Verbal Facility scale and one each from the Compliance and Interpersonal Affect scales. The fourth component also consisted of four Verbal Facility items, as did the third but had two items from the Extroversion scale and none from the Interpersonal Affect scale instead of one from each. The fifth component consisted of eleven Compliance items, three Interpersonal Affect, two Extroversion and one Verbal Facility item.

Table 4

Varimax Rotated Component Pattern for the Forty-Three Item SPAF

Item	Component					Theoretical scale
	1	2	3	4	5	
1					-.56	Co
2		-.70				Ex
3	-.52					IA
4	.75				.73	IA
5						Co
6	-.45					VF
7					.70	Co
8		-.60		-.48		Ex
9	.72					IA
10			.67			VF
11					-.63	Co
12				.64		VF
13			.76			VF
14					.83	Co
15		.45				Ex
16	.62					IA
17	.53				.62	Co
18		-.50				Ex
19	.49	-.48				IA
20			.83			VF

Table 4. Continued

Item	Component					Theoretical scale
	1	2	3	4	5	
21					.78	Co
22		.61				Ex
23	.63					IA
24	.85					IA
25					-.80	Co
26					-.85	Co
27		-.57			.54	Ex
28	.78					IA
29				.56		VF
30					.77	Co
31				.66		Ex
32	.71					IA
33	.81					IA
34		-.73				Ex
35			-.58			IA
36					.72	Co
37					.64	Co
38	.75					IA
39				.60		VF
40					-.72	VF
41		-.57				Ex

Table 4. Continued

Item	Component					Theoretical scale
	1	2	3	4	5	
42		.48		.48		VF
43		.51			-.59	Ex

Note. Items loading $< .45$ are deleted.

Note. $n = 185$

Table 5

Cross Classification of Items Between Components and Content Defined
Scales of the Forty-Three Item SPAF

Theoretical scale	Component				
	1	2	3	4	5
Compliance	1	0	1	0	11
Extroversion	0	9	1	2	2
Interpersonal Affect	11	1	0	0	3
Verbal Facility	1	1	4	4	1

Table 4 employed a .45 cutoff. Using .40 as a cutoff resulted in all items loading on at least one component. However, it also resulted in ten items loading on two components. Therefore, the results of using a more stringent cutoff was examined. Using .45 as the cutoff resulted in all items loading on at least one component and only six items loading on two. Using .50 as a cutoff resulted in only three items loading on two components but also resulted in three items not contributing to any component. This analysis indicated that .45 would result in the greatest ease of interpretation.

Examination of Table 4 shows that items 8, 17, 19, 27, 42, and 43 load on two components and may be measuring a complex concept. It also shows, however, that eleven of the twelve items designed to measure Interpersonal Affect load on Component 1; nine of the ten items designed to measure Extroversion load on Component 2; and eleven of the twelve items designed to measure Compliance load on Component 5. The only hypothesized scale which was not well delineated was Verbal Facility. However, three of the original five Environmental Sensitivity items (from Study one) load on component 3. A fourth item (# 6) had a .42 loading and could be considered as loading on Component 3. Component 4 was defined by three of the four items added to measure Verbal Facility.

Scale scores were developed based on the results of these two studies. Scale scores are the unweighted sum of all items allocated to that scale. Items with negative loadings are reflected before summation. Scale scores, unlike component scores, are not necessarily uncorrelated. However, this disadvantage is more than compensated for by the ease of computation, interpretability, and robustness of unit scoring across samples.

Table 6 shows the means, standard deviations, interscale correlations and Alpha coefficients for the Compliance, Interpersonal Affect and Extroversion scales. These scales included the following items:

Compliance: 1,5,9,13,17,21,25,29,33,37,40,42

Extroversion: 3,7,11,15,19,23,27,31,35,39

Interprsonal Affect: 2,6,10,14,18,22,26,30,34,38,41,43

These data are not shown for Verbal Facility since it was not reproduced.

The interscale correlations indicate that there is a moderate correlation between the Compliance and Extroversion scales, a moderate to high correlation between the Compliance and Interpersonal Affect scales but little relationship between the Extroversion and Interpersonal Affect scales.

Table 6

Means, Standard Deviations, Alpha Coefficients and Interscale
Correlations of the SPAF Scales

Theoretical scale	Mean	Standard deviation	Alpha	Correlation	
				Co	Ex
Co	32.5	11.15	.92		
Ex	26.5	7.71	.83	.38	
IA	33.0	8.83	.86	.66	-.03

Note. n = 185

The Compliance score can range between 12 and 60, with a midpoint of 36. The obtained mean was just below this with a high score indicating compliance. The Extroversion score can range between ten and 50 with a midpoint of 30. The obtained mean for the Extroversion scale was 26.5 with a high score indicating extroversion. The Interpersonal Affect scores can range between 12 and 60 with a midpoint of 36. The obtained mean was 33 with a high score indicating well developed interpersonal adjustment.

The Alpha coefficients, which are a measure of internal consistency, ranged from .83 to .92 with the Compliance scale being highest and Extroversion the lowest. This indicates that the scales have adequate internal consistency.

Discussion.

The PCA of the revised version of the SPAF resulted in three clearly identifiable components which demonstrates that the Compliance, Interpersonal Affect and Extroversion scales can be reproduced across different samples of both subjects and items. The scales based on these three components also demonstrated good internal consistency. The results also show, however, that the Environmental Sensitivity or Verbal Facility scale(s) has still not been

clearly defined. Yet, the items which contributed to the original scale continue to be delineated as a component. It can also be seen that the items added to further define Verbal Facility suggested a separate component. However, both need further study and development before they can be used in validity studies.

These results demonstrated adequate factorial validity for the three scales. Subsequent research on construct or criterion-related validity used a 34 item version of the SPAF which measured Compliance, Extroversion, and Interpersonal Affect.

Study Two: Relationship of the SPAF and the Devereux Elementary School Behavior Rating Scale (DESB)

Study Two examined construct validity of the SPAF by examining the relationship of the three SPAF dimensions as described above and the Devereux Elementary School Behavior Rating Scale (DESB) (Spivack & Swift, 1966). The DESB is an extensively researched and potentially useful instrument. It was not developed as a measure of personality but provides a profile of overt behaviors which have been judged to be related to academic achievement. It was chosen for that reason as the instrument to use in examining the relationship between the behaviors related to academic achievement and the personality constructs defined by the SPAF. By relating the personality constructs to the behaviors defined by the DESB, it may be possible to identify students who may begin to exhibit maladaptive behaviors prior to the onset of those behaviors.

Method

Subjects

The subjects for this study were kindergarten, first and second grade children in a centrally located Rhode Island school district. The sample was selected by asking

each teacher of the appropriate grade level to complete two questionnaires for each of ten randomly chosen students. The students were chosen by generating random numbers according to class size and selecting students according to their placement on the class list. The class lists were separated by sex, therefore, the selection of subjects was alternated by sex. Thirteen teachers from four schools completed both questionnaires for 104 children.

Instruments

Study Two utilized two instruments. One was the 34 item SPAF developed in Study One. The SPAF was constructed to reflect three scales: (1) Compliance, (2) Interpersonal Affect and (3) Extroversion. The second instrument was the Devereux Elementary School Behavior Rating Scale (DESB; Spivack and Swift, 1966). The DESB is made up of 47 items designed to measure eleven dimensions of school behavior. The eleven dimensions are: (1) Classroom Disturbance: the extent to which a child's behavior is active, inappropriately social and disruptive, (2) Impatience: the extent to which a child begins work before directions are completed and fails to consider the results of certain behaviors, (3) Disrespect-Defiance: the extent to which a child shows open disrespect for or resistance to school, (4) External Blame: the extent to which a child blames outside sources for his problems, (5) Achievement Anxiety: the

extent to which a student outwardly exhibits worry concerning the inability to meet the demands of the teacher or school situation, (6) External Reliance: the extent to which a child depends on others to make decisions or take action, (7) Comprehension: the extent to which a child is understanding the day to day work, (8) Inattentive-Withdrawn: the extent to which a child is able to maintain contact with and attend to classroom activities, (9) Irrelevant-Responsiveness: the extent to which a child's responses are irrelevant or inappropriate, (10) Creative- Initiative: the extent to which a child is motivated and personally involved and (11) Need for Closeness to the Teacher: the extent to which the child values the teacher as a person. There are also three items on the DESB which are not part of a common factor. The items were retained because they are related to the same four factors: Impatience, External Blame, Comprehension and Inattentive-Withdrawn. High scores on these items indicate that the child is being taught far beyond his/her instructional level. These three items will not be used in the analysis of the relationship between the DESB and SPAF.

Normative data for the DESB was obtained by having thirty two kindergarten through sixth grade teachers rate 809 children in thirteen elementary schools. Test-retest reliability was determined on ratings of 128 children approximately one week after completion of the initial

ratings. Correlation coefficients for these ratings ranged from .85 to .91 with the median coefficient being .87. Validity of the scale has been determined through the use of factor analytic techniques, teacher conferences to select the behaviors to be rated and examination of the relationships between the items or behavior factors and school achievement. In all, 1719 ratings were completed on 1546 children by 147 teachers.

Procedures

Each teacher was told that the project was investigating the relationship between personality and behavior. Instructions for completion were included on the front of each questionnaire (appendix E). The questionnaires were given to the kindergarten, first and second grade teachers of five of the six elementary schools in April, 1982 and collected as they were completed. Participation was voluntary and the teachers in one school did not participate. The SPAF forms completed were those used in the 43 item factorial validity study described in Study One. There were 185 SPAF forms completed by 17 teachers. The median number of forms completed by each teacher was ten. There were 104 DESB forms completed by 13 teachers. The number of forms completed by each teacher ranged from three to ten. The median number was ten. There were 46 forms completed by 4 kindergarten teachers, 43 by

seven first grade teachers and 15 by two second grade teachers.

Results

The data generated in this study were evaluated using univariate and multivariate correlation analyses. The zero order correlations indicated the relationship between each of the SPAF scales and each DESB factor. The multiple regression analyses looked at the relationship of each SPAF scale with all of the DESB factors thus examining DESB redundancy. A canonical variate analysis was also completed to examine the relationship of the three SPAF dimensions to the eleven DESB factors to consider the redundancy of both scales.

Zero Order Correlation

Analysis of the zero order correlations (Table 7), indicates that the Compliance scale has a high negative correlation with the DESB factors of Classroom Disturbance, Impatience, Inattentive-Withdrawn and External Reliance; a high positive correlation with Comprehension; low to moderate negative correlations with Irrelevant Responsiveness and External Blame; and a low to moderate positive correlation with Creative Initiative.

Table 7

Pairwise Correlation Coefficients of SPAF Scales and DESB Factors

DESB factors	SPAF scales		
	Co	EX	IA
Classroom Disturbance	-.66*	.05	-.80*
Impatience	-.60*	-.23	-.42*
Disrespect Defiance	-.17	-.11	-.23
External Blame	-.29*	-.12	-.46*
Achievement Anxiety	.05	-.19	.00
External Reliance	-.50*	-.36*	-.22
Comprehension	.58*	.38*	.25*
Inattentive-Withdrawn	-.78*	-.36*	-.51*
Irrelevant Responsiveness	-.31*	-.16	-.27*
Creative Initiative	.31*	.30*	.06
Teacher Closeness	.18	.11	.15

*P < .05

The Extroversion scale has a moderate and negative correlation with External Reliance and Inattentive-Withdrawn; and a low to moderate positive correlation with Creative Initiative and Comprehension.

The Interpersonal Affect scale has a high negative correlation with Classroom Disturbance; a moderate negative correlation with Impatience, External Blame and Inattentive-Withdrawn; a statistically significant but low negative correlation with Irrelevant Responsiveness and low positive correlation with Comprehension.

The results indicated that a student with a low score on the Compliance scale, which indicates non-compliance, is characterized by the following behaviors: (1) active, inappropriately social and disruptive behavior (Classroom Disturbance); (2) impulsivity and sloppiness in work habits (Impatience); (3) inability to maintain a focus on what is happening in class (Inattentive-withdrawn); (4) reliance on others to make decisions and provide direction (External Reliance); and (5) inability to comprehend the work in the classroom (Comprehension). It is also characterized to a lesser degree by: (1) attributing success or failure to external control (External Blame); (2) responding in an inappropriate, irrelevant manner (Irrelevant Responsiveness); and (3) not contributing constructively with behavior or ideas in the class (Creative Initiative).

Extroversion was not strongly characterized by any of the DESB factors. However, a student with a low score, which indicates introversion, was defined to a moderate degree by the following behaviors: (1) reliance on others to make decisions and provide direction (External Reliance); (2) inability to maintain a focus on classroom functions (Inattentive-Withdrawn); (3) concern with not being able to meet the requirements of the Teacher (Achievement Anxiety); (4) inability to comprehend the work in the classroom (Comprehension); and (5) not contributing constructively with behavior or ideas in the class (Creative Initiative).

A student with a low score on the Interpersonal Affect scale was characterized to a high degree, by active, inappropriately social and disruptive behavior (Classroom Disturbance). They were also characterized moderately by the following behaviors: (1) inability to maintain a focus on the workings of the class (Inattentive-Withdrawn); (2) impulsivity and sloppiness about class work (Impatience); and (3) attributing success or failure to external controls (External Blame). In addition, they were also characterized to a slight degree by: (1) inappropriate, irrelevant responding (Irrelevant Responsiveness); (2) inability to comprehend the demands of work in the class (Comprehension; and (3) reliance on others for decisions and directions (External Reliance).

The zero order correlations presented show that some of the specific behavioral factors, as measured by the DESB, are related to the personality variables measured by the SPAF. A multiple regression analysis (MRA) was employed to investigate the complex multivariate relationship as defined by combinations of the behaviors.

Multiple Regression Analysis

The MRA studied the relationship between each of the SPAF scales as criterion variables and all eleven of the DESB factors as predictors. This serves to indicate the level of redundancy. It is possible, for example, that the portion of the Classroom Disturbance factor which correlates with Compliance is redundant with that portion of the Impatience factor that correlates with Compliance. If this were the situation, the resulting multiple regression coefficients would be relatively low or at least very similar to the zero order correlation of the best predictor. If the scales are not redundant, they should supplement each other, resulting in a high multiple correlation coefficient.

Table 8 shows the results of multiple regression analysis (MRA) using the DESB factors as predictor variables and each of the personality scales of the SPAF as a criterion variable. Stepwise multiple regression was employed. The predictor which is entered first with the

Table 8

Multiple Regression Coefficients and Order of Inclusion of Variables
into the Stepwise Regression

	SPAF		
	Co	Ex	IA
Order	In-Wd	Compre	ClsDis
of	Compre	In-Wd	ExtBlm
Inclusion	Cls Dis	Cls Dis ^a	Tch Clo ^a
of	Tch Clo	Tch Clo ^a	Dis Def
DESB	Impat	Ach Anx ^a	In-Wd
Factors	Irr Res	Dis Dep ^a	Compre
	Ach Anx	Cre Ini	Ext Rel
	Cre Ini	Ext Rel	Cre Ini
	Ext Blm	Impat	Irr Res
	Ext Rel	Irr Res	
	Dis Def	Ext Blm	
Multiple R	.88	.70	.86
Multiple R ²	.78	.49	.74
Adjusted R ²	.75	.43	.72

^a Included as a suppressor variable.

Compliance scale is Inattentive-Withdrawn, followed by Comprehension, Classroom Disturbance, Teacher Closeness and Impatience. These five scales predict 77 percent of the variance. The remaining six scales account for only one percent of the variance.

The only variable which is surprising to find contributing significantly to the prediction equation is the Teacher Closeness scale. The other four were shown to correlate individually to a moderate to high degree with Compliance. However, Teacher Closeness had a low positive correlation (.18). What this suggests is that Teacher Closeness may be serving as a suppressor variable. Defining a suppressor variable as a situation in which "the Pth predictor is a suppressor variable if and only if $|R_{0.123}| > |R_{0.12} + r_{03}|$ " (Velicer, 1978, pg 956) does not support the contention that Teacher Closeness is a suppressor variable. Therefore, it must be accounting for a sufficient amount of variance not redundant with the other variables to be included as a direct predictor.

The Extroversion scale is best predicted by the Comprehension factor of the DESB followed by the Inattentive-Withdrawn, Classroom Disturbance, Teacher Closeness, Achievement Anxiety, Disrespect-Defiance and Creative Initiative factors. The Classroom Disturbance, Teacher Closeness, Achievement Anxiety and

Disrespect-Defiance factors are functioning as suppressor variables. Velicer's (1978) definition of a suppressor variable given above works equally well whether applied to the predictor or suppressor variable and identifies when a suppressor variable is present but not how it is functioning. In other words, it is unclear whether Teacher Closeness suppresses some of the extraneous variance of the Classroom Disturbance scale or whether it suppresses the extraneous variance of the Inattentive-Withdrawn scale. What is indicated is that the Extroversion scale is defined to a large degree by the Comprehension and Inattentive-Withdrawn scales with the next four scales reducing the irrelevant variance followed by the Creative Initiative scale which accounts for a small amount of the variance. It should be noted, that the addition of some of the suppressor variables increases the variance predicted by a moderate to high degree and the addition of others increases it minimally. The addition of the Classroom Disturbance scale increases R Square by .13. The addition of the Achievement Anxiety scale increases it by .11. However, the level of prediction is increased only slightly with the addition of Teacher Closeness (.03) and Disrespect-Defiance (.02).

The Interpersonal Affect scale is more simply defined than the Extroversion scale. It is highly correlated with the Classroom Disturbance factor of the DESB. Therefore,

the variance accounted for is increased only slightly with the addition of the External Blame, Teacher Closeness and Disrespect-Defiance factors. Here, as with the Extroversion scale, Teacher Closeness is functioning as a suppressor variable. In this case, Teacher Closeness has a low correlation with Extroversion and a high correlation with External Blame. This indicates that Teacher Closeness is reducing the extraneous variance of External Blame. It is also apparent that there is a high degree of redundancy between the DESB variables, in that, many of the scales which showed a significant zero order correlation did not contribute to the multiple regression equation.

Examination of the results of the MRA has shown that the DESB factors share much of the variance accounted for in the SPAF. However, the analysis says little about the overlap of the SPAF scales in their relationship to the DESB. The question to be considered is whether the dimensions of the SPAF can be identified separately in their relationship to the DESB factors. For example, does Compliance show an orthogonal linear relationship to the DESB factors or do the Interpersonal Affect and Compliance dimensions overlap to such a degree that there is a combined linear relationship? The question of redundancy will be examined in the next section by computation and analysis of a Canonical Correlation. The hypothesis, in this case, would be that the SPAF and DESB would have three dimensions

in common. Any fewer dimensions would indicate that the SPAF scales were overlapping and redundant in their relationship with the DESB.

Canonical Variate Analysis

The Bartlett's tests (App J) indicate that all three canonical variates identified with the SPAF are significant ($P < .01$). There is controversy concerning the interpretation of the dimensions of the CVA. Darlington, Weinberg and Walberg (1975) suggest analyzing the canonical variable loadings instead of the standardized coefficients if there is a large divergence between the two with a moderate sample size. There was considerable divergence between the two with these results. Therefore, the canonical variable loadings were analyzed. Table 9 shows the pattern resulting from the CVA.

The first canonical variate is defined by a negative loading on the Compliance and Interpersonal Affect scales. This linear composite is defined by the scales of Classroom Disturbance, Impatience, Disrespect Defiance, External Blame, External Reliance, lack of Comprehension, Inattentive-Withdrawn, Irrelevant Response and lack of Creative Initiative.

Table 9

Canonical Variable Loadings

SPAF scales	Canonical variables		
	1	2	3
Compliance	-.986	.161	.051
Extroversion	.313	.669	-.674
Interpersonal Affect	-.798	-.544	-.260
DESB			
DESB scales			
Classroom Disturbance	.882	.456	-.029
Impatience	.812	-.016	-.226
Disrespect Defiance	.602	.539	.144
External Blame	.534	.367	.253
Achievement Anxiety	.040	.012	.612
External Reliance	.728	-.425	.009
Comprehension	-.700	.467	-.160
Inattentive Withdrawn	.908	-.251	-.032
Irrelevant Response	.850	.151	.047
Creative Initiative	-.367	.497	-.371
Teacher Closeness	-.031	-.065	-.069

Note. n = 104

The second canonical variate is defined positively by the Extroversion scale and negatively by the Interpersonal Affect scale for the first data set and by the Classroom Disturbance, Disrespect Defiance, and External Blame scales and a negative relationship to the External Reliance, Comprehension, and Creative Initiative scales.

The third canonical variate was defined by a negative loading on the Extroversion scale of the first data set and by the DESB scales of Achievement Anxiety and a lack of Creative Initiative in the second data set.

Discussion

The univariate correlation analysis indicated that the Compliance scale was significantly correlated with eight of the eleven DESB scales. This lends support to the definition of the Compliance scale, discussed in Study One, as reflecting the extent to which a child shows a general orientation toward academic achievement. Since a low score reflects non-compliance, the specific scales which correlate also lend support to the definition. It should also be noted that the DESB scales which correlate with the Compliance scale are discussed by Spivack & Swift (1967) as reflecting two patterns often found with children showing poor school achievement. The first pattern involves high

External Reliance, high Inattentive-Withdrawn and low Comprehension scores and are associated with the child who needs much external direction and help. The second pattern involves high scores on the Classroom Disturbance, Impatience, External Blame and Irrelevant Responsiveness scales. This pattern is associated with the child who is poorly controlled and tends to act out in class. The Compliance scale correlates in the expected direction with these scales and also positively with the Creative Initiative scale. This suggests that a low score on the Compliance scale reflects a personality style which results in a child who needs much external control and help and who tends to act out. Conversely, a high score reflects a personality style which is generally achievement oriented.

The univariate analysis also supported the definition of the Extroversion scale. It showed a moderate negative correlation with the Inattentive-Withdrawn and External Reliance scales and positive correlation with the Comprehension and Creative Initiative scales. Thus, a low Extroversion score (indicating introversion) would reflect a child who withdraws from the classroom activities and is unable to make independent decisions. The moderate correlation suggests that there are other behaviors of equal importance evidenced in the personality dimension which are not measured by the DESB.

The definition of the Interpersonal Affect scale was also supported by the univariate analysis. It had a high negative correlation with the Classroom Disturbance scale and was moderately related to other scales such as External Blame and Impatience, which reflect interpersonal skills and attitudes.

Multiple Regression Analysis (MRA) indicated that there is considerable overlap among the DESB scales. This is demonstrated by the fact that several of the scales which had a significant zero order correlation enhanced the multiple correlation only slightly. There were also several scales which functioned as suppressor variables.

The results of the MRA indicated that the Compliance and Extroversion scales are both defined by their relationship to the Inattentive-Withdrawn and Comprehension scales of the DESB. However, it seems that the SPAF scales are being defined by different aspects of the DESB factors since Compliance was further defined by the Classroom Disturbance and Teacher Closeness scales whereas the same scales served as suppressor variables with the Extroversion scale. The Interpersonal Affect scale is defined to a high degree by the Classroom Disturbance and External Blame scales.

These results, then, support the validity of the SPAF in that: (1) a low score on the Compliance scale (indicating non-compliance) reflects inattentive-withdrawn, poorly controlled behavior with a lack of comprehension and no feeling of closeness to the teacher, (2) a low score on the Extroversion scale (indicating introversion) reflects inattentive-withdrawn behavior and a concomitant lack of comprehension without the poorly controlled, acting out behaviors and (3) a low score on the Interpersonal Affect scale (indicating a lack of interpersonal skills) reflects poorly controlled acting out behaviors in conjunction with blaming others for problems and disrespectful defiant behavior with no feeling of closeness to the teacher. These results, then, support the theoretical definitions of the SPAF scales.

The results of the Canonical Variate Analysis support the theoretical definition of the SPAF as consisting of three distinct scales. The CVA did not reproduce the three clearly identifiable scales produced with PCA and Varimax rotation. The relationship between the SPAF and DESB is also somewhat different than that noted with MRA. This is to be expected, however, since the purpose of CVA, MRA and PCA are different. With CVA, a linear composite in the first set of data is defined such that its correlation with a linear composite in the second set of data is maximized. With MRA, a linear composite is defined with respect to only

one external variable. With PCA a linear composite is only defined internally to maximize variance. The CVA results indicate that the three SPAF scales collectively are accounting for different aspects of the DESB and lend strong support to the validity of the SPAF. This is shown by the nature of the dimensions produced. The first canonical variate consists of negative loadings on the Compliance and Interpersonal Affect scales. This linear composite describes a non-compliant personality style with poorly developed interpersonal affect. This dimension is defined by all of the DESB scales except Achievement Anxiety and Need for Teacher Closeness. It is also negatively correlated with Comprehension and Creative Initiative. This seems to be a description of a general behavior problem dimension with a concomitant lack of academic understanding and initiative.

The second canonical variate consists of Extroversion and a lack of interpersonal skills (a negative relationship to Interpersonal Affect). This dimension is defined by a positive correlation with the scales of Classroom Disturbance, Disrespect-Defiance and External Elance but also Comprehension and Creative Initiative. It is negatively correlated with the External Reliance factor. This linear composite, then, is described by an extroverted personality style with poorly developed interpersonal affect which is behaviorally defined by control problems but not by an

inability to understand the material presented in the classroom.

The third canonical variate is defined by the Extroversion scale (indicating introverted tendencies) and as such is defined by Achievement Anxiety and a lack of Creative Initiative. This seems to reflect an introverted personality style depicted by fear of not being able to meet teacher expectations and an inability to contribute to classroom activities.

The correlation analysis, then, has demonstrated criterion-related validity with both the univariate and multivariate methods. This is, of course, only one method of considering validity. These results must also be examined with the possibility of a "halo effect" in mind. Since the teachers completed the SPAF and DESB during the same time frame, it is possible that the rating on one form influenced the rating on the other. This possibility can be considered by examining other aspects of validity. One technique, which would consider whether a "halo effect" was effecting the results would be to have two sets of people complete different forms, which measure similar constructs. In this way, convergence and divergence could be considered. This is accomplished in the next study.

Study Three: Relationship of the SPAF

Completed by Teachers and Parents

This study examined another aspect of the construct validity of the SPAF. Study One, described earlier, was designed to determine the factorial validity of the instrument. Study Two looked at the relationship of the SPAF and certain behaviors as measured by the DESE. Study Three will examine the convergent and discriminant validity (Campbell & Fiske, 1967) of the components described in Study One. This was accomplished by examining the relationship between the SPAF when completed by teachers compared to the same instrument completed by parents. Thus the same dimensions were assessed in a different environment by a different rater. The teacher report is considered as one method. Parent report is considered as a second. The teacher sees the child in the academic setting where he/she is interacting with other children and responding to various academic pressures. The parent sees the child in the home, interacting with siblings or a limited number of peers in a setting with few demands to perform academically. If the variance accounted for by the SPAF is trait variance there will be a high degree of correlation between the dimensions when completed by the two methods, thus demonstrating convergent validity. If the variance accounted for is

method variance, then, a low correlation would result.

For example, a low correlation could result if contamination was influencing the rating by the teacher. If the teacher has already made a decision about a child's personality, for example, based on reports from previous teachers and rates the child according to that construction, then, the parent and teacher would be viewing the child differently and the low correlation would be due to method variance.

A second situation which could result in a low convergent correlation would be if the parents and teachers view different aspects of the same personality traits as the salient features of that trait. For example, the salient features of Interpersonal Affect could conceivably be exhibited differently at home and school. In that case, a low correlation could be reflecting sampling inadequacies of the rating form.

A third situation which could result in a low level of convergence would be if the limited number of children for the parents to compare their child with were to reduce the level of variance in the parent completed forms relative to the teacher completed forms. Alternatively, the parents internal anchors, i.e. what is "average" could be distorted because of the limited sample of behaviors available. In

either case, a low level of convergence would result.

An equally important aspect of validity is the demonstration of a lack of relationship between the variable and other theoretically unrelated variables. The discriminant validity of the SPAF will be evaluated by examining the relationship between each dimension of the SPAF and the other theoretically unrelated dimensions when the instrument is completed by parents as compared to teachers. There are three types of discriminant validity which can be examined. The first type of discriminant validity requires that the correlation between the same trait as measured by different methods should be higher than different traits measured by different methods. In other words, the correlation between the trait of "Extroversion" as measured by two methods should be higher than the correlation of the traits of "Extroversion" and "Compliance" as measured by different methods.

The second type of discriminant validity is more stringent than the first. It requires that the correlation between the same trait as measured by different methods should be higher than different traits measured by the same method. For example, the correlation between "Extroversion" as measured by parents and teachers rating should be greater than the correlation between "Extroversion" and "Compliance" measured by the same method.

The third requirement is that patterns of relationships among the traits be comparable across methods. If "Extroversion" is correlated moderately with "Interpersonal Affect" and has a slight negative correlation with "Compliance" when teacher ratings are examined, then the same pattern would be expected when parent ratings are examined. The magnitude of the correlation coefficients would not have to be equal but the pattern should be similar.

Method

Subjects

The parents and teachers of children in kindergarten, first or second grade participated in this study. The schools involved were in a centrally located Rhode Island town. The teacher completed SPAFs were the same as those used for the analysis of the 43 item SPAF described in Study One. There were 185 questionnaires completed by 17 teachers from four schools. However, teacher questionnaires were used only if there was a matching parent form. The final sample, for which there were matching parent questionnaires, consisted of 104 children.

Instruments

There were two instruments used in this study. The SPAF, as described in study One and Two was used as one instrument. It is a 34 item questionnaire which measures Compliance, Extroversion and Interpersonal Affect using the original items designed to delineate these scales plus additional items as discussed in Study One

The second instrument, called the Parent Student Personality Assessment Form (ParSPAF) is a version of the SPAF modified to reflect the home rather than the school setting. Changes were made when an item referred to class seating arrangements, or the teacher and were kept to a minimum. The ParSPAF was completed as a 43 item questionnaire. All 43 items were included in the PCA as they were with the SPAF described in Study One. However, only 34 items were used in subsequent analysis.

Procedure

As indicated in Study Two, each teacher was asked to complete a SPAF for each of ten randomly chosen students. Student selection was accomplished by generating random numbers according to class size and selecting students by their placement on the class list. The class lists were separated by sex, therefore, subject selection was

alternated by sex.

As the teachers completed the SPAF, a ParSPAF was mailed to the parents of that child, along with a self addressed, stamped envelope. A cover letter (Appendix G) was attached explaining why and by whom the research was being done. The instructions for completion (Appendix H) were also attached. A space was provided for the parents to indicate whether they would like information concerning the results of the study sent to them. A follow up letter (Appendix I) and another ParSPAF were mailed to the parents who had not responded at the end of the fourth week. The data analysis involved two steps. First, a principal component analysis was performed to determine if the structure of the ParSPAF was comparable to the structure of the SPAF. Second, the data generated were used to examine the convergent and discriminant validity of the SPAF by a multitrait-multimethod matrix (Campbell & Fiske, 1967).

Results

Factorial Validity of the ParSPAF

Before the convergent and discriminant validity of the SPAF was considered, the factorial validity of the ParSPAF needed to be determined. This was accomplished by computing a principal component analysis with Varimax rotation on the 43 by 43 interitem correlation matrix. The minimum average partial (MAP) (Velicer, 1976) and Scree (Cattell, 1966) methods were used to determine the number of components to retain. The MAP method suggested the retention of five components. The Scree method indicated that four or five components should be retained.

The four component solution seemed to be an under extraction. The resulting pattern was very similar to the five component solution. However, there were five items which did not load on any component and four items which loaded on two components. This was compared to four items not loading and three items loading on two components when five were retained. In addition, the variance accounted for was reduced from 44.7 to 39.6 percent. Therefore, the five component solution was considered the better one.

Examination of Table 10 shows that the item pattern of the five component solution for the ParSPAF very closely resembles the SPAF item pattern. Table 11 shows the cross classification of items for the five component solution resulting from the PCA of the ParSPAF. With a .40 cutoff, the first component consisted of ten Compliance items, and one each from the Interpersonal Affect and Verbal Facility scales. The second was made up of seven items from the Extroversion scale and two from the Verbal Facility scale. The third had two items from the Verbal Facility scale and one each from the Compliance, Extroversion and Interpersonal Affect scales. The fourth consisted of ten items from the Interpersonal Affect scale. The fifth had three items from the Verbal Facility scale and one each from the Compliance, Extroversion and Interpersonal Affect scales.

These results indicate support for the factorial validity of the SPAF whether completed by parents or teachers. This is only true for the hypothetical constructs of Compliance, Extroversion and Interpersonal Affect. It was apparent, once again, that Verbal Facility was not being delineated as hypothesized. Therefore, subsequent analysis was completed using only the three most stable scales. Since this analysis supported the factorial validity of the ParSPAF, the convergent and discriminant validity (Campbell & Fiske, 1967) was examined using the data from the teacher completed SPAF as one method and the parent completed SPAF

Table 10

Varimax Rotated Component Pattern for Forty-Three Item ParSPAF

Item	Component					Theoretical scale
	1	2	3	4	5	
1	-.45					Co
2		.67				Ex
3					.55	IA
4				.51		IA
5	.69					Co
6						VF
7	.59					Co
8		.66				Ex
9				.70		IA
10					.42	VF
11	-.57					Co
12						VF
13						IA
14	.65					Co
15		-.65				Ex
16				.55		IA
17	.51					Co
18					-.60	Ex
19				.53		IA
20					.42	VF

Table 10. Continued

Item	Component					Theoretical scale
	1	2	3	4	5	
21	.75					Co
22		-.68				Ex
23				.52		IA
24				.41		IA
25					.60	Co
26	-.49					Co
27		.56				Ex
28				.73		IA
29					.50	VF
30	.85					Co
31		-.58				Ex
32	.49			.53		IA
33				.44		IA
34		.74				Ex
35				-.42		IA
36	.69					Co
37			-.86			Co
38			-.87			IA
39	-.41		.52			VF
40		.44	.40			VF
41			.63			Ex

Table 10. Continued

Item	Component					Theoretical scale
	1	2	3	4	5	
42		-.56				VF
43						Ex

Note. Items loading $< .40$ are deleted.

Note. $n = 104$

Table 11

Cross Classification of Items Between Components and Content DefinedScales of the ParSPAF

Theoretical scale	Component				
	1	2	3	4	5
Compliance	10	0	1	0	1
Extroversion	0	7	1	0	1
Interpersonal Affect	1	0	1	10	1
Verbal Facility	1	2	2	0	3

as the second.

The results of the analysis were judged to adequately replicate the results with the SPAF. As a result, the same scoring system was employed to form scale scores for the three scales. Table 12 shows the means, standard deviations, interscale correlations and Alpha coefficients for the Compliance, Extroversion and Interpersonal Affect scales.

The interscale correlations indicate a moderate correlation between Compliance and Interpersonal Affect and a low correlation between Extroversion and both Compliance and Interpersonal Affect. The scale scores range from 12 to 60 for Compliance and Interpersonal Affect and ten to 50 for Extroversion. The respective mid-point scores are 36 for Compliance and Interpersonal Affect and 25 for Extroversion. The obtained means of Compliance and Interpersonal Affect are slightly below the middle score. The mean of the Extroversion scale was slightly above the middle score. The Alpha coefficients for the scales ranged from .84 to .75 indicating adequate internal consistency.

Table 12

Means, Standard Deviations, Alpha Coefficients and Interscale
Correlations of the ParSPAF Scales

Theoretical scale	Mean	Standard deviation	Alpha	Correlation	
				Co	Ex
Co	28.9	6.74	.84		
Ex	26.3	6.12	.80	.14	
IA	28.0	5.45	.75	.49	.25

Note. n = 104

Analysis of Convergent and Discriminant Validity.

Table 13 is a multitrait-multimethod matrix showing the correlation of the SPAF and ParSpaf on the same traits (heteromethod-monotrait), on different traits and different methods (heterotrait-heteromethod), and on different traits with the same method (heterotrait-monomethod). The monotrait-heteromethod coefficients demonstrate the extent of convergent validity and are indicated by superscript c in Table 13. The discriminant validity is determined by the heterotrait-heteromethod (indicated by superscript b) and heterotrait-monomethod (indicated by superscript a) correlation coefficients.

Convergent validity.

Examination of the SPAF/ParSPAF interscale correlation coefficients, which are noted by superscript c on Table 13, reveals that the Compliance and Extroversion scales demonstrate a moderate level of convergent validity and that the Interpersonal Affect scale has a low-moderate level of convergent validity. This indicates that parents and teachers rate the children in a somewhat similar manner on all three scales.

Table 13

Inter- and Intrascale Correlation Coefficients for the SPAF and
ParSPAF

		SPAF			ParSPAF	
		Comp	Ext	IA	Comp	Ext
SPAF	COMP					
	EXT	.45 ^a				
	IA	.65 ^a	.09 ^a			
	Comp	.44 ^c	.15 ^b	.18 ^b		
ParSPAF	Ext	.01 ^b	.34 ^c	.04 ^b	.14 ^a	
	IA	.31 ^b	.11 ^b	.23 ^c	.49 ^a	.25 ^a

Note. n = 104

- ^a Heterotrait - monomethod coefficients
- ^b Heterotrait - heteromethod coefficients
- ^c Monotrait - heteromethod coefficients

Discriminant validity.

The first discriminant validity question to be considered is whether the monotrait-heteromethod correlation (superscript c) is greater than the heterotrait-heteromethod correlation (superscript b). The monotrait-heteromethod coefficients should be greater than the heterotrait-heteromethod coefficients in the same row and column. The correlation of the Compliance and Extroversion scales meet the criterion of being greater than their correlation with the other two traits. The Interpersonal Affect scale does not meet that standard only because of the slightly higher correlation between the SPAF Compliance and ParSPAF Interpersonal Affect scales.

The second aspect of discriminant validity is illustrated by determining whether the convergent validity coefficient (superscript c) is greater than the heterotrait-monomethod coefficients (superscript a). No scale met the more stringent requirements of this second aspect of discriminant validity. The correlation between the SPAF and ParSPAF Compliance scale was surpassed by the correlation of the Compliance scale with the Extroversion and Interpersonal Affect scales on the SPAF and the Interpersonal Affect scale on the ParSPAF. The Extroversion convergence coefficient was lower than that of the SPAF Compliance-Extroversion correlation and the correlation of

the two Interpersonal Affect scales was less than that of the Compliance and the Interpersonal Affect scales for both instruments and the Extroversion-Interpersonal Affect correlation of the ParSPAF.

The third aspect of discriminant validity is determined by looking at the pattern of relationships among the traits. This is accomplished by examining the different trait-same method and different trait-different method triangles. Examination of Table 13 shows that these patterns are similar. The major difference between the two sets of patterns is shown by the Compliance-Extroversion and Interpersonal Affect-Extroversion relationships of the heterotrait-monomethod coefficients (superscript a). The heterotrait-heteromethod coefficients (superscript b) show a slight relationship between the Compliance-Interpersonal Affect scales and very little or no relationship for Compliance-Extroversion and Extroversion-Interpersonal Affect coefficients. This indicates that the SPAF-ParSPAF relationship meets the requirements for the third aspect of discriminant validity to a limited degree.

Discussion

Study three considered the convergent and discriminant properties of the SPAF. Prior to this, the factorial validity of the ParSPAF had to be determined. A PCA with Varimax rotation resulted in the replication of the Compliance, Extroversion and Interpersonal Affect scales.

The convergent validity coefficients of the SPAF/ParSPAF scales were not high although they were statistically significant. There are several possible reasons for this, among them: (1) the parents or teachers were poor observers, (2) the parents or teachers were biased, (3) there was a selection bias created by voluntary participation by the parents, (4) the salient behavioral features of the personality dimensions vary from school to home or (5) the teachers were rating their personal construct of the child, not "reality". Each of these possibilities is discussed below.

If either the parents or teachers were poor observers, the factorial structure of the scale would vary from one administration to the next. Poor observation skills would result in random response patterns rather than similar components. The component pattern created when teachers complete the instrument has been very stable. This suggests that the sample of teachers surveyed thus far are not poor

observers. The component pattern resulting when the parents completed the form was very similar to that found with the teacher's. This suggests that they are also adequate observers. This hypothesis is not consistent with the data.

The question of whether teacher or parent rating errors result in the low level of convergence can be considered by examining the means and standard deviations of the respective scales. One type of rating error would be a leniency error where a group of respondents tend to rate the children high and the other doesn't. In that case, the means would vary significantly. In the current study, the means were very similar which suggests that neither group was demonstrating a leniency error. A second type of rating error would be a central tendency error as a result of the parents having access to a limited comparison group and would be indicated by less variance. Less variance would occur as a result of the parents not using the extremes of the rating scale. This occurred with this sample. Therefore, a partial explanation of the limited convergence appears to be explained by a Central Tendency rating error resulting from the lack of a large comparison group such as the teachers have.

The third question, whether a selection bias was operating through the voluntary participation of the parents can not be dealt with in this study. However, the results

of data analysis not included in this paper indicated that screening test results for the sample of children whose parents completed questionnaires was similar to that of the total group. This suggests that a selection bias was not contributing to the limited convergence.

The question of whether the salient features of the personality dimensions vary from school to home is a question which has been discussed for years by psychologists. This discussion has taken many forms, ranging from the view that behavior is completely situation specific to the view that behavior is controlled by genetic and immutable traits, with most personality theorists falling between these extremes. The current study did not address the question directly. However, predictions can be made concerning the expected interscale correlations if the traits are viewed as being a biological or genetic predisposition toward a behavioral style but which are not immutable nor unaffected by the environment (Thomas & Chess, 1977). It could be predicted that the Compliance scales would result in the highest correlation since the scale is a measure of general achievement orientation which involves perseverance and a lack of impulsivity. These are behaviors which would be similar at home or school. The next highest correlation would be expected between the Extroversion scales. As mentioned above, the parents would lack a large comparison group, but would have a chance to observe how

their child usually reacts to new situations and people as well as how other children react. The lowest correlation would be expected between the Interpersonal Affect scales. The parents would have little opportunity to observe their child interacting with a large group of children or with various individuals in a large group. These predictions were supported by the current research and other studies (Field & Greenberg, 1982). This suggests, then, that a difference in the salient features of the personality dimensions in question contribute to the low convergence between the parent and teacher ratings.

The question of whether teachers are rating their personal construct of the child rather than reality is not addressed directly by this study. However, it can be considered by examining the correlation between the teacher and parent ratings and other measures, such as grades and achievement test scores. This is dealt with in the next study where the results indicate that this is not a supportable hypothesis.

The low convergent validity coefficients, then, are potentially the result of: (1) a Central Tendency rating error by the parents due to a limited comparison group, and (2) a difference between home and school in the salient features of the personality dimensions in question.

The three SPAF scales had more difficulty demonstrating discriminant validity than convergent validity. This may be the result of (1) a halo rating error, which is a tendency to rate the person in the direction of the rater's general impression rather than how the person is on each specific item or (2) a logical rating error, which is shown by a tendency to give similar ratings on items which appear to be related even though the traits being measured are independent. In either case, the results would be a lack of divergence in the ratings which would explain the low discriminant validity. However, with the halo error, all three scales should be highly correlated whereas with the logical rating error, traits which seem similar should be highly correlated and less similar appearing traits should be less correlated. These results suggest that a logical rating error decreased the discriminant validity in this study. This is shown by the Compliance and Interpersonal Affect scales being highly correlated with both instruments which suggests that parents and teachers see items reflecting these dimensions as being similar and may have tried to rate their children consistently. Teachers also seem to view Extroversion items as being related to Compliance items. Parent ratings do not reflect this rating error. Conversely, teachers appear to see little relationship between Extroversion items and Interpersonal Affect items and parents see a stronger relationship between these items. One method of decreasing the effect of this

error would be to emphasize in the instructions that each item is to be considered independently and even if an item seems similar to another it may be rating a different aspect of personality. A second method of dealing with this error would involve altering items to make them seem less similar.

In summary, the SPAF has met some and failed to meet some of the requirements of convergent and discriminant validity. The requirement of convergence was met but indicated a need for further refinement of the scale to deal with a central tendency rating error, as well as further definition and identification of the salient features of the three dimensions. The difficulty in meeting the discriminant validity requirements reflected a logical rating error. This problem may be diminished by altering the instructions to emphasize considering each item independently and by refinement of the items to decrease content similarity.

Another question which must now be considered is whether the personality dimensions being investigated have any relationship to academic success.

Study Four: Relationship of the SPAF and Other Scales to Academic Achievement

This study examined the relationship between personality as measured by the SPAF and school achievement as rated by: (1) a cognitively oriented screening test, (2) a standardized achievement test, (3) grades and (4) conduct evaluations by the teachers.

The relationship between the screening test and the scales of the SPAF and ParSPAF was examined by multiple regression analysis to determine the amount of variance of the screening test predicted by the personality scale. Research suggested that there would be a low level of correlation between the two measures (Carey, 1970; McDevitt & Carey, 1978; Thomas & Chess, 1977). A second set of multiple correlation coefficients were computed to examine the ability of the screening test to predict scores on the achievement test, grades and conduct ratings. A third and fourth set of multiple correlation coefficients were computed to determine if the results of the personality ratings by parents or teachers could enhance the predictive ability of the screening test. A fifth and sixth set of coefficients were computed to examine the relationship between the SPAF and ParSPAF dimensions and achievement test

scores, grades and conduct ratings. A seventh MRA looked at the relationship between a behavior rating scale and achievement or grades. An eighth examined the relationship of the personality scales, screening test, and behavior rating scale when combined, to achievement test scores and grades.

There were five a priori predictions made concerning these variables. The first a priori prediction was that the MSSST would predict the objective measures best, followed by the DESB and then the SPAF. This ordering of predictor sets is from the most cognitive measures to least cognitive measures. However, this prediction must be modified in consideration of the fact that the MSSST results were compared to the achievement test scores completed one or two years later. To control for this, the 1981 administration of the SPAF can be employed to predict academic measures the following year. The correlations of the MSSST and 1981 SPAF could then be compared, with the assumption being that the results would reflect their relative ability to predict academic measures if concurrent scores and grades had been available. It was predicted, then, that the MSSST would predict objective academic measures better than the 1981 SPAF but not as well as the DESB or 1982 SPAF because of the difference in elapsed time between administration of the predictive and criterion measures, and that the DESB would predict the measures better than the 1982 SPAF.

A second a priori prediction was that the DESB would predict the subjective academic measures of Reading and Math grades better, followed by the SPAF and MSSST. The latter two predictor sets should be equally capable of predicting those subjective measures. The DESB was designed to correlate with behaviors which either inhibit or enhance academic achievement. Since grades are determined by both behaviors and completed academic assignments, the DESB should predict grades better than a personality rating scale or cognitive screening test. Conversely, the SPAF and MSSST should predict equally well since the SPAF does not consider cognitive variables, and the MSSST does not consider personality or behavioral variables.

A third a priori prediction was that the DESB would predict the subjective Conduct grade best, followed by the SPAF and then the MSSST. Conduct grades are based on overt behaviors and overt behaviors are measured by the DESB. The SPAF should be the next best predictor, since it is a personality measure and its rating is based on general behaviors. The MSSST should be least able to predict conduct grades since it is based on cognitive processing skills which should be related to Conduct grades only to a limited degree.

A fourth a priori prediction was that the SPAF-DESB-MSSST combination would be the best set of predictors for both objective and subjective academic measures, followed by the SPAF-DESB combination and finally the SPAF-MSSST combination. The SPAF-DESB-MSSST combination should be the best predictor set since they are measuring overt behaviors, personality variables and cognitive variables. The SPAF-MSSST combination should next since there is little redundancy between these measures. However, the predictability of these sets will be reduced since the MSSST is being compared to grades and achievement test scores one or two years later. The SPAF-DESB combination should be the least able to predict objective and subjective achievement measures since they were shown to be highly redundant in earlier studies.

A fifth a priori prediction was that the subjective measures would be predicted more accurately than the objective. This would be a reflection of a halo rating error, which is a tendency to rate a person in the direction of the rater's general impression rather than objectively on each item. This is predicted since the teachers completed the SPAF and DESB and also decide on grades.

The SPAF was not designed to be used as the prime test in predicting children who are at risk of developing academic problems. It may, however, prove useful in

conjunction with other tests as there are many studies which suggest a relationship exists between personality and academic achievement (Rubin & Balow, 1979; Thomas & Chess, 1977) It is this relationship which is discussed below.

Method

Subjects

The data for this study was gleaned in part from each of the studies discussed above. The SPAF data was produced by the 1981 administration of the 90 item form (Study One) or the 1982 administration of the 43 item form (Study One, Section 2). The ParSPAF data was produced in conjunction with Study Three. The MSSST was administered as a routine kindergarten screening instrument and was available for many of the students. The DESB was completed as a part of Study Two described above. The CTBS was administered to all children in the target school district as are reading, math and conduct grades. All of the information pertained to kindergarten, first or second grade children in a centrally located Rhode Island School District. The total sample size was 286. Twenty one teachers from five schools participated.

Instruments

The instruments used in this study were the SPAF, ParSPAF, Meeting Street School Screening Test (MSSST; Hainsworth & Sigueland, 1969), Devereux Elementary School Behavior Rating Scale (DESB; Spivack & Swift, 1966), Comprehensive Test of Basic Skills (CTBS; McGraw-Hill, 1982) and grades.

The SPAF is a 34 item questionnaire which has been designed to measure Compliance, Extroversion and Interpersonal Affect. The second instrument, called the Parent Student Personality Assessment Form (ParSPAF) is a version of the SPAF modified to reflect the home rather than the school setting. Changes were made when an item referred to class seating arrangements, or the teacher and were kept to a minimum. The ParSPAF was used in the same manner as the SPAF.

The MSSST is a screening test designed for early identification of children with learning disabilities. It is based on an information processing model which states that before information is useful to a child, the child must be able to "orient himself, selectively take in the appropriate information, precisely discriminate its details and retain their pattern or sequence long enough to associate with past learning in the appropriate category or

set. Formulating or planning an appropriate action.....requires selecting the appropriate modality and pattern for action and planning and moving precisely through the steps of execution, while checking the appropriateness of the action through its effect on the environment (Hainsworth and Sigueland, 1969, pg 7). The purpose of using the MSSST is to examine how effectively a child is processing information via three modalities: gross motor, visual-perceptual-motor, and language. In line with this purpose, there are three subtests: (1) Motor Patterning, which examines bilateral large motor movement and spatial body awareness, (2) Visual-Perceptual-Motor, which examines visual discrimination and memory, fine motor skills, and ability to follow instructions requiring understanding of spatial and directional concepts, and (3) Language which examines listening skills, short term rote memory for verbal material and expressive language. The scaled scores for each of the three subtests were utilized for subsequent analysis. The total score was not used since it is a composite of the three subtest scores.

The MSSST was standardized using 500 Rhode Island children who were representative of a cross section of the general population according to the 1966 U.S. Census. Test-retest reliability (two to four weeks apart) of the total MSSST score was .85 and for the subtest scores was .75 to .85. Inter-rater reliability coefficients were

consistently above .95 for the three scores. Predictive validity studies conducted with the normative sample (220 kindergarten, and 274 first grade children) indicated a significant correlation between achievement and the MSSST. The correlations between MSSST scores and readiness scores at the end of kindergarten and achievement scores at the end of first grade was .66 and .63 respectively for the kindergarten sample. Similar results were found for the first grade sample where the correlation was .53 with first grade achievement and .46 for second grade achievement. Multiple regression analysis indicated that the MSSST and IQ scores accounted for approximately 50 per cent of the achievement variance whereas IQ by itself accounted for only 25 per cent. The MSSST was selected for inclusion in this study since it is representative of the "better" screening tests. "Better" refers to the availability of reliability, validity and normative data.

The DESB is made up of 47 items designed to measure eleven dimensions of school behavior. The eleven behavior factors are: (1) Classroom Disturbance, (2) Impatience, (3) Disrespect-Defiance, (4) External Blame, (5) Achievement Anxiety, (6) External Reliance, (7) Comprehension, (8) Inattentive-Withdrawn, (9) Irrelevant Responsiveness, (10) Creative Initiative, and (11) Need for Closeness to the Teacher. A more complete description of the DESB is given in Study Two.

The CTBS is a norm referenced achievement test which measures reading, math, spelling, language, science and social studies. It is a timed, group administered test. The kindergarten test requires 1 hour, 8 minutes, first grade 1 hour 59 minutes and second grade 3 hours, 38 minutes for administration. Scores are reported as national percentiles, local percentiles, stanines or grade equivalents. A confidence interval is shown graphically and varies by test area and grade level. This study used the stanine scores since, by definition, they are distributed normally. The mean stanine score, for this sample, (N=101) was 4.1 for both reading and math with a standard deviation of 2.7 and 2.6 respectively.

The grades and attitude ratings were those given at the end of the fourth quarter. There are two grading systems used by the teachers in this study. The first and second grade teachers give grades of A, B, C, D, or F for reading and math. In this study, the letter grades were coded as: A=5, B=4, C=3, D=2, F=1. The mean reading grade for this sample (N = 129) was 3.95, with a standard deviation (SD) of .91. The mean math grade was 4.15 with an SD of .89.

The kindergarten teachers are not required to grade their students. However, three classes in this study were graded, using G to indicate "very good", S to indicate "satisfactory" I for "needs improvement" and U for

"unsatisfactory". These grades were coded as: G=4, S=3, I=2, and U=1. The mean reading grade for the kindergarten sample (N = 43) was 2.86 with the SD equal to .71. The mean math grade was 2.91, with an SD of .61.

The grading system for conduct was the same for the kindergarten, first and second grade, where: G = very good, S = satisfactory, I = needs improvement and U = unsatisfactory. The coding system was the same as that used for the kindergarten reading and math grades. The mean conduct grade, for this sample, (N = 169) was 3.15 with an SD of .56.

Procedure

As indicated previously, each teacher was asked to complete a SPAF for each of ten randomly chosen students. Student selection was accomplished by generating random numbers according to class size and selecting students by their placement on the class list. The class lists were separated by sex, therefore, subject selection was alternated by sex. The personality questionnaires were completed in April and May, 1981 and April and May, 1982. The teachers also completed a DESB on the same children as the SPAF for the 1982 administration. As the 1982 SPAF forms were completed, a Parent SPAF was mailed to the parents of the child to whom it pertained. The MSSST was

administered by volunteers to each child in kindergarten as a part of a routine screening program. Since it was administered while the child was in kindergarten the data for the MSSST reflected the child's functioning one or two years prior to the completion of the SPAF for first and second graders respectively. The achievement test scores and grades reflected the child's current functioning for the SPAF completed in 1982. The SPAF completed in 1981 was correlated with achievement test scores and grades given the following year. These procedures resulted in a conservative estimate of the correlation of the 1981 SPAF with grades and achievement test scores. However, it also resulted in being able to compare the correlation of the SPAF to current grades and achievement and to grades and achievement one year after completion of the form.

Results

Relationship of the SPAF, ParSPAF and MSSST

Multiple regression analyses (MRA) were computed to examine the relationships of the variables in this study. Table 14 shows the relationship of the SPAF and ParSPAF with the MSSST scaled scores. It is obvious that there is very little shared variance between the SPAF or ParSPAF and the MSSST. The zero order correlation coefficients ranged from

Table 14

Pairwise and Multiple Correlation Coefficients for SPAF and ParSPAF
With MSSST Scaled Scores

1982 SPAF ^a	Motor patterning	Visual- perceptual-motor	Language
Compliance	.11	.13	.09
Extroversion	.02	.05	.05
Interpersonal Affect	.04	.05	.00
Multiple R	.13	.15	.13
Multiple R ²	.02	.02	.02
Adjusted R ²	.00	.00	.00
ParSPAF ^b			
Compliance	-.14	-.23	-.11
Extroversion	.10	.07	.04
Interpersonal Affect	-.18	-.22	-.16
Multiple R	.23	.29	.18
Multiple R ²	.05	.08	.03
Adjusted R ²	.02	.05	.00

^a n = 101

^b n = 98

.00 to .23. The multiple correlation coefficients ranged from .13 to .29. These results support other research (McDevitt & Carey, 1978; Thomas & Chess, 1977) showing low correlations between personality instruments and cognitively oriented tests. It is apparent from this analysis that the relationship between the cognitively oriented screening test and SPAF or ParSPAF is extremely limited, whereas the results of Study Two indicated that the correlation of the behavior rating scale and the SPAF was high. With that in mind, the relationship of the SPAF and other instruments, including the cognitive and behavior rating scales, with achievement and grades was considered.

Relationship of the Individual Predictor Sets to Achievement and Grades

Table 15 shows the results of twenty five different MRAs. The 1981 SPAF, 1982 SPAF, ParSPAF, DESB, and MSSST were employed individually as the set of predictor variables. The CTBS reading stanine, CTBS math stanine, reading grade, math grade and conduct grade were employed as criterion variables. Only the final multiple correlation coefficient is presented.

Examination of the results for the two Achievement tests, CTBS reading and CTBS math, shows that the SPAF administered in 1982 produces the highest multiple

Table 15
Multiple Regression Coefficients for Individual Predictor Sets With
Achievement and Grades

Predictor set	Criterion variable				
	CTBS reading	CTBS math	Reading grade	Math grade	Conduct grade
1981 SPAF ^a	.28	.26	.37	.40	.39
1982 SPAF ^b	.57	.44	.24	.16	.36
ParSPAF ^c	.18	.19	.11	.10	.10
DESB ^d	.48	.37	.52	.51	.59
MSSST ^e	.16	.21	.38	.42	.23

^a n = 86

^b n = 101

^c n = 98

^d n = 91

^e n = 101

correlation coefficients. The DESB produced the second highest multiple correlation coefficients. The 1982 SPAF and the CTBS were administered during the same time period. The SPAF administered in 1981 resulted in low to moderate correlation coefficients. The ParSPAF and MSSST resulted in very low correlation coefficients. The pattern of correlation coefficients are nearly the same for CTBS math achievement scores as for reading achievement scores, but lower.

The pattern of coefficients involving the reading grades is different than the pattern with math and reading achievement scores. As Table 15 shows, the DESB yields the highest coefficient, followed by the 1981 administration of the SPAF. The 1982 administration of the SPAF does not predict reading grades as well as the SPAF completed the year before by a different teacher. The MSSST predicts reading grades as well as the 1981 SPAF, showing a moderate correlation. The ParSPAF did not predict reading grades.

The pattern for math grades is similar to that produced with reading grades. The best predictor set was the DESB with the next highest coefficient being produced by the MSSST followed by the 1981 SPAF. The coefficients for the MSSST and 1981 SPAF were very similar. The 1981 administration of the SPAF correlated with 1982 math grades by a different teacher resulted in a higher coefficient,

once again, than the SPAF completed by the same teacher giving the grade. The 1982 SPAF and ParSPAF both resulted in very low coefficients.

Table 15 reveals that the pattern of coefficients with conduct grades is similar to that produced by reading and math grades. Again, the predictor set which yields the highest coefficient is the DESB. The conduct grade is predicted better than the reading or math grade. The 1981 SPAF yields the next highest coefficient but, in this case, it is only slightly higher than the 1982 SPAF. The MSSST produced a low to moderate coefficient. The ParSPAF once again, resulted in a low coefficient.

Relationships of Combined Predictor Sets to Achievement and Grades

The separate predictor sets were intermixed in the following three combinations: (1) SPAF (1982), DESB, MSSST, (2) SPAF (1982), DESB, and (3) SPAF (1982), MSSST. These combinations were employed for several reasons: (1) The SPAF was the instrument of major interest, therefore, it was used in each combination, (2) the rating by the teacher giving the grade would reflect the use of the instrument in a practical situation, therefore, the 1982 rather than 1981 SPAF was used, (3) the SPAF and ParSPAF were developed to be redundant and were therefore not combined and (4) the

ParSPAF did not predict achievement or grades, therefore, it was not combined with other instruments.

Examination of the multiple correlation coefficients of various combinations of instruments with the CTBS reading achievement score (Table 16) shows that: (1) the SPAF, DESB and MSSST combination produced the highest multiple correlation coefficient, followed closely by (2) the SPAF and DESB. The SPAF and MSSST combination, also, resulted in a moderate to high correlation. Examination of the multiple correlation coefficients produced with CTBS math achievement as a criterion score when the scales are combined shows a pattern very similar to that produced with reading achievement scores in that: (1) The SPAF-DESB-MSSST is the best combination, followed by (2) the SPAF-DESB combination. Both coefficients are lower than the corresponding correlation with reading achievement. The SPAF-MSSST combination resulted in a slight increase over the SPAF by itself (see Table 15).

The MRA of the reading grade with the combined scales resulted in the DESB-SPAF-MSSST combination, once again, producing the highest coefficient followed by the DESB-SPAF combination. The SPAF-MSSST combination produced only a moderate correlation. The SPAF, DESB and MSSST combination also produced the highest multiple correlation with the math grade. The SPAF-DESB combination produced a slightly lower

Table 16

Multiple Correlation Coefficients for Combinations of Predictor Sets

With Achievement Tests and Grades

Predictor set	Criterion variable				
	CTBS	CTBS	Reading	Math	Conduct
	reading	math	grade	grade	grade
SPAF/DESB/MSSST ^a	.69	.55	.63	.66	.67
SPAF/DESB ^b	.67	.53	.57	.59	.64
SPAF/MSSST ^c	.61	.50	.44	.44	.43

^a n = 78

^b n = 78

^c n = 101

coefficient. The SPAF-MSSST combination resulted in a moderate coefficient. Similar results were obtained with the conduct grade. The SPAF-DESB-MSSST combination resulted in the highest correlation coefficients. The SPAF and DESB without the MSSST resulted in a coefficient which was only slightly lower. The SPAF-MSSST combination resulted in a low to moderate correlation coefficient.

Discussion

Study Four consisted of several multiple regression analyses with the 1981 SPAF, 1982 SPAF, ParSPAF, DESB or MSSST as predictor variables and the CTBS reading score, CTBS math score, reading grades, math grades, and conduct grades as criterion variables. The predictor variables were utilized individually and in various combinations to determine their relationship to the criterion variables. The variables of major interest in this study were the 1982 SPAF, the DESB, and the MSSST and their relationship to objective academic measures (achievement test scores) and subjective academic measures (grades).

The results did not support the first a priori prediction of the MSSST being the better predictor of objective academic measures. Rather, the 1981 SPAF was able to predict both CTBS reading and math scores slightly better than the MSSST. The relationship of the 1982 SPAF and DESB

to the objective measure also failed to satisfy the prediction. Instead, the SPAF was the better predictor followed by the DESB. This suggests that personality, as measured by the SPAF, may be more important to academic achievement than previously assumed. It should be noted that the difference between the MSSST and 1981 SPAF coefficients is small as is the difference between the 1982 SPAF and DESB coefficients. These differences could be occurring because of sampling error. The results do demonstrate that the SPAF is just as capable of predicting achievement as are more cognitive or behaviorally oriented instruments. This implies that personality variables as measured by the SPAF may affect learning in the early school years as much as overt behaviors or cognitive abilities. In other words, a child's personality, as reflected by his/her general achievement orientation, level of extroversion, and interpersonal affect may effect academic progress to a great extent during the early school years.

The second a priori prediction was supported as stated. The DESB was the best predictor of reading and math grades, followed by the 1981 SPAF and MSSST which were approximately equal. The fact that the DESB met the assumptions of the second a priori prediction supports the supposition that grades are determined by behavior as well as measures of completed academic assignments. In addition, the fact that the 1981 SPAF and MSSST were approximately equal in their

ability to predict reading and math grades also supports the supposition that grades are based on more than personality or cognitive ability by themselves. An unusual aspect of these results was that the 1981 SPAF predicted the subjective academic measures given by a different teacher the following year better than the 1982 SPAF did. This may have occurred due to differences in the samples. The 1981 SPAF sample consisted of more first and second grade than kindergarten students whereas the 1982 sample was equally split between kindergarten and first or second grade students. The higher predictability of the 1981 sample seems to be a reflection of increased accuracy of grades mirroring ability in the first and second grade. This occurs because first and second grade students are involved in a formal reading and math curriculum whereas kindergarten students are not.

The third a priori prediction was also supported. The DESB was the better predictor of conduct grades followed by the SPAF and then the MSSST. The 1981 and 1982 administrations of the SPAF were approximately equal in their ability to predict conduct grades. These results indicate that conduct grades are determined to a large extent by overt behaviors relating to academic success (indicated by the DESB). They, also, indicate that personality variables play a part in the determination of conduct grades. The similarity between the 1981 and 1982 SPAF administrations suggest that: (1) the relationship

between personality variables and conduct grades may be stable or (2) that conduct grades are as accurate for kindergarten children as they are for first or second grade students.

The fourth a priori prediction, that the SPAF-DESB-MSSST combination would be the best predictor, the SPAF-MSSST would be next and the SPAF-DESB combination would be third in their ability to predict both objective and subjective academic measures was not met. The SPAF-DESB-MSSST combination was the best predictor set, the SPAF-DESB combination was next and the SPAF-MSSST combination was third. Similar results were found with the subjective measures. As predicted, the SPAF-DESB-MSSST combination was the best predictor set. However, the next best predictor set was the SPAF-DESB, instead of the SPAF-MSSST. The results are somewhat different, however, if the relative contribution of each predictor set is considered because of the difference in lag time between administration of the predictor sets and criterion measures. Since the MSSST was compared to the criterion measures given one or two years later, a better comparison with the SPAF and DESB is possible by considering the change in the multiple correlation coefficients when each predictor set is added. The correlation of the SPAF-DESB combination with the CTBS reading score is increased by .10 over the individual SPAF coefficient. Adding the MSSST increases it

only .02. Similar results are obtained with the CTBS math score. Different results are obtained with the subjective measures. The MSSST increases the coefficient obtained with the best predictor (DESB) as much as the SPAF. The results examined in this way still only partially support the fourth a priori prediction. The objective measures are predicted better by the personality and behavior rating scales with little contribution from the cognitive screening instrument. Conversely, subjective measures are predicted better by a behavior rating scale and cognitive screening instrument than by the personality scale. These results imply that the redundancy between the SPAF and DESB is evident in the limited increase in the coefficients of the subjective measures. However, the SPAF and DESB appear to be measuring different aspects of achievement with the objective measures. This suggests that redundancy is not an appropriate explanation of the high coefficients obtained with both the SPAF and DESB and objective measures.

The fifth a priori prediction, that subjective measures would be predicted better than objective, was not fulfilled. Instead, the subjective and objective measures were predicted approximately equally. This suggests that a halo rating error was not affecting the ratings by the teachers. This conclusion supports the results of Study Three which suggested that a logical rating error was reducing the discrimination of the SPAF-ParSPAF scales.

The results of this study serve to demonstrate the complexity of academic achievement whether measured objectively or subjectively. It has been recognized for many years that cognitive abilities affect academic progress and it has recently been recognized that other factors affect that progress just as much. Behavioral factors have been attracting considerable attention with studies such as those by Spivack and Swift (1969). One area which has been neglected, however, is the effect of personality on specific areas in the educational setting. This study has made a move in that direction and has suggested that personality variables predict objective measures of academic progress more than the subjective measures. These results must be corroborated and extended by future studies.

An extension of this study could include having parents rate their children as they believe they would be in school to determine if the relationship between parent rating and teacher ratings would increase and to see if the parent ratings would correlate more highly with achievement test scores and grades. Another area which needs to be investigated is the determination of personality profiles thru the use of cluster analysis. The relationship of the clusters and achievement could then be examined. Another area which needs to be considered is how the personality data can be used to enhance educational quality in the school.

In conclusion, this study was undertaken to determine the ability of the SPAF and several other scales to predict objective and subjective achievement measures. It has been shown that both individually and in combination the SPAF, DESB and MSSST can account for a considerable amount of the variance. There were also suggestions made concerning research ideas arising from these results. The implications of this study and the three discussed before it now need to be considered as a whole.

Discussion and Implications

This project was begun as the result of a study of the psychometric properties of the Teacher Temperament Questionnaire (TTQ) (Thomas & Chess, 1977) conducted by Baker and Velicer (1982) which identified three stable dimensions and a fourth less stable dimension of childrens personality instead of the eight hypothesized scales. These were labeled Compliance, Extroversion, Interpersonal Affect and Environmental Sensitivity . The general purpose of the current project was to identify other personality dimensions, incorporate them into a rating scale, and then examine the psychometric properties of the scale. Four complementary studies were designed to accomplish this. These four studies are somewhat lengthy and complex. Therefore, the first part of this section is a summary of the results and comments followed by a discussion of: (1) the theoretical implications for the field of child personality, (2) empirical implications for the field of child personality, (3) practical implications for the field of education. and (4) limitations and future directions.

The first study involved developing and examining the construct validity of the Student Personality Assessment Form (SPAF) To do this, a 90 item questionnaire was constructed to identify the traits of: (1) Compliance, (2)

Extroversion, (3) Interpersonal Affect, (4) Environmental Sensitivity, (5) Organization, (6) Aggression, and (7) Impulsivity. A principal component analysis with Varimax rotation resulted in the original three stable scales being reproduced as well as the fourth scale, which was poorly defined once again. Items which had been added to delineate Organization and Impulsivity loaded with Compliance items while the Aggression items loaded with the Interpersonal Affect items. The original Environmental Sensitivity items formed a component but the new items designed to define the scale more completely did not load on it.

Since the SPAF dimensions were not reproduced as expected, a refined SPAF was developed and analyzed. This version of the SPAF was designed to measure Compliance, Extroversion and Interpersonal Affect as before. However, items were added to those scales from the Organization, Impulsivity and Aggression item pools. The Environmental Sensitivity items were re-examined and new items were added to see if the scale was measuring verbal facility. The scale was aptly renamed Verbal Facility.

A principal component analysis with Varimax rotation reproduced the now embellished scales of Compliance, Extroversion, and Interpersonal Affect but not Verbal Facility. These results supported the factorial validity of the three scales of the SPAF across different subjects and

items. The scales also demonstrated good internal consistency. Subsequent research, on construct and criterion-related validity, was, therefore, conducted using the 34 items designed to reflect the three well defined scales.

Study two examined construct validity via univariate and multivariate correlation analyses of the relationship between the SPAF and DESB. The results of this study supported the validity of the SPAF. The univariate and multiple regression analyses showed a low score on the Compliance scale (indicating non-compliance) to be related to inattentive, poorly controlled behavior, with a concomitant lack of understanding of material presented in class. A low score on the Extroversion scale (indicating introversion) reflected inattentive, withdrawn behaviors without the lack of control noted with non-compliance. The Extroversion scale was only moderately correlated with the DESB which indicates that the Extroversion scale is not well defined by the behaviors described by the DESB factors. A low score on the Interpersonal Affect scale (indicating poorly developed interpersonal skills) reflected poorly controlled, acting out behaviors in conjunction with a tendency to blame others for problems and to be disrespectful and defiant.

The Canonical Variate Analysis (CVA) also supported the theoretical definitions of the SPAF scales. The linear composites produced presented a behavioral description of the personality dimensions commensurate with their definitions.

The results of Study Two supported the construct validity of the SPAF by demonstrating a relationship between the three scales and certain specific behaviors as measured by the DESB. Both instruments, however, were completed concurrently by the same person. These results, then, could be a reflection of a halo rating error rather than a reflection of objective behaviors. Study Three was designed to control for this.

Study three considered the factorial validity of the parent version of the SPAF (ParSPAF) and examined the convergent and discriminant validity of the SPAF by a multitrait-multimethod matrix (Campbell & Fiske, 1967).

The PCA with Varimax rotation resulted in the ParSPAF reproducing the same three stable dimensions that were found with the teacher completed SPAF, namely, Compliance, Extroversion, and Interpersonal Affect. The Verbal Facility scale could not be reproduced with the ParSPAF or SPAF. Since these results confirmed the expectation that the ParSPAF consisted of at least the same three stable

dimensions found with the SPAF, it was possible to examine the convergent and discriminant validity of the SPAF.

A moderate level of convergence was found between all three scales of the SPAF and ParSPAF. This level of convergence was not as high as was hoped. However, they were similar to those found in other research comparing teacher and parent personality ratings (Field & Greenberg, 1982; Hubert, Wachs, Peters-Martin & Gandour, 1982). There were several possible reasons considered for the low convergence. Those reasons which appeared to be contributing were: (1) a central tendency rating error by parents due to a limited comparison group and (2) differences between home and school in the salient features of the personality dimensions being measured. Some aspects of discriminant validity were satisfied and others were not, resulting in a limited level of discriminant validity overall. The limited discrimination appeared to be the result of a logical rating error in that the parents and teachers were attempting to rate the children in a logically consistent manner instead of rating each item independently. It was apparent from Study Three that there was support for considering the SPAF dimensions as reflecting external personality variables and not just method variance. There were also indications of areas needing refinement. The SPAF scales also needed to be compared to other external variables to determine if a relationship existed between

ratings and other external criteria.

Study four used multivariate regression analyses to determine the relationship of the SPAF, ParSPAF, DESB, and MSSST individually and in various combinations to CTBS reading scores, CTBS math scores, reading grades, math grades and conduct grades.

The results indicated that the ParSPAF was unrelated to achievement test scores or grades. The SPAF was found to predict achievement test scores and conduct grades moderately well. However, its ability to predict reading and math grades was rather limited. Conversely, the DESB was able to predict grades better than achievement test scores. The MSSST did not predict achievement test scores or conduct grades as well as the 1981 SPAF but did predict reading and math grades better than the 1982 SPAF. The potential utility of the SPAF, DESB and MSSST together was demonstrated. These results suggested that personality ratings may be more useful in the academic setting than previously recognized and that behavior rating and personality scales are not totally redundant.

There have been conclusions and implications discussed at the end of each of the studies. Now, the issue which must be considered is what conclusions can be reached and implications for the project as a whole.

Theoretical Implications

One of the theoretical implications of this project which must be considered pertains to trait attribution and the use of rating scales to identify those traits. It has been stated by many theorists, who disagree with trait theories, that what is being measured by a trait rating scale is not within-the-person behaviors but the raters perception of the person. It has also been stated that the perception bears little resemblance to the behavior of the person being rated (Mischel, 1968). Before this issue can be considered, several aspects of trait ratings must be discussed.

One of the difficulties in using a trait rating scale is determining whether the people using the scale define the trait terms in the same way. In other words, if children are being rated on the trait of distractibility, do all of the raters define the term in a similar manner? This is a question of implicative meaning. There are many studies which suggest that there is a considerable amount of consistency in the way trait terms are defined by raters (Wiggins, 1973). The question of similarity of definition deals with only one aspect of a trait scale, the internal structure. The internal structure of the SPAF was not considered directly. This could be accomplished by asking

teachers to rate hypothetical children. This doesn't appear to be necessary with the SPAF because there have been factor analyses completed on various samples completed by different groups of teachers and parents. These have all resulted in similar factors being reproduced. It should be recognized, that a certain amount of consistency in the internal structure of an instrument must be present for this to occur. The SPAF can be considered, then, to have demonstrated adequate internal structure indirectly.

Another aspect of trait ratings is the external structure of the scale. This pertains to the relationship of the items when applied to an actual situation. The same or similar dimensions should be identified when the results of ratings by a variety of observers are factor analyzed. As mentioned above, this has been accomplished several times with the SPAF. The results each time demonstrated the stability of its' external structure. The consistency of these results, then, indicate that the SPAF trait descriptors can be applied to an actual situation, i.e. ratings of children by teachers. However, there is a third aspect of trait descriptors which must be considered, the semantic meaning (Norman & Goldberg, 1966).

Consistency in the semantic meaning of a trait descriptor is shown by high inter-observer agreement. A high level of agreement indicates that different observers

apply the same trait descriptors to a similar degree to a single individual. For this project, inter-observer agreement was determined between parent and teacher ratings. The results did not show a high degree of agreement. Evidence suggested that this occurred because of a central tendency rating error by the parents due to a limited comparison group and a difference in the salient features of the personality dimensions demonstrated at home and school. This could be considered as an indication that the ratings reflect the raters perception and have little relationship to the attributes of the student being rated. However, before this can be concluded, external data need to be considered. If the SPAF were reflecting only the teacher's perceptions of the students, there would be little relationship between objective academic measures and a moderate to high relationship with grades. The opposite was found in this project. The ratings by the teacher showed a high correlation with achievement test scores and a limited relationship to math and reading grades. Conversely, the parent ratings showed no relationship to grades or achievement test scores. One conclusion which can be made from this is that it may be inappropriate to compare parent and teacher ratings of students from an applied research point of view (Fiske, 1975). An analogy could be made between comparing parent and teacher ratings and comparing a football coach's ratings of his players with those of the player's mothers on a trait such as Aggression. There might

be some correlation but it would probably not be germane to how the football player did in a game. Similarly, the ratings by parents and teachers may show a low to moderate correlation but the parent rating may not be pertinent to how a child does in school.

The SPAF has shown good external structure as well as a strong relationship to an external source of data other than ratings. It did not demonstrate generalizability to raters other than teachers. The results do support the idea that what is being measured by the SPAF is an attribute of the student and not just a perception of the teacher. This implies that trait ratings are not just a reflection of the raters perception but are a measure of actual behaviors reflecting underlying dimensions of the person.

The results also have implications for another theoretical issue. Support was demonstrated for an interactionist theory which states that a child acts on as well as reacts to his environment. (Rothbart & Derryberry, 1982; Thomas & Chess, 1980) The personality dimensions as measured by parent and teacher ratings, were similar in structure. However, the scale was designed to measure personality styles pertinent to the educational setting. The results demonstrated that the teacher ratings predicted achievement test scores very well but parents ratings showed little relationship. This implies that the salient features

of the personality constructs measured by the SPAF are demonstrated in different ways as the environmental structures and expectations change. The expectations of the teacher and the responses to the child at school are different from the expectations of the parent and the responses to the child at home. This results in changes in the salient personality features demonstrated by the child as he/she acts upon and reacts to the different environments, including people.

The results of this project, then, have implications for several theoretical issues. There are also implications for empirical aspects of the results which will be discussed next.

Empirical Implications

The Baker and Velicer (1982) research which found that the TTQ was measuring three well defined dimensions and a fourth less clearly defined dimension was replicated in the results of the first study of this project. Of major interest, however, was the fact that four dimensions were reproduced when an attempt had been made to expand the SPAF to seven. Instead of reproducing seven dimensions, the added scales served to further define the Compliance and Interpersonal Affect scales. It became apparent that

Compliance included organization and a lack of impulsivity and that Interpersonal Affect included a lack of aggression. Subsequent research in Study One reproduced the four dimensions. The other studies in this project dealt with various aspects of validity of the SPAF. The fact that four dimensions were reproduced even when the scale was designed to measure seven is of major interest for empirical implications. Therefore, these results will be considered with respect to: (1) implications, for this project, of studies using personality type scales developed from research other than the NYLS (Thomas, Chess, & Birch, 1968) data base (both child and adult) and (2) implications, of this project, for child personality scales based on the NYLS research.

There are results from several studies which need to be considered in conjunction with the results of this project. The first series of studies was conducted by Tupes and Christal (1961) with adults. They factor analyzed data from peer rating studies using 35 bipolar rating scales developed by Cattell (1945) from 171 trait names. The samples varied in length of acquaintance, type of relationship, type of subject and type of rater. The data were factor analyzed with an orthoganol rather than oblique rotation which Cattell (1945) used. There were five factors which consistently emerged. These were labeled: (1) Extroversion, (2) Agreeableness, (3) Dependability

(conscientiousness) (4) Emotional Stability and (5) Culture. It was concluded from these results that five orthogonal factors were adequate for the expression of personality variables in peer rating studies.

The similarity between the Tupes and Christal (1961) studies and this project is unmistakable. The Extroversion factor is similar to the SPAF Extroversion scale, The Agreeableness factor to the SPAF Interpersonal Affect and the Dependability (conscientiousness) factor resembles the SPAF Compliance scale. The fourth SPAF scale, which was never clearly defined, but was tentatively labeled Environmental Sensitivity and in a subsequent study Verbal Facility, may be related to the Emotional Stability factor. The Culture factor did not resemble any factors in this project.

An additional study by Flint, Hick, Horan, Irvine, and Kukak (1980) examined the construct validity of the California Preschool Social Competency Scale and the Classroom Behavior Inventory. Although neither is a personality scale per se, both were designed to measure habitual behaviors exhibited by the children, which are often included in personality scales. However, the child is rated by degrees of competency on the California scale and on affective behavior with the Classroom Behavior Inventory.

Principal component analysis with Varimax rotation was performed on the 30 by 30 interitem correlation matrix resulting from the California Preschool Social Competency Scale ratings of 1723 children ranging in age from 45 to 77 months. The Scree method (Cattell, 1966) indicated that five factors should be retained. Factor one was labeled Considerateness and resembled the SPAF Interpersonal Affect scale; factor two was labeled Task Orientation and resembled the SPAF Compliance scale; factor three was labeled Extroversion and resembled the SPAF scale by the same name; factor four was labeled Verbal Facility and resembled the factor reproduced by items added to the SPAF Verbal Facility scale to further define it. The fifth factor seemed to be the least well defined and was labeled Response to the Unfamiliar. Items on this scale were concerned with unfamiliar or changing situations but could also reflect a child's emotional stability. The Environmental Sensitivity scale may be related to this scale but research would be necessary to determine this.

The principal component analysis of the 18 by 18 interitem correlation matrix for the Classroom Behavior Inventory resulted in three factors being retained. They were: (1) Task Orientation/ Distractibility, (2) Hostility/Consideration and (3) Introversion/Extroversion. They seem very similar to the SPAF Compliance, Interpersonal Affect and Extroversion scales respectively. The structure

of this and the other instruments utilized in the studies just described are obviously very similar. What must be considered are the reasons for and implications of that similarity.

There are several hypotheses which can be considered. One explanation could be that the similarity occurs due to a logical rating error. If this were occurring, items would be rated to maintain a logical consistency. This error was found in earlier research and can be assumed to be contributing to the similarity discussed here. However, if a logical rating error accounted for all of the similarity there should be no consistent relationship between the personality rating and external variables. The results of the current project do not support this hypothesis.

A second hypothesis is that personality can be described by a relatively small number of traits. This hypothesis is not supported, however, when research by Jackson (1976) is considered. He has developed the Jackson Personality Inventory (JPI) which measures 16 traits which are reproducible by factor analytic techniques. This suggests that items need to be developed very carefully and refined to avoid ambiguity and overlap. One consideration, though, is that the JPI was developed for use with adults rather than children and for purposes far broader than describing classroom performance.

A third and related hypothesis is that children's personality can be described by a relatively small number of traits which are relevant to the rater in a specific environment. This would imply that the salient features of the traits would vary with the different environments. This project provided partial support for this hypothesis. The ParSPAF and SPAF had similar factorial structure but moderate convergence due in part to differences in the salient features. It seems that the underlying personality dimension interacts with the situational demands to produce specific behaviors.

There are two assumptions which can be developed from the hypotheses discussed above: (1) personality in children may be described more simply than adult personality and (2) the salient features or behavior patterns demonstrating children's personality may vary with the environment.

There are, also, other implications stemming from this project that apply to instruments developed from the NYLS research. The findings of this study as well as the earlier research of Baker and Velicer (1982) suggest that the scoring systems being used by many child personality scales be examined carefully. The importance of this is demonstrated by a recent review of the literature pertaining to early temperament by Hubert et al. (1982). They point out that a generally accepted theoretical definition of

temperament has not been developed. However, most of the instruments available operationally define the construct of temperament by "how" the child behaves rather than focusing on "what" the child does. The review examined reliability (test-retest, internal consistency and interrater), stability over time and validity. Test-retest data was generally based on small samples, with moderate correlations produced. Some instruments reported no test-retest data. Internal consistency was generally moderate and interrater reliability was generally high. Estimates of stability over time varied very much depending on the instrument. This data was available for only six scales, all of which were compared to instruments measuring the NYLS categories. No validity data was available for nearly half the instruments reviewed. Methodological problems were very apparent with several others. Convergent validity coefficients were generally low for those studies which did not have obvious methodological problems. Results from concurrent validity studies were generally inconsistent. Predictive validity studies were limited and yielded generally moderate coefficients. Hubert et al. (1982) concluded that "Psychometrically, there appears to be no single satisfactory measure of infant or child temperament" (pg 578). They also point out that the majority of personality scales applicable to children utilize the theoretical model developed from the results of the New York Longitudinal Study (NYLS) (Thomas, Chess & Birch, 1968). Of the 26

instruments reviewed by Hubert et al. (1982), 17 were based on the NYLS results. Of the 11 which were not based on the NYLS, only six were applicable to the early school age child. The other five were developed as infant measures. Of those instruments based on the NYLS, few have been subjected to factorial validity studies. Of the few studies utilizing factor analysis, those by Thomas, Chess and Birch (1970) and Thomas and Chess (1977) are reported to support their theoretical position. However, even those studies do not indicate that the nine categories of temperament are reproduced. Instead, three factors were delineated which was interpreted as support for the constellations of "Difficult Child", "Easy child" and "Slow-To-Warm-Up Child" (Thomas & Chess, 1977, pg 22-23). The question which must be addressed here, is one of interpretation. Thomas and Chess (1977, pg 24) indicate that factor analysis with Varimax rotation resulted in three factors one of which describes the Difficult Child at one pole and the Easy Child at the other. The other two factors are not described. The described factor closely resembles the dimension labeled Compliance in the current study and previous research (Baker & Velicer, 1982). It would seem from this that since the Difficult and Easy Child constellations are descriptions of the opposite poles of a single dimension they need to be renamed to indicate that. In line with this suggestion, Compliance seems to be an adequate descriptor.

As a part of this discussion, it seems that a decision must be made as to whether to interpret data qualitatively or quantitatively. The data on temperament which has been gathered by the NYLS can be interpreted and utilized on a qualitative basis with much support. However, when objective measures, such as questionnaires, are utilized, then, the data must be scrutinized quantitatively and subjected to the rigors of psychometric analysis. When this occurs, the interpretations and scoring methods developed from the NYLS must be re-examined and altered to fit the results of that analysis. The current study, then, suggests a reinterpretation of the Difficult or Easy Child constellations. With this in mind, the concept of three constellations must be reconsidered.

A related topic which has received much attention recently is whether the Difficult Child temperament exists in the child or whether it is a parental perception (Bates, 1980; Kagan, 1982; Plomin, 1982; Thomas, Chess & Korn, 1982). It seems, then, that the question must be reconsidered, since, the Difficult Child concept is a misinterpretation and should instead be viewed as how a child scores on a continuum of three dimensions, viz: Compliance, Interpersonal Affect and Extroversion. For example, instead of a Difficult Child characterization, the rating scale could be interpreted as indicating that a child scores significantly below the mean on the Interpersonal

Affect scale, significantly above the mean on the Extroversion scale and significantly below the mean on the Compliance scale. This profile might be associated with children who are high risk for failure in school or need to be referred for psychological therapy. Research using cluster analytic techniques would be required before a definitive statement could be considered.

There are two recommendations which can be made from this discussion: (1) that many of the current personality scales should consider revising their scoring criteria, and (2) that the Difficult, Easy, and Slow-To-Warm-Up constellations should be revised to reflect psychometric analysis. Consideration can now be given to uses for the SPAF.

Practical implications

The purpose of developing the SPAF was twofold. The first was to produce an instrument which provided a measure of child personality and was psychometrically sound. The second was to determine if there was a relationship between personality and academic achievement.

The first purpose was met since the SPAF demonstrated the ability to measure three personality dimensions consistently and demonstrated adequate validity. There were

also two other dimensions suggested but not clearly defined.

The second purpose was fulfilled by demonstrating a relationship between achievement test scores and the SPAF. There have also been other studies which found a relationship between personality and school functioning. For example, Korn (1977), in a limited study, found a significant relationship between non-adaptability and withdrawal, and low academic achievement test scores. He interpreted that as an indication of a significant relationship between academic achievement and the constellation of the Slow-To-Warm-Up child. Another study, (Cary, Fox, & McDevitt, 1977) using the NYLS categories found that adaptability could predict school adjustment and that non-adaptability and non-persistence were related to impulsivity. However, they did not find a relationship between the diagnostic subgroups (Difficult, Easy, Slow-To-Warm-Up) and school adjustment.

The results of this project and the studies mentioned indicate that there is a relationship between academic achievement and personality as well as school adjustment and personality. It can be assumed from this that there may also be a relationship between behavior problems and personality. The DESB was designed to measure behaviors which affect academic progress and a strong relationship was demonstrated between the DESB and SPAF. However, this

project did not address the relationship between personality and behavior problems and actions which can be taken directly.

However, a recent study by Rubin and Balow (1979) addressed the question of prevalence of behavior problems and can be extrapolated to the use of a personality rating scale. They examined problem behavior and school functioning by following a group of 1586 children from kindergarten to sixth grade. By the final year of the project, 68 percent of the boys and 51 percent of the girls, who had been rated by six teachers, were classified as a behavior problem by at least one teacher rating. Each year, approximately 35 percent of the boys and between 19 and 26 percent of the girls were rated as behavior problems. The authors also found a significant relationship between academic achievement and whether a child was rated as: (1) no behavior problem consistently, (2) a behavior problem inconsistently or (3) a behavior problem consistently.

The authors gave several alternative interpretations of these findings. They were: (1) behavior problems in children are transient so that each year new problems occur and disappear, (2) children's behaviors are relatively constant but teachers observational abilities differ which leads to differing perceptions of problems, (3) children's behavior remains relatively constant but teachers vary in

their opinion as to what constitutes a problem, (4) teachers vary greatly in the classroom environment they create which has a variable effect on children, (5) teacher-child interactions result in different children being a behavior problem within different settings or (6) a combination of the above (Rubin & Balow, 1979, p 536).

When the Rubin and Balow (1979) study is considered in conjunction with the current project, the alternative explanations have different implications. If alternative (1) is correct and behavior problems in children are transient and arise and disappear each year then a personality rating scale would be inefficient but a behavior rating scale would be more efficient since it would yield information concerning appropriate contingency management techniques and would not need to be utilized for the prediction of later behavior. If alternative (2) is correct and children's behaviors are relatively constant but teacher's observational skills differ, then a behavior rating scale would be called for again. If alternative (3) is correct and teachers opinions vary as to what constitutes a behavior problem, then, teacher training is needed. Alternatives (4) and (5) which state respectively that: (1) the classroom environment and (2) teacher-student interactions result in different children exhibiting problem behaviors each year are correct, then, a personality rating scale is needed. If, in conjunction with the assumption

that alternative 4 or 5 are correct an interactionist position is taken which states that the child acts on and affects the environment (including the teacher) while the environment is acting on and affecting the child then a personality rating scale (such as the SPAF) might be used to determine which personality styles fit with which teachers and environments. However, much research is still needed before the SPAF could be used for this purpose. A cluster analysis of the SPAF would be required as well as an evaluation of the effects of various teaching styles on the resulting profiles.

Two implications which can be drawn from this project and studies just discussed are: (1) that there is a relationship between SPAF personality dimensions and academic achievement which can be used to predict achievement test scores and (2) that the SPAF could be used to determine which personality styles fit with which classroom environments as well as teaching styles and to help teachers understand the effects of their behavior on children with various personality styles. More research will be necessary, however, before this can be done efficiently. This and other future directions are discussed in the next section.

Limitations and Future Directions

This project has demonstrated that the SPAF has the potential to provide useful information to researchers in the field of child personality as well as to practitioners in the school. However, there are limitations which must be considered as well as areas which need further exploration.

One of the major limitations of this project is the small sample size and restricted population characteristics utilized in some of the studies. The structure of the SPAF has been replicated by studies using ratings by teachers and parents with a total sample size of approximately 750. However, the samples have come from mostly middle socio-economic status, white populations from Rhode Island or Massachusetts. This limits generalizability. The sample sizes for the validity studies are much smaller and therefore even less generalizable.

These difficulties point out the first area which needs more research. The validity studies need to be replicated with a larger sample which is more representative of the current U.S. population. The replication could also be extended to become a longitudinal study. With this design, ratings of a student by different teachers could be compared as well as the stability of the personality variables over time.

A second limitation is the number of dimensions being measured. Three dimensions have been reproduced consistently. There were two other dimensions suggested but never clearly defined. It is apparent from this, that other dimensions could be measured. The items would have to be chosen carefully and then refined to avoid overlap. Scale extension then, is another potential research area.

Another limitation which must be recognized concerns the use of the scale. The SPAF was designed as a measure of normal personality dimensions. As such, it was not designed to reveal or delineate psychopathological symptoms. it's major use was to provide information about personality to school psychologists, counselors, special education teachers or regular education teachers. In line with that, it has been used on an experimental basis as a kindergarten screening device in conjunction with other scales. A possible use, which has not been researched is as a screening instrument to indicate children who might need a more indepth evaluation using a more clinically oriented device such as the Personality Inventory for Children (PIC; Wirth & Lachar, 1981). Research is needed to determine the feasibility of this.

A positive aspect of this project was the technique used to develop the SPAF. There were several steps followed whch combined a theoretical and empirical approach. The

trait constructs were defined and then items were developed to measure the construct. Its' factorial validity was then analyzed, appropriate changes were made, and factorial validity was again determined. Other aspects of validity were then examined, including concurrent, multitrait-multimethod and predictive. Each of these steps are a necessary but not sufficient basis for the development of a personality scale.

One area of research which was noted from the sequential scale development process was indicated by the limited convergent and discriminant validity coefficients. It was apparent from this that the items need to be refined to identify salient features which are pertinent in more than one environment or to see if this is possible. A related topic would involve determining the effect of having the parent complete the form from the viewpoint of the child's school behavior. This might increase the ability of a parent rating to predict achievement.

Another area which needs more research is the relationship of personality patterns to achievement. The current project demonstrated that the SPAF could predict achievement moderately well but did not deal with which patterns would be indicative of impending or potential problems. Research dealing with this subject needs to be conducted. It must be noted, however, that this approach

might not yield definitive information since a child interacts with and affects the environment while the environment (including the teacher) affects the child. Therefore, personality patterns will probably need to be considered in relation to the environment.

A related area which needs to be investigated is what teaching settings and styles result in optimal academic progress for children exhibiting various personality patterns. The study by Rubin and Balow (1979) indicates the importance of this. If a child is considered a behavior problem by a teacher, which is likely to happen during the first six years of school, then that child is treated as a behavior problem for the better part of a year. This could easily have far reaching ramifications and effects on a child's later progress. A behavior rating scale would be useful for dealing with the problem after it occurs, whereas, the SPAF could be useful in identifying personality styles which fit with certain teaching styles or classroom environments. This could prevent the behavior problem from occurring. Research is needed, however, to determine which personality patterns mesh with which environments and teachers. Cluster analysis techniques could be used to determine this.

In conclusion, this project has demonstrated that personality variables can predict academic achievement. It has also demonstrated that a trait rating method can be used to measure those variables. It seems, then, that it is time for psychologists and other researchers to begin exploring how personality as measured by traits can be used to enhance education rather than ignoring the effect of personality on education. It is hoped that this project and the recommendations and conclusions stemming from it will help those research efforts.

- Carey, W. B. (1970). A simplified method for measuring infant temperament. The Journal of Pediatrics, 77, 188-194.
- Carey, W. B., Fox, M., & McDevitt, S. (1977). Temperament as a factor in early school adjustment. Pediatrics, 60, 621-624.
- Cattell, R. B. (1945). The principal trait clusters for describing personality. Psychological Bulletin, 42, 129-161.
- Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 140-161.
- Cattell, R. B., & Coan, R. A. (1957). Child personality structure as revealed in teachers ratings. Journal of Clinical Psychology, 13, 315-327.
- Coan, R. W., & Cattell, R. B. (1966). Early School Personality Questionnaire. Champaign, IL: Institute for Personality and Ability Testing.
- Comrey, A. L. (1970). Manual for the Comrey personality Scales. San Diego: Educational and Industrial Testing Service.

- Cox, B., & Sargent, H. (1950). TAT responses of emotionally disturbed and emotionally stable children: Clinical judgement versus normative data. Journal of Projective Techniques, 14, 61-74.
- Darlington, R. B., Weinberg, S. L., & Walberg, H. J. (1975). Canonical variate analysis and related techniques. In D. J. Amick & H. J. Walberg (Eds.), Introductory multivariate analysis. Berkeley, Ca: McCutchan.
- Digman, J. M. (1963). Principal dimensions of child personality as inferred from teacher's judgements. Child Development, 34, 43-60.
- Field, T., & Greenberg, R. (1982). Temperament ratings by parents and teachers of infants, toddlers and preschool children. Child Development, 53, 160-163.
- Fiske, D. W. (1975). A source of data is not a measuring instrument. Journal of Abnormal Psychology, 84, 20-23.

Flint, D.L., Hick, T.L., Horan, M.D., Irvine, D.J., & Kukak, S. E. (1980). Dimensionality of the California Preschool Social Competency Scale. Applied Psychological Measurement, 4, 203-212.

Garside, R. F., Birch, H., Scott, D. M., Chambers, S., Kolvin, I., Tweddle, E. G., & Barber, L. M. (1977). Dimensions of temperament in infant school children. In S. Chess & A. Thomas (Eds.), Annual progress in child psychiatry and child development. New York: Brunner/Mazel.

Hainsworth, P. K., & Sigueland, M. L. (1969). Early Identification of Children with Learning Disabilities: The Meeting Street School Screening Test. East Providence, RI: Meeting Street School.

Hubert, N. C., Wachs, T. D., Peters-Martin, P., & Gandour, M. J. (1982). The study of early temperament: Measurement and conceptual issues. Child Development, 53, 571-600.

Jackson, D. N. (1976). Jackson Personality Inventory: Manual. Port Huron, MI: Research Psychologists Press Inc.

- Kagan, J. (1963). Reflection, impulsivity and reading ability in primary grade children. Child Development, 36, 609-628.
- Kagan, J. (1982). The construct of difficult temperament: A reply to Thomas, Chess and Korn. Merrill-Palmer Quarterly, 28, 21-24.
- Kerlinger, F. N. (1973). Foundations of behavioral research (2nd Ed.). New York: Holt, Rinehart, & Winston.
- Korn, S. (1977). Temperament and academic achievement. In A. Thomas & S. Chess. Temperament and development. New York: Brunner/Mazel.
- McDevitt, S. C., & Carey, W. B. (1978). The measurement of temperament in 3 - 7 year old children. Journal of Child Psychology, Psychiatry, and Allied Disciplines, 19, 245-253.
- McGraw-Hill. (1982). The Comprehensive Tests of Basic Skills (Form V). Monterey, Ca: Author.
- Murray, H. A. (1943). Thematic Apperception Test. Cambridge: Harvard University Press.

- Mussen, P. H., Conger, J. J., & Kagen, J. (1974). Child development and personality (4th Edition). New York: Harper and Row.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner K., & Bent, D. H. (1975). SPSS: Statistical Package for the Social Sciences (2nd ed.). New York: McGraw-Hill.
- Norman, W. T., & Goldberg, L. R. (1967). Raters rates, and randomness in personality structure. Journal of Personality and Social Psychology, 4, 681-691.
- Plomin, R. (1982). The difficult concept of temperament: A response to Thomas, Chess and Korn. Merrill-Palmer Quarterly, 28, 25-33.
- Porter, R. B., & Cattell, R. B. (1959). Children's Personality Questionnaire. Champaign, Il: Institute for Personality and Ability Testing.
- Rothbart, M. K., & Derryberry, D. (1982). Theoretical issues in temperament. In M. Lewis & L. T. Taft (Eds.), Developmental disabilities: Theory, assessment, and intervention, New York: Spectrum Publications.

- Rubin, R. A., & Balow, B. (1979). Prevalence of teacher identified behavior problems: A longitudinal study. In S. Chess & A. Thomas (Eds.), Annual progress in child psychiatry and child development. New York: Brunner/Mazel.
- Rutter, M., & Graham, P. (1968). The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. British Journal of Psychiatry, 114, 563-579.
- Spivack, G., & Swift, M. S. (1966). The Devereux Elementary School Behavior Rating Scale: A study of the nature and organization of achievement related disturbed classroom behavior. Journal of Special Education, 1, 71-91.
- Spivack, G., & Swift, M. S. (1967). Devereux Elementary School Behavior Rating Scale. Devon, Pa: Devereux Foundation.
- Thomas, A. & Chess, S. (1977). Temperament and development. New York: Brunner/Mazel.
- Thomas, A., & Chess, S. (1980). The dynamics of psychological development. New York: Brunner/Mazel.

Thomas, A., Chess, S., & Birch, H. (1968). Temperament and behavior disorders in children. New York: New York University Press.

Thomas, A., Chess, S., & Birch, H. (1970). The origin of personality. Scientific American, 223.

Thomas, A., Chess, S., & Korn, S. (1982). The reality of difficult temperament. Merrill-Palmer Quarterly, 28, 1-20.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF ASD Technical Report, No. 61-53. (From Wiggins, J. S. (1973). Personality and prediction: Principles of personality assesment. Reading, Ma: Addison-Wesley.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 321-327.

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. Educational and Psychological Measurement, 38, 953-958.

Wiggins, J. S. (1973) Personality and prediction:
Principles of personality assessment. Reading Mass.:
Addison-Wesley.

Wirth, R. D., & Lachar, D. (1981). The Personality
Inventory for Children: Developmental and clinical
applications. In P. McReynolds (Ed.), Advances
in psychological assessment (Vol. 5). San Francisco:
Josey-Bass.

Zwick, W., & Velicer, W. F. (1982). Factors influencing
four rules for determining the number of components to
retain. Multivariate Behavioral research, 17,
253-269.

Appendix A

Limited Review of Representative Research

Limited Review of Representative Research

A major concern in education and psychology for the past several years has been the early identification of children who are at risk of developing learning or behavior problems. This has been demonstrated by state and federal requirements for kindergarten screening and Child Find programs. The implication of such programs is that a child who is at risk of developing learning or behavior problems can be identified by certain characteristics. It is also implied that learning or behavior problems can be prevented or reduced via appropriate intervention. This indicates that the problem is internal to the child or at least can be identified and altered at that level. The tests used to identify these children are typically considered as readiness or screening tests and can be quickly and easily administered. The content emphasis of most of these tests reflects a cognitive, perceptual or psychomotor orientation. A small number of tests are available with a behavioral orientation and even fewer with a personality orientation. The psychometric data on these tests vary from those for which it is not available to those for which it is adequate. However, there are few readiness or screening tests with a personality orientation which meet rigorous psychometric standards.

The following is a discussion of several studies pertaining to the predictive validity of a few representative tests. The discussion is divided according to the orientation of the test into three categories: (1) cognitive, (2) behavioral or (3) personality. This is not intended as a review of the readiness/screening literature. It is a reflection of some of the difficulties in early identification of children with learning problems.

Cognitively Oriented Instruments

Flook and Velicer (1977) examined the predictive validity of the Preschool Screening System (PSS; Hainsworth & Hainsworth, 1974) using teacher ratings of academic performance as the criterion measure. The PSS was chosen because the authors felt the tasks comprising the scale were representative of the types of variables included in most tests of school readiness. In that study, the PSS was administered to 245 kindergarten children. Two months later, the Rhode Island Pupil Identification Scale (RIPIS; Novack, Bonaventure, & Merenda, 1973) was completed by the five kindergarten teachers. The RIPIS is a teacher rating form for both behavior and written work. A multiple regression analysis was used to investigate the relation between the four subscores of the PSS (Information Processing, Draw-A-Person Concept, Draw-A-Person Control and

Verbal Reasoning), sex, and age in months and the composite score of the RIPIS. The multiple correlation coefficients ranged from .67 to .80, indicating good predictive validity. It must be recognized, however, that the criterion measure is not capable of predicting with complete accuracy to more objective measures of academic performance. A conclusion which appears warranted, then, is that the PSS is capable of predicting to a high degree, the results of the RIPIS and that the RIPIS is "moderately predictive of more objective measures of school performance" (Flook & Velicer, 1977, p. 140). This conclusion suggest that many of the children who are not ready for kindergarten and are going to develop academic problems will be identified by the PSS. However, it seems that many of them will also not be identified. This conclusion reflects the usual result with the screening/readiness tests currently available.

Another study which resulted in a more pessimistic conclusion was that by Flynn and Flynn (1978). They examined the relationship of five screening measures administered during kindergarten and a criterion measure administered during second grade. The screening instruments were: (1) the Slosson Intelligence Test (Slosson, 1963), (2) the Peabody Picture Vocabulary Test (Dunn, 1959), (3) the Goodenough-Harris Drawing Test (Harris, 1963), (4) the Developmental Test of Visual Motor Integration (Beery, 1967) and (5) the Metropolitan Readiness Test (Hildreth, Griffiths

& McGouvran, 1969). The criterion measure was the California Achievement Test (Tiegs & Clark, 1970). The data were analyzed using multiple regression techniques. Only one of the predictors was significantly related to the achievement test administered in the second grade. This was the Metropolitan Readiness Test. Even this relationship was not strong, however, accounting for only 10 percent of the variance. There was a significantly greater length of time between the administration of the predictor variables and the criterion measure in this study than in the study cited above (Flook & Velicer, 1977). It would be expected that the predictive ability of the screening measures would be reduced as a function of the greater length of elapsed time. However, these results suggest that the idea of screening for at risk children should be examined very carefully prior to the implementation of an intervention program.

This warning is also given by Satz and Fletcher (1979) in their examination of some common misinterpretations of results of screening tests for high risk children. They give an example of a misinterpretation or misleading conclusion of a comparison of two screening instruments through the use of detection tables. The two predictor variables in the article were the Student Rating Scale (SRS) and the deHirsch Predictive Index (HPI). These instruments were administered at the end of kindergarten and used to predict to the criterion measure of reading outcome at the

end of first grade. The children were divided into a group of good readers and a group of poor readers according to the results of the Gates-MacGinitie Reading Test. The data of the original study which was being critiqued indicated that the total hit rate (valid positive and valid negative) for the SRS was 77 percent whereas the HPI had only a 73 percent hit rate. The SRS had a 30 percent valid positive rate, a 70 percent false negative rate, a three percent false positive rate and a 97 percent valid negative rate. The HPI had a 26 percent valid positive rate, a 74 percent false negative rate, a seven percent false positive rate and a 93 percent valid negative rate. Satz and Fletcher (1979) felt that further analysis would be useful in determining the predictive utility of both instruments. They determined the conditional probability that a child would become a poor reader given that the SRS or HPI predicted that event. These data indicate that the SRS is a better predictor of reading problems with a probability of correct prediction being .83. The 70 percent false negative rate of the SRS indicates that if a child is predicted to be a poor reader there is a good probability that the child will be. However, there are going to be more children missed than found. The results with the HPI were worse. The authors suggest that these results show the advantages of screening instruments but also show that the instruments have little usefulness in identifying the majority of children who could benefit from an early intervention program.

The unifying factor in the three studies just reviewed is the difficulty they have in predicting which children are at risk of developing academic problems and who would, therefore, profit from an early intervention program. The problems discussed are found with most cognitively oriented screening/readiness tests available at this time. For these and other reasons, many researchers have attempted to approach from a behavioral orientation. Some of the difficulties encountered with this orientation are discussed next.

Behavioral Orientation

As a part of the research on early identification of children at risk of developing learning or behavior problems some efforts have focused on delineation of specific behaviors which affect academic functioning. This research has focused on several levels of observation. The following discussion reviews three. These were chosen because they are fairly representative of behavioral research. The first study (Hops & Cobb, 1973) used natural observation procedures to identify precise academic behaviors which enhanced learning. The ultimate goal was to increase those behaviors through specific behavioral techniques. The second study (Behar, 1977) describes the development of a behavior rating form to identify specific behaviors which indicate the development of learning or behavior problems.

The third method (Swift & Spivack, 1975) used factor analytic techniques to identify behavioral dimensions. Their goal was to help the teacher develop alternate teaching strategies.

The purpose of the Hops and Cobb (1973) study was to demonstrate that first grade students who were trained in specific survival skills which enhance learning would make greater gains in reading achievement than students not trained. The survival skills had been identified in earlier research by the authors. The skills were Attending, Volunteering, and Looking Around. In that study, the authors divided Attending into Attending and Work. The intervention involved training the teachers in behavioral techniques and operationally defining the target behaviors. The teachers used group contingency procedures to increase the occurrence of the Attending, Volunteering and Working minus the frequency of Looking Around divided by the total frequency of all behaviors. This indicated the proportion of survival skill behaviors to all behaviors. The assumption was that the higher the proportion the more the student would be learning. The second criterion measure was reading achievement as measured by the Gates-MacGinitie Reading Readiness and Primary A Tests.

The intervention program was designed to increase the mean proportion of survival skills. The program resulted in a significantly greater increase in the use of those skills in the experimental group compared to the control group. However, the difference in the mean proportion of survival skills emitted by both groups was only .03. One reason for this was that the mean proportion of survival skills in the control group was higher than the experimental group at the onset of the study. The Gates-MacGinitie test also showed a significantly greater increase for the experimental group. The difference in Standard scores on the post test was small with only .3 points difference between the two groups. The experimental group increased 11.9 points whereas the control group increased only 7.5 points. Although these figures are statistically significant, the practical significance must be questioned. The authors stated that more than one half hour of consult time per child was required. This amount of time appears to be less than optimally cost effective. There may also be variables in the classroom which can alter the proportion of survival skills without a direct intervention to increase those skills. One of those variables may be the teachers's level of awareness and style of interaction with the children. It can be concluded, however, that direct behavioral techniques may be effective but may not be as efficient as other techniques.

An example of a more efficient method is that developed by Behar (1977). A screening instrument was developed to identify preschool age children with apparent or developing behavior problems. The instrument, The Preschool Behavior Questionnaire, is a modification of the Children's Behavior Questionnaire (Rutter, 1967). The Pre-school Behavior Questionnaire is a 30 item rating scale to be completed by teachers using a three point scale. The points are: (1) does not apply, (2) applies sometimes and (3) frequently applies. The questionnaire was standardized on a sample of 598 children. Of these, 496 were children from regular school and 102 were from specialized treatment centers. The latter children had been previously diagnosed as emotionally or behaviorally disturbed. Analysis of the data from the standardization sample indicated that the instrument was able to discriminate between the two groups. A principal component analysis with Varimax rotation resulted in three components. They were labeled: (1) Hostile-Aggressive, (2) Anxious-Fearful and (3) Hyperactive-Distractible. To determine the reliability of the instrument, a new study was completed using 80 children from normal preschools and nine from a therapeutic school. Interrater reliability between teacher and aide ratings was .84 for the total test, .81, .71 and .67 for the three components. Test-retest reliability over a three to four month period was .87 for the overall scale, .93, .60 and .94 for the components. The scale was also able to discriminate

between the normal preschool group and therapeutic preschool group.

In summary, the Preschool Behavior Questionnaire demonstrated adequate reliability and the ability to discriminate between a normal and diagnosed behaviorally disturbed group. However, there is no information concerning the ability of the instrument to identify children in the preschool setting who will later develop school problems. It is possible that the instrument would demonstrate the same problems as other screening devices as discussed above. It would appear, then, that this instrument may be capable of discriminating between emotionally disturbed and normal children but more work must be done before this instrument will be useful in a preventative or predictive model. The research discussed next may be more useful in this respect.

Swift and Spivack (1975) took a different approach than the two discussed above. They helped teachers develop methods of altering behaviors in the classroom which detract from learning. As such, their focus was on more global behaviors than the survival skills described above and did not follow the "medical" model used by Behar (1973).

Swift and Spivack (1975) determined the prevalence of learning or behavior problems and found that a typical rate was 36 to 43 percent of the children between kindergarten and twelfth grade. This suggests that a significant proportion of a teacher's time may be spent in disciplining rather than teaching. They next identified specific behavioral dimensions which were related to learning. This was accomplished by having teachers rate a large sample of elementary school age children and then identifying the specific behavioral dimensions. The dimensions were: (1) inattentiveness and withdrawal, (2) external reliance and intellectual dependence, (3) poor rapport with the teacher, (4) impatient or unreflective, (5) lack of personal initiative and involvement, (6) irrelevant talk, (7) negative feelings and actions, (8) achievement anxiety, (9) restless, disturbing and (10) external blaming. The authors then described specific alternative teaching strategies which could be used to change the behavioral dimensions being observed. This is a useful method of identifying those children who are exhibiting behavioral problems which are interfering with their educational advancement and developing teaching styles which can change the problem behaviors.

However, one possibility not considered is a "personality" conflict between teacher and student. Personality in this context implies a fairly stable

behavioral style exhibited over time and across situations. There are several studies to support this (Olweus, 1980; Thomas & Chess, 1977; Thomas, Chess, & Birch, 1968). However, a study by Rubin and Balow (1979) seems especially germane. They gathered data on a group of 1586 children from kindergarten to sixth grade. Among the data gathered was whether the child was a behavior problem. In any single year, from 23 percent to 31 percent of the children were regarded as behavior problems which is similar to the level found by Swift and Spivack (1975). However, of interest was the fact that those children for whom complete data was available, i.e. six teacher ratings over the period of the study (n=439), 60 percent were considered a behavior problem by at least one teacher. This suggests that the norm is for a child to be considered a behavior problem by at least one teacher between kindergarten and sixth grade. Therefore, rather than looking for aberrant behavior in the children, consideration should be given to the idea of looking at personality variables and their effect on teacher/student interactions. Studies with this orientation will be discussed in the next section.

Personality Orientation

As pointed out above, another factor which may contribute to academic success or failure is personality. Some major problems arise when an attempt is made to examine

the effects of personality on success in school. One of the major problems is in defining "personality" and determining what view of personality to use in examining the factors involved in academic success or failure or behavior problems.

Monte (1977) discusses four theoretical approaches to personality: (1) Interpersonal, (2) Trait, (3) Psychoanalytic and (4) Social Learning (Behavioral). Magnusson and Endler (1976) would add a fifth, Interaction Theory. Maddi (1968) discusses personality theory as being expressed by three models: (1) Conflict, (2) Fulfillment and (3) Consistency. It appears that there is little agreement as to what constitutes the important aspects of personality. However, in reviewing the literature, there seems to be at least one common aspect to many of the various positions. Most of the statements about personality include a suggestion of the presence of stable behavioral styles or traits. Maddi (1968) describes the syndromes attributed to Sullivan who is described as an "interpersonal" theorist by Monte (1977). These syndromes imply the presence of consistent behavior over time and could be viewed as traits. Maddi (1968) also describes the character types comprised of traits inherent in Freud's theory of personality. Magnusson and Endler (1976) in promoting an interactionist view of personality use measured traits on the person side of the person-situation

interaction. Mischel (1968) is often quoted as sounding the death knell for trait theory, yet he has altered his position in a later article (Mischel, 1979) to state that trait categorizations can have certain advantages and are useful for a variety of purposes.

A conclusion which may be reached from this discussion is that trait theory has been and still is extremely important in the development of personality theory. However, screening instruments have seldom been used to examine the effect of personality traits on school functioning. The interest in the relationship of personality and academic progress has been increasing which has resulted in an increase in research. In line with that, research on personality for pre-school and lower grade school children has relied on self-report as well as alternative methods such as: (1) interview of the subject (2) interview of persons knowledgeable of the subject, (3) projective techniques conducted with the subject, or (4) questionnaires completed by persons knowledgeable of the subject. Examples of these approaches will serve to illustrate the advantages and disadvantages of each.

An example of the self report approach is the scales developed by Cattell and his associates (Coan and Cattell, 1966; Porter and Cattell, 1959) such as the Children's Personality Questionnaire (CPQ) and Early School Personality

Quiz (ESPQ). These instruments are completed by the children and demonstrate adequate test-retest reliability. However they are rather lengthy (60 to 100 minutes in two sessions for the ESPQ) and have little established external validity (Buros, 1978). A general difficulty with this approach is that the verbal skills limit the type of items that can be employed.

An example of an approach using direct interview of the subject is that conducted by Rutter and Graham (1968). They used a standardized interview format with nonclinic children and classified them as showing no psychiatric abnormality, some abnormality or definite psychiatric abnormality. Using this classification format resulted in good overall agreement between raters ($r=.84$). However, the agreement on specific items was much lower. Another part of that study looked at test-retest reliability with psychiatric patients and found the agreement on overall psychiatric state with four weeks between interviews to be .61. Again, agreement on specific items was much lower. These results indicate that moderately reliable and valid classifications as to gross psychiatric condition can be made with standardized interview procedures. However, there is less agreement on clinical inferences which could be useful in determining diagnosis and treatment. This process is also inefficient, requiring a significant amount of professional time.

An example of a method using an indirect interview of persons knowledgeable of the subject was that developed by Garside, Birch, Scott, Chambers, Kolvin, Tweddle, and Barber (1976) for use with parents of young children. The technique used a semi-structured, open-ended interview conducted with the mother. The results were then converted to scores on scales which had been previously defined. The emphasis of the interview was on "how" the child interacted with his environment across situations to produce data on temperament. The interview was conducted two times at approximately a one month interval. Principal component analysis resulted in four main components across the two administrations. They were labeled: (1) withdrawal, (2) assertiveness, (3) high activity, and (4) negative mood. Inter-rater reliability on these dimensions was good, ranging from .92 to .98. However, test-retest reliability was not as good. An interview conducted with the same parents approximately one month later showed that "about 50 percent of the 39 items reliabilities were above 0.8 and 90 percent above 0.06" (Garside et al, 1976, pg 38). This procedure is also inefficient, again requiring a significant amount of professional time.

Projective techniques have been used for many years in the examination of personality with both adults and children. An example of this technique is the Thematic Apperception Test (TAT; Murray, 1943) which is widely used

with both children and adults. With the TAT, the subject is asked to make up stories about pictures depicting various situations. These stories purportedly reveal needs or motives of the person being tested. Advantages of this technique are that the individual is allowed to reveal the underlying personality structures which reflect his organizational patterns and they are simple to administer. A major problem with this and other projective techniques is the limited reliability and validity of the method created by the difficulty interpreting the results (Achenbach, 1974). For example, a study by Cox and Sargent (1950) resulted in 11 of 15 normal subjects being classified as emotionally disturbed when their protocols were evaluated by experienced clinicians.

An alternative approach that avoids the inefficiency problems as well as the difficulty produced by the limited language development of young children is the utilization of a rating scale completed by an individual familiar with the subject. Carey (1970) developed a questionnaire to be completed by parents which was designed to assess the temperamental styles of infants between the ages of four and eight months. The temperamental categories used were from the New York Longitudinal Study (Thomas, Chess and Birch, 1968). Test-retest reliability data was collected with an extremely small sample ($N=3$) and no reliability coefficients were given. Much work needs to be completed on this

questionnaire before it is used extensively. It is also not applicable to the preschool and early school age child.

A questionnaire which was developed for children between the ages of three and seven years by McDevitt and Carey (1978). This instrument, the Behavioral Style Questionnaire (BSQ), is completed by parents. The authors report test-retest reliability for the complete instrument as being .89. The coefficients for the individual categories of temperament ranged from .67 to .94. Internal consistency coefficients ranged from .47 to .80. From this information, the BSQ appears to be adequately reliable, but is lacking internal consistency in some categories. No construct validity information is available..

Another instrument which could be used in screening is the Teacher Temperament Questionnaire which was based on the New York Longitudinal Study (NYLS; Thomas, Chess and Birch, 1968) The NYLS utilized three techniques of assessment. Data collection was begun in 1956 and has been continued to the present. It has been gathered from interviews with parents and teachers, observations of the children and from objective testing. Qualitative analysis of the data resulted in the delineation of nine temperamental categories of reactivity or traits. The categories were: (1) Activity Level, the motor component present in a given child's functioning; (2) Rhythmicity (regularity), the

predictability of a child; (3) Approach or Withdrawal, the nature of the initial response to a new stimulus; (4) Adaptability, the ease with which a child's responses are modified in the desired direction; (5) Threshold of Responsiveness, the level of stimulation necessary to evoke a response; (6) Intensity of Reaction, the energy level of a response; (7) Quality of Mood, the amount of pleasant behavior contrasted with unpleasant behavior; (8) Distractibility, how effective external stimuli are in interfering with ongoing behavior; and (9) Attention Span and Persistence, two related categories. Attention Span concerns the length of time a child stays with an activity. Persistence refers to the continuation of an activity in the face of obstacles to the maintenance of the activity. From these categories three temperamental constellations were identified. First was the Easy Child, who showed regularity, positive approach to new stimuli, high adaptability and a mild or moderately intense mood which is usually positive. Second was the Difficult Child, who showed irregularity, negative withdrawal responses to new stimuli, non-adaptability or slow adaptability, and intense mood expressions which were often negative. The third behavioral style was the Slow-to-warm-up Child, who showed negative responses of mild intensity to new stimuli and slowly adapted after repeated contact.

In conjunction with the study just described, the NYLS research group developed a questionnaire (Thomas & Chess, 1977) for use with children from three to seven years of age which can be completed by teachers in a short time. The Teacher Temperament Questionnaire (TTQ) was designed to delineate eight of the nine categories of temperament described in the NYLS research. Rhythmicity is not identified because it requires knowledge of the child's functioning over a 24 hour period. The questionnaire was developed originally for parents by culling items of behavior from the NYLS interview protocols which seemed to be typical and unambiguous and did not overlap with other categories of temperament. These items were then given to twenty mothers with three to seven year old children and were then revised using comments and suggestions made by these mothers. The questionnaire was reduced to 132 questions using this procedure. The TTQ was developed by modifying these questions and making them applicable to the school setting. Some items were dropped as totally inappropriate. Sixty questionnaires were then completed by a group of teachers. The results were reviewed and refined to determine whether the item should be included in the final form which was reduced to eight items for each of the eight traits.

Limited data exists concerning the reliability or validity of this instrument. In the development of the parent questionnaires, fifty mothers were asked to complete the questionnaire and were then interviewed using the NYLS interview format two weeks later. Another fifty mothers were interviewed and then completed the questionnaire two weeks later. The data generated were used to reduce the parent questionnaire from 132 items to 72 by keeping those items which showed significant correlation between the interview and the question. These data were not generated for the TTQ as there was no interview completed. The questions for the TTQ were reviewed by four judges who sorted each item into categories defined by the temperamental trait. Items were deleted if the four judges did not agree on category. This provides minimal content validity. The only other reliability study reported for the TTQ was Hubert, Wachs, Peters-Martin, and Gandour (1982) which cited unpublished data and indicated that reliabilities for the categories ranged from .69 to .81. A recent concurrent validity study reported low- moderate correlations ($r = .18$ to $.46$) between the categories of the TTQ and other instruments with similar categories (Billman and McDevitt, 1980).

Research by Baker and Velicer (1982) examined the reliability and construct validity of the TTQ. Construct validity was assessed by determining to what extent the

eight theoretical scales could be reproduced empirically. The sample (N=116) was made up of kindergarten, first and second grade children in a southeastern Massachusetts town. Their teachers were asked to complete a TTQ for each child in their class. Four weeks later, they were asked to complete a second TTQ. A principal component analysis (PCA) with Varimax rotation was performed on both administrations. There were four stable components identified from each administration. They were labeled: (1) Compliance, which involved attentiveness, distractibility and persistence, (2) Extroversion, which involved shyness, comfort in novel situations and social interactions, (3) Interpersonal Affect, which involved competitiveness, cooperation with others and a tendency to be argumentative and (4) Environmental Sensitivity, which involved light and temperature sensitiveness and responsiveness to stories. Each of these components was composed of a mix of items from the theoretical scales of the NYLS (Thomas and Chess, 1977).

Test-retest reliability coefficients were computed for each of the four components. The results were: Component 1: .76, Component 2: .73, Component 3: .75, and Component 4: .86. Internal consistency was also determined, resulting in Alpha Coefficients of .97, .89, .88 and .45 for Components 1, 2, 3, and 4 respectively for the first administration and .97, .87, .85, and .76 for the second administration. The results of this study, then, did not

support the theoretical scales of Thomas and Chess (1977) but did suggest that an alternative method of scoring could be used to produce scales with adequate reliability and validity.

Besides the analysis of the two administrations, Baker and Velicer (1982) presented a partial replication using a shortened version of the TTQ. This version included fifteen items from each of the previously described scales of Compliance, Interpersonal Affect and Extroversion. Five teachers rated 155 children as a part of a kindergarten screening. A principal component analysis with Varimax rotation resulted in four components which accounted for 65 per cent of the variance. The first component was composed of eighteen items, fourteen of which were from the Compliance scale. The second component was made up of twelve items, eleven of which were from the Interpersonal Affect scale. The third component included nine items, all from the Extroversion scale. The fourth component was composed of four items, three from the Extraversion scale and one from the Interpersonal Affect scale. As can be seen from these results, 34 of the 45 items were included on the same scale as the original study. Two of the remaining eleven items loaded on more than one component and can be excluded. Two items did not load on any component and four items were assigned to the fourth component. The other three items loaded on different components than those

predicted from the original study.

It appears then, that the questionnaire developed by the NYLS project is capable of identifying stable personality traits. However, the research by Baker and Velicer (1982) indicated that the TTQ was delineating three clearly identified traits and a fourth which was less clearly identified rather than the eight hypothesized scales. The four components were made up of a mix of items drawn from the theoretical scales and demonstrated good test-retest reliability. They were labeled Compliance, Extroversion, Interpersonal Affect, and Environmental Sensitivity. This research indicated that the structure of child personality being identified by this instrument was complex and perhaps not limited to the four dimensions just described.

It appears from these results and from reviewing other studies that there is no personality screening instrument available which meets the criteria stated at the beginning of this paper. Namely: (1) ease of administration and scoring and (2) good psychometric properties. There does appear to be a need for one however.

Conclusion

It was pointed out in the beginning of this paper that a major concern in education and psychology has been the early identification of children who currently exhibit or who are at risk of developing learning or behavior problems. Three theoretical orientations were discussed: (1) cognitive, (2) behavioral and (3) personality. Representative examples of research on instruments or techniques available indicated shortcomings in all three areas. There are cognitively oriented instruments available which have adequate reliability and validity yet miss a large percentage of the high risk children. There are behavioral techniques available which increase the occurrence of educational survival skills but are inefficient. There is research indicating that a child's personality can affect academic functioning but there are no instruments currently available which have been thoroughly researched and are also easily administered.

What appears to be needed at this time is an instrument which can identify stable behavioral styles and can demonstrate adequate psychometric properties. The purpose of the instrument would be to help the teacher learn more appropriate teaching styles and techniques for the individual children rather than identifying the high risk

child specifically. The high risk child may still be identifiable with such an instrument but this would not be its primary goal. An instrument which may meet this need is described in this project.

Appendix B

Student Personality Assessment Form (SPAF)

(Ninety Item)

Student Personality Assessment Form

1. Child tends to return to interrupted activities.
2. Child is shy with unfamiliar adults.
3. Child will accept a substitute if the work or play activity is unavailable.
4. Child does not comment about how pretty or ugly colors are.
5. Child says things to make other children feel bad.
6. Child keeps a messy desk.
7. Child gives first answer thought of when asked a question.
8. Child is easily distracted from work by noises or movements.
9. Child tells stories of personal experiences with enthusiasm and excitement.
10. Child shows little response if another child takes a toy or other item.
11. Child can be talked out of a bad mood easily.
12. Child considers others feelings before acting.
13. Child is late to school.
14. Child considers questions carefully before answering.
15. Child becomes impatient and goes to another task if the original task is difficult to understand or do.
16. Child prefers to watch new activities rather than participate.
17. Child argues with other children.

18. Child is aware of or comments about the class being hot or cold.
19. Child fights with other children.
20. Child likes to follow the same routine daily.
21. Child has trouble listening to directions.
22. Child will persist at one activity for a long period of time if allowed.
23. Child enjoys doing tasks for the teacher.
24. Child develops a good relationship with other children even if they originally did not get along well.
25. Child is very sensitive to bright light.
26. Child is able to settle disagreements without fighting.
27. Child is very careful about where his/her belongings are placed.
28. Child can sit for a short time without something to do.
29. Child starts but does not finish class work.
30. Child overcomes reluctance to participate in new activities very quickly.
31. Child becomes upset when another child wins a game.
32. Child listens attentively when stories are told or read.
33. Child breaks toys.
34. Child turns in neat work.
35. Child fidgets or wiggles even when involved in an activity.
36. Child can sit quietly for a longer period of time than others in class.

37. Child adapts slowly to changes in physical location such as seating arrangements, etc.
38. Child becomes overly upset in stressful situations.
39. Child is very aware of changes in the level of light in the room (brighter or dimmer).
40. Child is able to use breakable items without breaking them.
41. Child is interested only in turning paper in, not what it looks like.
42. Child stays with a task only for a short length of time before switching to another.
43. Child is not disturbed by extraneous sounds or other students when working.
44. Child readily becomes involved with new activities and situations.
45. Child complains about other children.
46. Child does not appear to be affected by special events or unusual situations such as field trips or visitors.
47. Child bullies other children.
48. Child finishes work before it is due.
49. Child is able to consider alternatives before acting.
50. Child can attend to what the teacher is saying even when other children are being noisy.
51. Child adjusts slowly to new situations.

52. Child communicates disagreement to other children by fighting and yelling.
53. Child notices new or unusual clothes worn by others.
54. Child is able to play without forcing others to do as he/she wishes.
55. Child seldom forgets things.
56. Child has trouble waiting to be called on to answer questions.
57. Child can be switched to other activities easily.
58. Child does not mind telling stories or doing other activities in front of the class.
59. Child becomes upset or angry when not allowed to engage in a desired activity.
60. Child is not concerned about minor cuts or bruises.
61. Child kicks or hits other children.
62. Child has trouble keeping track of hats or gloves.
63. Child does not talk out of turn in class.
64. Child changes activities often during free play periods.
65. Child is shy with children who are unfamiliar.
66. Child takes a loss in a competitive situation lightly.
67. Child is not disturbed by loud sounds such as bells ringing.
68. Child is careful to not hurt others.
69. Child is usually behind in school work.
70. Child has trouble making up his mind.
71. Child is inattentive and absent minded.

72. Child will share toys with others.
73. Child begins tasks without planning.
74. Child gives irrelevant answers to questions.
75. Child does not seem to be concerned about hurting himself when playing active outdoor games.
76. Child likes to do very few things.
77. Child needs a neat area to work well.
78. Child is cautious when doing something which could be harmful such as swinging or climbing on bars.
79. Child adapts easily to new or different ways of doing tasks.
80. Child can be trusted.
81. Child sometimes misses recess because of incomplete work.
82. Child's answers are usually accurate.
83. Child is distracted if brushed or touched even lightly by another child.
84. Child rejects hugs or other physical affection.
85. Child hands in incomplete work.
86. Child usually makes good decisions.
87. Child often complains of not feeling well.
88. Child teases other children cruelly.
89. Child is annoyed by messy work even if it is another child's.
90. Child is often first to finish papers.

Appendix C

Student Personality Assessment Form
Instruction Sheet (Ninety Item SPAF)

Student Personality Assessment Form

This questionnaire is designed to look at how a child behaves in various school situations. Each statement describes something which a child could be doing in school. You are to decide whether that behavior occurs hardly ever, infrequently, once in a while, sometimes, often, very often, or almost always. This decision can be indicated by circling the appropriate number from 1 to 7, where 1 indicates "hardly ever" and 7 indicates "almost always". The statements will usually involve making a judgement. These judgements should be based on how the child compares to other children. For example, when deciding the appropriate response to "Child will share toys", the amount of sharing compared to other children in that grade should be considered. The total frequency of "sharing" could be different when comparing a child in kindergarten and a child in second grade. Yet both children could be considered to share "often" when compared to their peers.

If you feel that some choice requires more explanation, please circle the response which seems to fit best and then write a note on the back of the last page of the questionnaire.

Appendix D

Student Personality Assessment Form (SPAF)
(Forty Three Item)

Student Personality Assessment Form (SPAF)

1. Child tends to return to interrupted activities.
2. Child will accept a substitute if the work or play activity desired is unavailable.
3. Child is shy with unfamiliar adults.
4. Child can be talked out of a bad mood easily.
5. Child is easily distracted from work by noises or movement.
6. Child argues with other children.
7. Child prefers to watch new activities rather than participate.
8. Child comments about the class being hot or cold.
9. Child becomes impatient and goes to another task if the original task is difficult to understand or do.
10. Child becomes upset when another child wins a game.
11. Child overcomes reluctance to participate in new activities very quickly.
12. Child is sensitive to bright light.
13. Child will persist at one activity for a long period of time if allowed.
14. Child becomes overly upset in stressful situations.
15. Child adapts slowly to changes in physical location such as seating arrangement, etc.
16. Child is aware of changes in the level of light in the room (brighter or dimmer).
17. Child starts but does not finish class work.

18. Child complains about other children.
19. Child readily becomes involved with new activities and situations.
20. Child notices new or unusual clothes worn by others.
21. Child fidgets or wiggles even when involved in an activity.
22. Child communicates disagreement to other children by fighting and yelling.
23. Child adjusts slowly to new situations.
24. Child enjoys doing tasks for the teacher.
25. Child can attend to what the teacher is saying even when other children are being noisy.
26. Child becomes upset or angry when not allowed to engage in a desired activity.
27. Child is shy with children who are unfamiliar.
28. Child reports accidents or rule infractions to the teacher.
29. Child is easily sidetracked.
30. Child takes a loss in a competitive situation lightly.
31. Child enjoys telling stories or doing other activities in front of the class.
32. Child can follow verbal instructions.
33. Child stays with a task for only a short time before switching to another.
34. Child says things to make other children feel bad.
35. Child prefers familiar toys and games to new play.
36. Child can explain his needs or desires to others.

37. Child finishes work before it is due.
38. Child bullies other children.
39. Child plunges into new activities and situations without hesitation.
40. Child gives irrelevant answers to questions.
41. Child kicks or hits other children.
42. Child misses recess because of incomplete work.
43. Child teases other children cruelly.

Appendix E

Student Personality Assessment Form (SPAF)

Instruction Sheet (Forty Three Item SPAF)

Student Personality Assessment Form

This questionnaire is designed to look at how a child behaves in various school situations. Each statement describes something which a child could be doing in school. You are to decide whether that behavior occurs: Never, Rarely, Occasionally, Often, or Always. This decision can be indicated by circling the appropriate number from 1 to 5, where 1 indicates "Never" and 5 indicates "Always". "Always" means that without exception the statement describes the child's behavior. "Frequently" means that repeatedly or periodically, but not always, the statement describes the child's behavior. "Sometimes" means that once in a while, but not frequently, the statement describes the child's behavior. "Rarely" means that seldom or infrequently the statement describes the child's behavior. "Never" means on no occasion does the statement describe the child's behavior.

The statements will usually involve making a judgement. These judgements should be based on how the child compares to other children. For example, when deciding the appropriate response to "Child will share toys", the amount of sharing compared to other children in that grade should be considered. The total frequency of "sharing" could be different when comparing a child in kindergarten and a child

in second grade. Yet both children could be considered to share "often" when compared to their peers.

If you feel that some choice requires more explanation, please circle the response which seems to fit best and then write a note on the back of the last page of the questionnaire.

Thank You

Appendix F

Parent Student Personality Assessment Form (ParSPAF)
(Forty Three Item)

Parent Student Personality Assessment Form

1. Child returns to interrupted activities.
2. Child will accept substitute if the work or play activity desired is unavailable.
3. Child is shy with unfamiliar adults.
4. Child can be talked out of a bad mood easily.
5. Child is distracted from work by noises or movement.
6. Child argues with other children.
7. Child prefers to watch new activities rather than participate.
8. Child comments about the house being hot or cold.
9. Child becomes impatient and goes to another task if the original task is difficult to understand or do.
10. Child becomes upset when another child wins a game.
11. Child overcomes reluctance to participate in new activities very quickly.
12. Child is sensitive to bright light.
13. Child will persist at one activity for a long period of time if allowed.
14. Child becomes overly upset in stressful situations.
15. Child adapts slowly to changes in physical location such as a new house, moving his/her bed, etc.
16. Child is aware of changes in the level of light in the room (brighter or dimmer).
17. Child starts but does not finish work.
18. Child complains about other children.
19. Child readily becomes involved with new activities and

situations.

20. Child notices new or unusual clothes worn by others.
21. Child fidgets or wiggles even when involved in an activity.
22. Child communicates disagreement to other children by fighting and yelling.
23. Child adjusts slowly to new situations.
24. Child enjoys doing tasks for you.
25. Child can listen to what you are saying even when other things (TV or other children) are making noise.
26. Child becomes upset or angry when not allowed to engage in a desired activity.
27. Child is shy with children who are unfamiliar.
28. Child reports accidents or inappropriate behavior to you.
29. Child is easily sidetracked.
30. Child takes a loss in a competitive situation lightly.
31. Child enjoys telling stories or doing other activities in front of people.
32. Child can follow verbal instructions.
33. Child stays with a task only for a short length of time before switching to another.
34. Child says things to make other children feel bad.
35. Child prefers familiar toys and games to new play.
36. Child can explain his needs or wants to other children.
37. Child finishes work around the house quickly.
38. Child bullies other children.

39. Child plunges into new activities and situations without hesitation.
40. Child gives irrelevant answers to questions.
41. Child kicks or hits other children.
42. Child must be reminded to complete work around the house.
43. Child teases other children cruelly.

Appendix G

Parent Student Personality Assessment Form (ParSPAF)

(ParSPAF Cover letter)

Dear Parent,

My name is E. H. Baker. I am a school psychologist with the Coventry School District and a PhD (Doctoral) student at the University of Rhode Island. Enclosed is a personality questionnaire which I would like for you to complete. This questionnaire is a part of a study being conducted for my doctoral dissertation and is not being paid for by the school system. It is also not a part of my job. An explanation of the study is given in the rest of this letter and I will be available to answer questions concerning the questionnaire and the project in general on Tuesday, July 27th between 10:00 A.M. and 1:00 P.M. If you have any questions call the Coventry Special Education Offices (Ph: 828-6114) and ask for E. H. Baker.

I am asking some of the parents of kindergarten, first, and second grade students to complete the questionnaire. The parents being asked to complete the questionnaire have been chosen randomly from the total group. I intend to use the information gathered to determine if parents and teachers see children's personality in the same way. The information will not be looked at on an individual basis. Rather, it will be analyzed using statistical techniques such as Factor Analysis and Multiple Regression Analysis.

Past research with teacher completed questionnaires suggest that three traits are measured. The first, Compliance/Organization, looks at a behavioral style which contributes to success in school. The second, Interpersonal Affect, looks at how well the child gets along with other children. The third, Introversion/Extroversion, looks at shyness and how quickly the child adjusts to new situations.

I want to see whether teachers and parents view children's personality in the same way for two reasons. The first is because research and my own experience in the schools suggests that intelligence plays only a small part in determining success in school (about 25 per cent). Personality appears to contribute nearly the same amount. If we can get an idea of how a child functions in school because of his/her personality the teaching method may be changed to fit the child's personality thus increasing learning.

The second reason is to increase knowledge of children's personality and how it affects school functioning. There is a very limited amount of research being done on personality in children and I feel more needs to be done. If teachers can learn to teach more effectively because of knowledge of children's personality, then children will be more comfortable in the learning process and will be able to learn more.

If you would like to receive a summary of the results of this study, check the box at the top of the questionnaire marked "Summary". Again, if you have questions regarding the study, please call me at 828-6114 on July 27th. If you can't call during that time, I can be reached at 789-3726 during the summer.

Thank you

E. H. Baker, MS
School Psychologist

Appendix H

Parent Student Personality Assessment Form (ParSPAF)
(ParSPAF Instruction Sheet)

Parent Student Personality Assessment Form

This questionnaire is designed to look at how a child behaves in various home situations. Each statement describes something which a child could be doing at home. You are to decide whether that behavior occurs: Never, Rarely, Occassionally, Often, or Always. This decision can be indicated by circling the appropriate number from 1 to 5, where 1 indicates "Never" and 5 indicates "Always". These alternatives are described as follows:

1. "Never" means on no occasion does the statement describe the child's behavior.
2. "Rarely" means that seldom or infrequently the statement describes the child's behavior.
3. "Sometimes" means that once in a while, but not frequently, the statement describes the child's behavior.
4. "Frequently" means that repeatedly or periodically, but not always, the statement describes the child's behavior.
5. "Always" means that without exception the statement describes the child's behavior.

The statements will usually involve making a judgement. These judgements should be based on how your child compares to other children his/her age. For example, when deciding the appropriate response to "Child will share toys", the amount of sharing compared to other children that age should be considered. The total amount of sharing could be different when comparing different age children.

If you feel that some choice requires more explanation, please circle the response which seems to fit best and then write a note on the back of the last page of the questionnaire.

Thank You

Appendix I

Parent Student Personality Assessment Form (ParSPAF)
(ParSPAF Follow-up Letter)

Dear Parent,

My name is E. H. Baker. I am a school psychologist for Coventry School District and a University of Rhode Island PhD student. I sent you a letter in June which contained a personality questionnaire and an explanation of the research I am doing.

I am sending this letter because I noted that I hadn't received a completed personality questionnaire from you. I have enclosed another questionnaire just in case the first one was misplaced or has been lost in the mail. I feel that this research will be helpful in answering some questions about the relationship between personality and education. Therefore, I hope that you will complete the questionnaire and return it to me.

If you have any questions please give me a call at 789-3726.

Thank you

E. H. Baker, MS

School Psychologist

Appendix J

Bartlett's Test for Remaining Eigen Values

Appendix J

Bartlett's Test for Remaining Eigen Values

Eigenvalue	Canonical correlation	Chi ²	DF	Tail probability
--	--	341.13	33	.00
.858	.926	166.75	20	.00
.749	.865	43.18	9	.00
.383	.618			

Bibliography

Bibliography

- Achenbach, T. M. (1974). Developmental psychopathology.
New York: Ronald Press.
- Baker, E. H., & Velicer, W. F. (1982). The structure and reliability of the Teacher Temperament Questionnaire. Journal of Abnormal Child Psychology, 10, 531-546.
- Bates, J. E. (1980). The concept of difficult temperament. Merrill-Palmer Quarterly, 26, 299-319.
- Beery, K. E. (1967). Administration and Scoring the Developmental Test of Visual Motor Integration.
Chicago: Follett.
- Behar, L. B. (1977). The preschool behavior questionnaire. Journal of Abnormal Child Psychology, 5, 265-275.
- Billman, J., & McDevitt, S. (1980). Convergence of parent and observer ratings of temperament with observations of peer interaction in nursery school. Child Development, 51, 395-400.

Buros, O. K. (1978). The eighth mental measurements yearbook. Highland Park, NJ: Gryphon Press.

Campbell, D. T., & Fiske, D. W. (1967). Convergent and discriminant validation by the multitrait-multimethod matrix. In D. Jackson & S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill.

Carey, W. B. (1970). A simplified method for measuring infant temperament. The Journal of Pediatrics, 77, 188-194.

Carey, W. B., Fox, M., & McDevitt, S. (1977). Temperament as a factor in early school adjustment. Pediatrics, 60, 621-624.

Cattell, R. B. (1945). The principal trait clusters for describing personality. Psychological Bulletin, 42, 129-161.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 140-161.

Cattell, R. B., & Coan, R. A. (1957). Child personality structure as revealed in teachers ratings. Journal of Clinical Psychology, 13, 315-327.

- Coan, R. W., & Cattell, R. B. (1966). Early School Personality Questionnaire. Champaign, Il: Institute for Personality and Ability Testing.
- Comrey, A. L. (1970). Manual for the Comrey personality Scales. San Diego: Educational and Industrial Testing Service.
- Cox, B., & Sargent, H. (1950). TAT responses of emotionally disturbed and emotionally stable children: Clinical judgement versus normative data. Journal of Projective Techniques, 14, 61-74.
- Darlington, R. B., Weinberg, S. L., & Walberg, H. J. (1975). Canonical variate analysis and related techniques. In D. J. Amick & H. J. Walberg (Eds.), Introductory multivariate analysis. Berkeley, Ca: McCutchan.
- Digman, J. M. (1963). Principal dimensions of child personality as inferred from teacher's judgements. Child Development, 34, 43-60.
- Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service.

- Field, T., & Greenberg, R. (1982). Temperament ratings by parents and teachers of infants, toddlers and preschool children. Child Development, 53, 160-163.
- Fiske, D. W. (1975). A source of data is not a measuring instrument. Journal of Abnormal Psychology, 84, 20-23.
- Flint, D.L., Hick, T.L., Horan, M.D., Irvine, D.J., & Kukak, S. E. (1980). Dimensionality of the California Preschool Social Competency Scale. Applied Psychological Measurement, 4, 203-212.
- Flook, W. M., & Velicer, W. F. (1977). School readiness and teacher ratings: A validation study. Psychology in the Schools, 14, 140-146.
- Flynn, T. M., & Flynn, L. A. (1978). Evaluation of the predictive ability of five screening measures administered during kindergarten. Journal of Experimental Education, 46, 65-70.

- Garside, R. F., Birch, H., Scott, D. M., Chambers, S., Kolvin, I., Tweddle, E. G., & Barber, L. M. (1977). Dimensions of temperament in infant school children. In S. Chess & A. Thomas (Eds.), Annual progress in child psychiatry and child development. New York: Brunner/Mazel.
- Hainsworth, P. K., & Sigueland, M. L. (1969). Early Identification of Children with Learning Disabilities: The Meeting Street School Screening Test. East Providence, RI: Meeting Street School.
- Harris, D. (1963). Goodenough-Harris drawing test manual. Chicago: Harcourt Brace.
- Hildreth, G. H., Griffiths, N. L., & McGouvran, M. E. Metropolitan readiness tests. New York: Harcourt, Brace, Jovanovich.
- Hops, H., & Cobb, J. A. (1973). Survival behaviors in the educational setting: Their implications for research and intervention. In L. A. Hamerlynk, L. C. Handy, & E. J. Mash (Eds.) Behavior change: Methodology, concepts and practice. Champaign, IL: Research Press.

- Hubert, N. C., Wachs, T. D., Peters-Martin, P., & Gandour, M. J. (1982). The study of early temperament: Measurement and conceptual issues. Child Development, 53, 571-600.
- Jackson, D. N. (1976). Jackson Personality Inventory: Manual. Port Huron, MI: Research Psychologists Press Inc.
- Kagan, J. (1963). Reflection, impulsivity and reading ability in primary grade children. Child Development, 36, 609-628.
- Kagan, J. (1982). The construct of difficult temperament: A reply to Thomas, Chess and Korn. Merrill-Palmer Quarterly, 28, 21-24.
- Kerlinger, F. N. (1973). Foundations of behavioral research (2nd Ed.). New York: Holt, Rinehart, & Winston.
- Korn, S. (1977). Temperament and academic achievement. In A. Thomas & S. Chess. Temperament and development. New York: Brunner/Mazel.

- Maddi, S. R. (1968). Personality theories: A comparative analysis. Homewood, IL: Dorsey Press.
- Magnusson, D., & Endler, N. S. (1977). Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.) Personality at the crossroads. Hillsdale, NJ: Erlbaum.
- McDevitt, S. C., & Carey, W. B. (1978). The measurement of temperament in 3 - 7 year old children. Journal of Child Psychology, Psychiatry, and Allied Disciplines, 19, 245-253.
- McGraw-Hill. (1982). The Comprehensive Tests of Basic Skills (Form V). Monterey, Ca: Author.
- Mischel, W. (1968). Personality and Assessment. New York: John Wiley.
- Mischel, W. (1979). On the interface of cognition and personality: Beyond the person-situation debate. American Psychologist, 34, 740-754.
- Monte, C. F. (1977). Beneath the mask: An introduction to theories of personality. New York: Praeger Publishers.

Murray, H. A. (1943). Thematic Apperception Test.
Cambridge: Harvard University Press.

Mussen, P. H., Conger, J. J., & Kagen, J. (1974). Child development and personality (4th Edition). New York: Harper and Row.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner K., & Bent, D. H. (1975). SPSS: Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill.

Norman, W. T., & Goldberg, L. R. (1967). Raters rates, and randomness in personality structure. Journal of Personality and Social Psychology, 4, 681-691.

Novack, H. S., Bonaventura, E. R., & Merenda, P. F. (1973).

A scale for early detection of children with learning problems. Exceptional Children, 40, 98-104.

Olweus, D. (In press). The consistency issue in personality

psychology revisited-With special reference to aggression. British Journal of Social and Clinical Psychology.

Plomin, R. (1982). The difficult concept of temperament: A response to Thomas, Chess and Korn. Merrill-Palmer Quarterly, 28, 25-33.

Porter, R. B., & Cattell, R. B. (1959). Children's Personality Questionnaire. Champaign, IL: Institute for Personality and Ability Testing.

Rothbart, M. K., & Derryberry, D. (1982). Theoretical issues in temperament. In M. Lewis & L. T. Taft (Eds.), Developmental disabilities: Theory, assessment, and intervention, New York: Spectrum Publications.

Rubin, R. A., & Balow, B. (1979). Prevalence of teacher identified behavior problems: A longitudinal study. In S. Chess & A. Thomas (Eds.), Annual progress in child psychiatry and child development. New York: Brunner/Mazel.

Rutter, M. (1967). A childrens behavior questionnaire for completion by teachers: Preliminary findings. Journal of Child Psychology and Psychiatry, 8, 1-11.

Rutter, M., & Graham, P. (1968). The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. British Journal of Psychiatry, 114, 563-579.

Satz, P., & Fletcher, J. M. (1979). Early screening tests: Some uses and abuses. Journal of Learning Disabilities, 12, 56-60.

Slosson, R. (1963). Slosson intelligence test manual. New York: Slosson Education Publications.

Spivack, G., & Swift, M. S. (1966). The Devereux Elementary School Behavior Rating Scale: A study of the nature and organization of achievement related disturbed classroom behavior. Journal of Special Education, 1, 71-91.

Spivack, G., & Swift, M. S. (1967). Devereux Elementary School Behavior Rating Scale. Devon, Pa: Devereux Foundation.

Swift, M. S., & Spivack, G. (1975). Alternative teaching strategies: Helping behaviorally troubled children achieve. Champaign, IL: Research Press.

Thomas, A. & Chess, S. (1977). Temperament and development. New York: Brunner/Mazel.

Thomas, A., & Chess, S. (1980). The dynamics of psychological development. New York: Brunner/Mazel.

Thomas, A., Chess, S., & Birch, H. (1968). Temperament and behavior disorders in children. New York: New York University Press.

Thomas, A., Chess, S., & Birch, H. (1970). The origin of personality. Scientific American, 223.

Thomas, A., Chess, S., & Korn, S. (1982). The reality of difficult temperament. Merrill-Palmer Quarterly, 28, 1-20.

Tiegs, E. W., & Clark, W. W. (1970). Manual of California achievement tests, 1970 Edition. New York: McGraw-Hill.

Tupes, E. C., & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. USAF ASD Technical Report, No. 61-53. (From Wiggins, J. S. (1973). Personality and prediction: Principles of personality assesment. Reading, Ma: Addison-Wesley.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 321-327.

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. Educational and Psychological Measurement, 38, 953-958.

Wiggins, J. S. (1973) Personality and prediction: Principles of personality assessment. Reading Mass.: Addison-Wesley.

Wirth, R. D., & Lachar, D. (1981). The Personality Inventory for Children: Developmental and clinical applications. In P. McReynolds (Ed.), Advances in psychological assessment (Vol. 5). San Francisco: Josey-Bass.

Zwick, W., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. Multivariate Behavioral research, 17, 253-269.