AN ABSTRACT OF THE THESIS OF

Tseng-Min Sun for the M.S. in H. Ec.
(Name) (Degree) (Major)

Date Thesis presented January 1939

Title The Development of a Controlled Observational Method for Measuring Selected Personality Traits of Preschool Children.

Abstract Approved: [redacted]
(Major Professor)

Twelve nursery school children were observed daily for a period of six weeks on their compliance with teachers. Three judges rated them on two scales of Merrill-Palmer Personality Rating, "Compliance with Routine" and "Response to Authority", while the observation was in process. Five more qualified judges were added after the completion of the observation. Each judge showed self-consistency on the rating of the two scales. Correlations between rating and observational scores ranged from .07 ±.19 to .59 ± .13 with a mean of .39 ±.16. The positive correlations, though low, suggest the possibility of assigning numerical values to children's compliance as revealed in observation of their behavior in natural nursery school situations. The present investigation with so few cases is suggestive only. Much further research is needed to determine the factors that cause disagreement between ratings and observation of personality traits.

All the judges except F showed high agreement with herself on the ratings of the two scales.

The observational method of measuring personality of children yielded differences in children's compliance with routine in school and also compliance with teachers.

It was found possible to assign quantitative measures to children's compliance.

Low correlations were found between the observational method and the rating scale method of measuring personality.

The results of the present study show no relationship seems to exist between compliance and intelligence.

****
****
**

THE DEVELOPMENT OF A CONTROLLED OBSERVATIONAL METHOD FOR MEASURING SELECTED PERSONALITY TRAITS OF PRESCHOOL CHILDREN

by

TSENG-MIN SUN

A THESIS

submitted to the

OREGON STATE COLLEGE

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

May 1939

APPROVED:

[redacted]

Professor of Child Development

In Charge of Major

[redacted]

Head of Department of Household Administration

[redacted]

Chairman of School Graduate Committee

[redacted]

Chairman of College Graduate Council

ACKNOWLEDGMENT

The writer wishes to express her deep gratitude to Dr. Vera Brandon who has given so generously of her time and constant guidance throughout the study. Dr. Katherine E. Roberts and Dr. Rachel Stutsman Ball have added heavily to her obligation by suggesting the general plan of the investigation and giving helpful criticisms during the preliminary study. She is also grateful to Misses Eleanor Spike, Jacqueline Brier, Floraine Simons, Bette Carothers, Mrs. Carolyn Sullivan and Mrs. Margaret Fleming for their cooperation in rating the children. Miss C. K. Lui helped greatly in checking statistical computations. Last, but not the least, she is indebted to Merrill-Palmer School which gave her pleasant and profitable experience that served as impetus to her interest in these rating scales.

TABLE OF CONTENTS

	Page
Introduction	1
Review of Some Studies in the Field of Personality Measurement by Means of Rating Scales	3
Problem	10
Subjects Included in the Present Study	11
Procedure	12
Observation	12
Construction of Blank Used in Study	12
Technique Used in Gathering Data	12
Situations Observed	14
Ratings of Subjects	16
Ratings by Three Judges	16
Ratings by Five Additional Raters	17
Mental Test	18
Treatment of Data and Results	20
Ratings	20
Ratings of Eight Judges	20
Self-consistency of Judges	25
Observational Data	26
Method of Classifying Behavior	26
Method of Assigning Numerical Values	28
Frequency of Each Degree of Compliance	29
Comparison of Data from Observation and Rating Scales	29

TABLE OF CONTENTS

Page

Comparison of Observational Data with Ratings by Three Judges 29

Comparison of Observation with Eight Ratings .. 31

Introspective Reports of Raters 34

Relationship between Intelligence and Compliance 35

Summary 36

Conclusions 37

References 38

Appendix 40

*****
***
*

LIST OF TABLES

			Page
Table	I	Age Range of the Twelve Children Included in the Study	11
Table	II	Number of Days Each Child was Observed and Total Number of Situations Recorded During the Observational Period	15
Table	III	Mental Measurements of Subjects as Shown by Scores on the Stutsman Mental Test ...	19
Table	IV	Judges' Evaluation of Childrens' Compliance with Routine as Measured by the Merrill-Palmer Personality Rating Scale and Reported in Terms of Scale Values ...	21
Table	V	Judges' Evaluation of Childrens' Response to Authority as Measured by the Merrill-Palmer Personality Rating Scale and Reported in Terms of Scale Values ...	22
Table	VI	Intercorrelations of Judges' Ratings for "Compliance with Routine"	23
Table	VII	Intercorrelations of Judges' Ratings for "Response to Authority"	24
Table	VIII	Self-Consistency of Each of the Eight Judges as Determined by Correlations Between the Scales "Compliance with Routine" and "Response to Authority"	25
Table	IX	Frequency of Each of the Five Degrees of Compliance for Each Child as Obtained from Observation	30
Table	X	Correlations Between Observational and Rating Scores for Each Judge on the Two Scales	31

*****
***
*

THE DEVELOPMENT OF A CONTROLLED OBSERVATIONAL METHOD FOR MEASURING SELECTED PERSONALITY TRAITS OF PRESCHOOL CHILDREN

INTRODUCTION

A glance at the voluminous literature on personality measurement will convince one of the growing interest in this new but promising field. Rating and observation are among the most widely used techniques concerning problems of personality study. Some writers, however, are still doubtful of the value of these two methods. Koos and Kefauver[1] say, "Ratings of character are highly subjective in nature and they possess all the weakness of subjective measures". Ruch and Stoddard[2] went as far as to say, "The widespread popular use of ratings can only be explained upon a basis of ignorance of the limitations of the method and the lack of better measures".

Goodenough[3] brings out the questioning of reliability of observations. She thinks it is very difficult to follow children who are running around without interfering with their activities.

---

[1]Koos, Leonard V. and Kefauver, Grayson N. Guidance in Secondary Schools, p. 355.

[2]Ruch, G. M. and Stoddard, George D. Tests and Measurements in High School Instruction, p. 313.

[3]Goodenough, F. The Observation of Children's Behavior as a Method in Social Psychology. Social Forces Vol. 15, No. 4, May 1937, pp. 476-479.

Most people established the reliability of observational data by checking the record of two raters who worked independently but simultaneously. There is no general agreement at present as to the statistical procedure best suited to show the extent of agreement.

REVIEW OF SOME STUDIES IN THE FIELD OF PERSONALITY MEASUREMENT BY MEANS OF RATING SCALES

Marston (9) employed rating and experiments in his study on extroversion-introversion of preschool children. The average correlation of the same children by different raters was .71 ±.09 and .54 ±.10 for boys and girls respectively.

Reynolds (12) reported a study on negativism of preschool children. Her subjects consisted of 229 children ranging in age from one year eleven months to five years six months with a mean of three years seven months. Their intelligence quotients ran from 60 to 170 with a mean of 114. She set up thirteen experimental situations in which the child was asked to do or not to do certain things such as "don't knock down" a tower or to imitate the examiner in clapping hands. A negative score of 1.0 was given for each failure to comply with the request in the stated time interval. Then teachers were asked to make judgments of the child's negativistic tendencies on a scale especially devised for this study. Correlations between teachers' rating were .84 ±.04 and .67 ±.08. Experimental negativistic scores correlated with teachers' rating scores .35 ±.12 for one group and .26 ±.06 for the other. The author believes the meaning of these low correlations is not at all clear. "So many unknown quantities are involved that

it is difficult to hazard even a guess as to interpretations."[1]

Newcomb (10) in his study on consistency of certain extrovert-introvert behavior patterns observed fifty-one problem boys over a period of five weeks at a summer camp. Daily record scores were checked against ratings. The mean correlation for Group I is .489 with a range from .23 to .73. For Group II the mean correlation is .409 with a range from .02 to .68.

Berne (1) combined with observation a rating scale of thirty paired traits and ten experimental situations for studying social behavior patterns in preschool children. When the scores on six experiments were used as a criterion, the ratings on the traits measured by the experiments were found to be valid measures. The correlations between rating and experimental scores ran from .58 $\pm$.05 to .91 $\pm$.01 with an average of .76 $\pm$.03. Observation scores agreed fairly high with rating scores and experimental scores.

Bridges' (2) Character Rating Chart consists of fifty items among which the following are similar to those contained in the two Merrill-Palmer School Personality Rating Scales employed in this study. They are, "Keeping order or confusion", "cooperative or not in routine", "adjusted to,

---

[1] Reynolds, Martha M. Negativism of Preschool Children. Teachers College, Columbia Univ., Cont. to Ed. 1928, No. 288, p.105.

or reacts against authority", "obedient or disobedient", "trustworthy or not in adult's absence", "controls tears or cries easily", "little or much disturbed by observation", and "does not show or shows temper". Scores on character rating were compared with intelligence test results. The rank difference coefficient of correlation between average rating scores and Stanford-Binet I.Q. were +.2 for the session 1928-29 and +.5 for the next session. The rating scores correlated with Stutsman Performance scores -.2 for one year and zero for the following year.

Conrad (3) studied the validity of personality ratings of preschool children by having three nursery school teachers rate thirty children aged 23 to 57 months in a behavior inventory of 231 trait descriptions devised by him. Each teacher was asked to star her ratings of any traits which she considered of dominating importance for the particular child being rated. The agreement among teachers concerning significant traits is high, for traits considered important by two judges the correlations ranged from .79 to .85; by three judges they ranged from .93 to .96. Reliability and validity of the inventory varied from child to child; he concluded some children can be rated with more accuracy than others. He also realized the absence of experimental or observational verification of the validity of the teachers' rating.

Jack (5) made a study of ascendant behavior in pre-school children. An experiment made up of a series pairings was devised. Validity of the experiment was established by its correlation of .81 ±.01 (rank difference) with a composite of three teachers' ratings.

Weiss (20) conducted an experiment on children's compliance with commands with seventy-seven children from two years, five months to five years, eight months of age. To each child she gave eight commands such as putting certain objects in certain places or bringing her certain things. Each child was given the same eight commands in each of three experimental situations. In the control situation the child had no toy to play with between commands. In the toy situation, toys were furnished; in the social situation, besides toys, a companion of the child's own age was present. Both time and behavior records were kept. Results showed age differences and differences between experimental situations were less evident in time records than in behavior records. Obedience was found to be complex. In determining compliance with commands, other individual factors are as significant as age, sex, and mental status.

Jersild and Markey (6) studied conflicts between nursery school children by means of observation. The objective data were then compared with teachers' ratings of various aspects of conflict behavior. The median correla-

tion between the various teachers' ratings is .57; the median correlation between composite teachers' ratings and observation data is .53. With regard to the later finding, the authors made the following comment and explanation. "Although the correlations between teachers' ratings and objective scores are far from perfect, it is still of interest to note all but one of the coefficients are positive, and that a majority of them range above .50. It is possible that a more carefully planned scheme for obtained teachers' ratings might have raised this correlation."[1] The ratings were obtained several months after the observation had started. "A series of ratings paralleling the observation might perhaps have yielded higher coefficients". Some of the raters had not known the children over a long period of time.

With regard to method of personality study, Olson (11) commented, "The trend from both direct observation and rating suggests that an eye-witness record of behavior or a rating judgment based on the impression from a series of events may possess more validity than an abstracted questionnaire or test instrument from which one must infer a relationship to the behavior. The disadvantage of the direct approach is the inconvenience of securing the record

---

[1] Jersild, Arthur T. and Markey, Frances V. Conflicts between Preschool Children. Teachers College, Columbia Univ., Child Develop. Monog. No. 21, 1931, p. 132.

and the inaccessibility of certain aspects of the experience and personality to direct observation.

Basing on Thurstone's technique of measuring attitudes, Roberts and Ball (13) developed a series of nine scales for studying personality in young children. The scales are Ascendance-submission, Sociability with Other Children, Compliance with Routine, Response to Authority, Respect for Property Rights, Tendency to Face Reality, Independence of Adult Affection and Attention, Attractiveness of Personality, and Physical Attractiveness. Each scale consists of a number of statements describing a child's behavior. Every trait measured was thought of being distributed along a linear continuum, with 1 representing the least degree of the characteristic measured, 6 the middle value, and 11 the greatest possible degree. From 50 to 70 judges were asked to sort the statements of each scale into 11 groups. The scale value of each statement was determined by using the median value assigned to it by the judges. The two scales, the validity of which the present investigation studies are "Response to Authority" and "Compliance with Routine". The former consists of 28 statements with the scale values ranging from 44 to 1. Forty-four showing the greatest response and 1 the least response. There are 31 statements in the latter scales, the values ranging from 11 to 99. In rating, the judges check the statement or statements that she feels really true of the child. The mean value of the different

items checked is the child's score for a given rater. The scales were developed in Merrill-Palmer School and have been used there for seven years.

PROBLEM

The present investigation was undertaken with an attempt to develop an observational method by means of which personality traits can be studied and which can be verified with another criterion. As preschool children are presumed to have simpler and more naive behavior than adults, the nursery school children were chosen as subjects. The next problem which the writer confronted was to select an aspect of behavior that lends itself to direct observation. Compliance was selected because of its importance in personality development and its availability for observation. Anyone who has experience with children knows that they might be compliant with some people but disobey others. As Conrad[1] pointed out relatively few experimental situations correspond to the genuine situations of life, it was thought better to observe children's reactions to teachers in everyday natural situations in nursery school than to put them through some artificial experiments and watch their response to experimenters whom the children do not tend to consider as of equal footing as their teachers.

---

[1] Conrad, Herbert S. The Validity of Personality Ratings of Preschool Children. Ped. Sem. and J. Genet. Psy., 1932, p. 41.

SUBJECTS INCLUDED IN THE PRESENT STUDY

The subjects of the present study were twelve children enrolled in the Government Nursery School at Corvallis, Oregon during the fall term of 1938. In August at the time the study was begun, the subjects ranged in age from 2 years 1 month to 5 years, with a mean of 3 years and 9 months (Table I). Eight of the children had been in the nursery school either one or two years; the other four, though enrolled during the first week of August, seemed well adjusted after the school had started three weeks and it was at this time that the writer began to collect her data. The number of subjects available were much fewer than the writer had planned. Unfortunately it was necessary for her to gather her data at the time the old college nursery school building was taken down and the new one was in the process of construction.

TABLE I

AGE RANGE OF THE TWELVE CHILDREN INCLUDED IN THE STUDY

Child	Age in Months	Child	Age in Months
1	58	7	28
2	59	8	53
3	60	9	44
4	49	10	25
5	38	11	37
6	52	12	39

PROCEDURE

Observation

Construction of Blank used in Study. In the Summer Session of the Nursery School of Oregon State College, a preliminary study was carried to determine the technique of collecting observational data. It was found that more data could be obtained by following the teacher than by following the children. A blank was devised which made it possible to record detailed information concerning each situation--the teacher's request and the child's response, also hour and date of the observation. Space was provided for scoring results. Preliminary observational work shows the feasibility of the blank in observing children's difference in response to authority. The observational forms were mimeographed and a separate sheet was used for each child's daily observation. A sample of the mimeographed form is included in the Appendix, page 40. The form shows how it was actually used.

Technique used in Gathering Data. On alternate days the writer followed each of the two teachers in the nursery school from 8:45 A.M. till 12:15 P.M. when the children went to the nap room after the dinner. The writer recorded all contacts that teacher had during the morning with each of the children under observation. Data were recorded immediately after its occurrence. The situations, teacher's

request, and the child's response were all entered into their proper places on the blank. Child-child contacts were not recorded unless they were concerned with certain teacher-child contacts under observation. The former data were however noted merely to give the entire pictures of the situation.

Any consciousness of being followed and having the words taken down verbatim would tend to make the teacher nervous and unnatural. The writer tried therefore to make her presence and following as inconspicuous as possible. Most of the time the writer kept a few steps away from the teacher so that she was within hearing distance but not sufficiently close to make the teacher aware of being followed. The teachers were merely told that the writer was studying the children and neither of the teachers asked the writer about the exact nature of her observation. There were two teachers in the nursery school and each was observed every other day. The children were accustomed to visitors and to being observed. The writer had been in constant attendance at the nursery school during the two weeks before actual recording was started so that the children became accustomed to her presence.

With this technique, undoubtedly the writer failed to record a few of the incidents which occurred due to failure on her part to see or to hear well. As the investigation

was designed to gather data from natural situations in nursery school, conditions were in no way controlled. Each of the children had had a different number of contacts with the teachers. This difference in contacts was due to one or more of the following causes: (1) Children were sometimes absent from the school; (2) A teacher happened to have more contacts with a certain child on a particular day than on another day; (3) One teacher took charge of the younger group more frequently than the other; (4) The teachers divided their daily duties; one inspected the children after arrival and the other supervised the nap room; (5) Children were assigned definite seats at meal time and they tended to be placed at the same table day after day with the same teacher.

Situations Observed. The number of situations the children had with both teachers and the days observed are presented in the following table.

Observations were made each school day during a period of six weeks, beginning on August 25 and ending on October 5. The writer planned for a longer observational period. This was made impossible because the college staff started to participate in the nursery school on September 24. Because the increase in personnel changed the environment of the children, it seemed best to stop the observation at the time the new authorities took over a large part in the direction of routine activities.

TABLE II

NUMBER OF DAYS EACH CHILD WAS OBSERVED AND TOTAL NUMBER OF SITUATIONS RECORDED DURING THE OBSERVATIONAL PERIOD

	Teacher A		Teacher B	
Subject Number	Number of Days Observed	Number of Situations Observed	Number of Days Observed	Number of Situations Observed
1	7	10	5	11
2	8	16	7	12
3	10	25	10	16
4	12	63	13	43
5	12	51	9	34
6	7	18	6	8
7	11	34	12	41
8	7	17	8	16
9	11	31	10	13
10	9	27	14	51
11	10	29	7	17
12	12	82	13	62
Average per Child	9.67	33.58	9.50	27.0

Rating of Subjects

Ratings by Three Judges. Two of Merrill-Palmer School Personality Rating Scales (For samples of scales see Appendix) were used to measure the children; "Compliance with Routine" and "Response to Authority". At the beginning of the investigation the writer gave each of the two teachers samples of the rating scales and told them that later on she would ask them to rate the children. This was done with the hope that a knowledge of the scales would help the teachers bear in mind the characteristics they should look for and evaluate in the children. Two weeks later, the teachers were given copies of the scale which is designed to measure the preschool child's "Compliance with Routine" and were asked to rate the children independently. As soon as the first set of scales were returned, the teachers were given the second set which is concerned with "Response to Authority". As a means of check up, the writer also rated the children. Her ratings were made before she analyzed the data from her observations. In this way her judgments were based on her observations as a whole instead of on an analysis of particular incidents observed.

The writer secured the ratings from the teachers or judges at the time the observational period was in progress, teachers therefore made their judgment on the behavior which was included in the writer's observations.

Qualifications of the judges are as follows: Judge A is a graduate nurse and has had charge of nursing in a children's hospital. She has also been a kindergarten teacher. Judge B is a college graduate and has been a nursery school teacher for five years. Judge C has had two years of experience in nursery schools and at the time the study was made was a graduate student in child study.

Ratings by Five Additional Raters. The data recorded on the observational blanks were converted into numerical scale values (technique described on page 28) and were compared with scores on the Merrill-Palmer Rating Scales. Correlations between the two measures by rank-order method ranged from .07 $\pm$.19 to .59 $\pm$.13 and from .30 $\pm$.18 to .48 $\pm$.15 on "Response to Authority" and "Compliance with Routine" respectively. As a means of check up and further analysis, five additional raters were secured. Four were staff members from the Oregon State College Nursery School who went to the government nursery school two or three times a week. Their nursery school experience ranged from two to eleven years. All are college graduates with specialized training in child development. The other rater was a senior student in home economics who has also had special training and experience in child development. All five judges had been in the nursery school eight weeks before making the ratings. They made their acquaintance with the children after the observational records were completed. Unfortun-

ately they did not see exactly the same behavior as was observed by the writer. They rated independently all the twelve children except child No. 5 who dropped from school shortly after the observations were completed. Like the other three judges, these five raters rated children first on the scale "Compliance with Routine" and later on the scale "Response to Authority".

Mental Test

The Stutsman mental test was given to all the children included in the study with the exception of child No. 5 who lived with his parents in a trailer and was not accessible at the time the tests were given. The mental tests were given after the observational work was completed. In interpreting the scores on this test, Stutsman-Ball recommends that, "For purposes of comparing the performance of children of different ages, or the performance of the same child at different ages, interpretation of the score by standard deviation in terms of score value and interpretation by percentile rank are preferable to interpretation in terms of the intelligence quotient".[1] Most people, however, are more familiar with the term intelligence quotient than with either the sigma value or the percentile rank. All three values are therefore presented in Table 3.

---

[1] Stutsman: Mental Measurement of Preschool Children, p. 235.

The range of intelligence quotients is from 92 to 126 with 105 as the median.

TABLE III

MENTAL MEASUREMENTS OF SUBJECTS AS SHOWN BY SCORES ON THE STUTSMAN MENTAL TEST

Child No.	Percentile Rank	Sigma Value	Intelligent Quotient
1	67	+0.5	115
2	56	0.0	106
3	25	-0.5	92
4	50	0.0	98
5	--	-.-	--
6	65	+0.5	105
7	85	+1.0	113
8	63	+0.5	105
9	63	+0.5	104
10	45	0.0	96
11	45	0.0	97
12	93	+1.5	126

TREATMENT OF DATA AND RESULTS

Ratings

Ratings of Eight Judges. The ratings of each of the eight judges together with the averages of all are presented in Tables 4 and 5, pages 21 and 22.

The intercorrelations though low were not lower than what should be expected. As Koos and Kefauver (7) said, "Ratings of character are low in reliability. There is wide variation in the ratings given by a number of different persons or by the same person at different times." Volker (18) and Webb (19) reported an average reliability of .55. Shen (15) also obtained an average of .55 from his 104 correlations. Kornhauser (8) had three independent ratings checked against three other independent ratings and got an average reliability of .44. The last author explained, "Disagreements between ratings of the student may indicate that the student is different in different activities--not that instructors are wrong in their estimates. What is called unreliability of ratings may really be in large measure a matter of the specificity of traits." Symonds (17) remarked, "The reliability of ratings has been found to be variable and disappointingly low; indeed, it has been found by various workers to be so low as to cast grave doubt on the value of rating as a method for gathering trustworthy data". According to Symonds, "A reliability coefficient

TABLE IV

JUDGES' EVALUATION OF CHILDREN'S COMPLIANCE WITH ROUTINE AS MEASURED BY THE MERRILL-PALMER PERSONALITY RATING SCALE AND REPORTED IN TERMS OF SCALE VALUES

Child No.	A	B	C	D	E	F	G	H	Average	**Percentile Rank
1	91.57	81.50	90.00	89.86	91.44	88.60	91.44	91.86	89.53	81.77
2	90.86	90.57	87.88	87.71	90.71	56.17	89.78	90.00	85.46	68.60
3	86.20	79.60	88.80	84.33	61.00	65.75	89.71	82.00	79.67	45.57
4	84.60	50.33	89.88	77.50	65.83	89.80	83.50	28.50	71.24	26.94
*5	81.29	76.50	53.83			60.00			67.91	29.85
6	75.67	59.25	72.50	75.67	88.11	87.40	90.38	29.67	72.32	28.13
7	64.57	74.67	46.00	59.67	40.00	65.00	63.33	32.25	55.68	12.10
8	56.00	44.50	86.00	60.88	85.38	66.50	53.16	51.67	63.01	18.35
9	51.83	80.14	86.00	76.80	87.38	85.00	77.60	84.25	78.63	55.43
10	48.60	34.33	51.67	83.00	25.67	52.00	70.14	28.00	49.18	6.55
11	48.17	58.50	52.57	53.25	83.71	76.75	59.67	44.50	59.64	17.05
12	35.60	56.00	37.33	45.78	32.75	35.50	52.11	51.67	43.34	3.67

*Judges contacts with child were too few to evaluate compliance with routine.

**Roberts, Katherine E. and Ball, Rachel Stutsman. A Study of Personality in Young Children by Means of a Series of Rating Scales. J. of Genetic Psy., March, 1938, p. 145.

TABLE V

JUDGES' EVALUATION OF CHILDREN'S RESPONSE TO AUTHORITY AS MEASURED BY THE MERRILL-PALMER PERSONALITY RATING SCALE AND REPORTED IN TERMS OF SCALE VALUES

Child No.	A	B	C	D	E	F	G	H	Average	**Percentile Rank
1	38.67	36.50	39.50	34.80	38.50	33.80	33.83	37.25	36.61	78.70
2	38.67	38.67	31.67	37.25	38.00	31.80	30.50	38.67	35.65	75.50
3	37.17	28.00	31.60	33.00	16.00	19.14	28.71	32.50	28.27	44.23
4	27.29	17.67	23.00	40.00	18.00	37.60	27.50	24.50	26.95	39.83
*5	36.43	31.60	12.50			33.40			28.48	44.93
6	26.25	22.00	24.33	37.25	31.00	19.40	32.75	17.67	26.33	37.77
7	35.60	29.75	13.83	14.20	9.33	20.71	21.17	12.75	19.67	19.18
8	27.00	14.40	31.67	32.50	23.00	26.67	13.00	34.00	25.28	34.27
9	20.33	30.40	34.25	16.75	20.00	30.40	16.17	38.00	25.79	35.47
10	8.50	23.67	10.50	33.50	8.33	10.67	12.00	12.50	14.95	8.25
11	24.00	23.00	19.20	10.40	34.00	13.40	11.80	29.50	20.66	21.65
12	20.00	15.00	11.71	14.56	10.67	23.67	9.14	16.40	15.14	8.53

*Judges contacts with child were too few to evaluate response to authority.

**Roberts, Katherine E. and Ball, Rachel Stutsman. A Study of Personality in Young Children by Means of a Series of Rating Scales. J. of Genetic Psy., March, 1938, p. 145.

TABLE VI

INTERCORRELATIONS OF JUDGES' RATINGS--COMPLIANCE WITH ROUTINE

Rater	A	B	C	D	E	F	G	H
	Correlations with Probable Errors							
A		.64 ±.115	.81 ±.067	.81 ±.070	.62 ±.125	.37 ±.168	.86 ±.053	.42 ±.166
B			.46 ±.153	.49 ±.150	.67 ±.112	.16 ±.190	.64 ±.120	.80 ±.073
C				.78 ±.080	.64 ±.120	.68 ±.109	.72 ±.098	.45 ±.162
D					.46 ±.160	.25 ±.190	.82 ±.067	.39 ±.172
E						.60 ±.130	.67 ±.112	.61 ±.128
F							.53 ±.146	.01 ±.203
G								.30 ±.185

TABLE VII

INTERCORRELATIONS OF JUDGES' RATINGS--RESPONSE TO AUTHORITY

Rater	A	B	C	D	E	F	G	H
				Correlations with Probable Errors				
A		.58 ±.129	.59 ±.127	.39 ±.172	.52 ±.148	.52 ±.148	.78 ±.080	.54 ±.144
B			.30 ±.177	.03 ±.203	.29 ±.186	.15 ±.199	.48 ±.156	.46 ±.160
C				.30 ±.185	.64 ±.120	.49 ±.150	.69 ±.106	.84 ±.060
D					.29 ±.186	.48 ±.156	.71 ±.101	.21 ±.194
E						.45 ±.162	.53 ±.146	.77 ±.083
F							.45 ±.162	.58 ±.135
G								.43 ±.166

of .55 can be said to be typical for rating personality traits by ordinary judgment methods".

In the present study there is no conclusive evidence to indicate whether this variability in rating scores is due to inconsistency of children's behavior at different times, to children's difference in response to different people, or to difference in raters' interpretation of the same or similar behavior.

Self-consistency of Judges. To study the self-consistency of judges, agreement between the scores of each judge was determined. Correlations were computed between the two scales, "Compliance with Routine" and "Response to Authority" by rank-difference method; these were then converted into product-moment correlations.

TABLE VIII

SELF-CONSISTENCY OF EACH OF THE EIGHT JUDGES AS DETERMINED BY CORRELATIONS BETWEEN THE SCALES "COMPLIANCE WITH ROUTINE" AND "RESPONSE TO AUTHORITY"

Rater	Correlation with P E	Rater	Correlation with P E
A	.92 ±.03	E	.94 ±.02
B	.86 ±.05	F	.39 ±.16
C	.84 ±.06	G	.95 ±.02
D	.72 ±.09	H	.84 ±.06

Correlations show that each of the judges agreed with herself. It appears, therefore, that each judge kept the same traits in mind for each child as she checked the two scales. These results agree with Roberts and Ball's[1] finding of a correlation of .75 ±.035 between these two scales, being the highest among the intercorrelations of the nine scales. They think it is possible to combine these three closely related scales, "Compliance with Routine", "Response to Authority", and "Respect for Property Rights", into a single scale which measures a general trait of compliance.

Observational Data

Method of Classifying Behavior. At the end of the observational period the children's behavior which had been recorded on the data sheets was divided into situations or items. An item, as used in this study means the child's response to a teacher's single verbal request or gesture. As some children comply with the first request that is made and others will not give the desired response before the third or the fourth request, it was thought best in assigning numerical score values to divide situation into small units of behavior rather than to attempt to classify responses on the basis of larger units. Thus, if it took three requests from the teacher to get the child to put a

---

[1] Roberts, Katherine E. and Ball, Rachel S. A Study of Personality in Young Children by Means of a Series of Rating Scales. J. of Genetic Psy., 1938, 52, p. 121.

toy away, this behavior was divided into three situation items.

A method of assigning numerical scale values was worked out. Classification of the child's behavior was based on his degree of compliance. A child's complete compliance to the first request was considered as falling at one extreme of a linear continuum and physical rebellion was placed at the other extreme. Other degrees of compliance or non-compliance were evaluated and assigned a position on a five point continuum scale.

The scale developed by the writer for classifying her observational data into the five apriori catagories is as follows:

Degree 5. Complete compliance--compliance with the first request.

Degree 4. Partial compliance.

a. Did part of the suggested activity.

b. Complied after neutral response.

c. Complied with hesitation or reluctance.

Degree 3. Neutral response.

a. No overt response, either positive or negative.

b. Went on with own activity as if no request had been made.

c. Did not do the suggested activity even though he was not engaged in play of his own.

Degree 2. Surrendered to second or later request.

a. Verbal argument as to why he should not do the suggested activity.

b. Surrendered after physical rebellion.

c. Evasion--by talking or by action.

d. Second or later neutral response.

Degree 1. Physical objection.

a. Temper-tantrum.

b. Ran away when called.

c. Cried.

d. Second or later refusal.

e. Clung to certain objects and refused to go.

The reliability of sorting children's situations of behavior was determined by having raters C and E classify independently into categories three children's observational records which were picked up at random and consisted of data covering seven days each. The reliability of this technique is shown by correlations of .93 ±.02, .94 ±.02, and .96 ±.01 (by product-moment method) for each of the children.

Method of Assigning Numerical Values. In order to make the observational data scores comparable with the rating scales, the following technique was devised. The Merrill-Palmer scale on "Compliance with Routine" consists of 31 statements with scale values ranging from 11 to 99. These values were arranged on a linear continuum according to the

scale value assigned to each statement. The continuum was then divided into five equal intervals. The mid-point value for each interval was used as the scale value for each unit of behavior classified in that interval. In this way the writer weighted the behavior falling into the respective intervals. From the extreme end of the scale which represents complete compliance, the values for each of the five step intervals are 91, 73, 55, 37, and 19. In the scale "Response to Authority" there are 28 statements. The scale values range from 1 to 44. The same technique as was used in the scale "Compliance with Routine" was employed for this scale. Values for each interval are 40.50, 31.50, 22.50, 13.50, and 4.50 respectively.

Frequency of Each Degree of Compliance. Percentages of the frequency of each degree of compliance are shown in the following table. Percentages were computed to the nearest whole number.

Comparison of Data from Observation and Rating Scales

Comparison of Observational Data with Ratings by Three Judges.

As Hughes pointed out validity of a rating can be checked when there is agreement between the actual reactions of the individual and the results of the rating, scores from the observational data which were based on the contacts of judges A and B with children were compared with each of these judges own ratings. It would seem that the children's

TABLE IX

FREQUENCY OF EACH OF THE FIVE DEGREES OF COMPLIANCE FOR EACH CHILD AS OBTAINED FROM OBSERVATION

Child No.	Teacher A Frequency of Degree of Compliance Reported in Percentages						Teacher B Frequency of Degree of Compliance Reported in Percentages					
	5	4	3	2	1	Total	5	4	3	2	1	Total
1	80.00	10.00	10.00	0.00	0.00	100.00	81.82	0.00	9.09	9.09	0.00	100.0
2	87.50	6.25	0.00	0.00	6.25	100.00	83.34	8.33	0.00	8.33	0.00	100.0
3	40.00	24.00	24.00	12.00	0.00	100.00	81.25	6.25	12.50	0.00	0.00	100.0
4	68.25	15.87	14.29	1.59	0.00	100.00	62.79	18.60	11.63	6.98	0.00	100.0
5	60.78	13.73	3.92	13.73	7.84	100.00	67.65	17.65	14.70	0.00	0.00	100.0
6	50.00	22.22	27.78	0.00	0.00	100.00	87.50	0.00	12.50	0.00	0.00	100.0
7	82.35	11.77	5.88	0.00	0.00	100.00	56.10	26.83	9.75	7.32	0.00	100.0
8	76.47	11.76	5.88	5.88	0.00	100.00	81.25	6.25	12.50	0.00	0.00	100.0
9	67.74	9.68	12.90	9.68	0.00	100.00	69.23	15.38	15.38	0.00	0.00	100.0
10	18.52	22.22	22.22	14.82	22.22	100.00	58.82	17.65	9.80	7.84	5.88	100.0
11	55.17	24.14	20.69	0.00	0.00	100.00	70.59	17.65	5.88	5.88	0.00	100.0
12	28.05	25.61	23.17	15.85	7.32	100.00	53.24	19.36	9.67	14.51	3.22	100.0

behavior as recorded on the observational blanks should be the bases on which the judges made their estimates, because these measures were made simultaneously. Ratings made by the writer who served as judge C were also compared with the average score obtained from the observations of judges A and B; the latter two judges being the nursery school teachers in charge. The writer's opinion of a child should be based on what that child did in the observed situations for either or both of the judges. Since the cases were few in number, correlations were determined by the rank-order method, then by means of tables were converted into product-moment equivalents.

TABLE X

CORRELATIONS BETWEEN OBSERVATIONAL AND RATING SCORES FOR EACH JUDGE ON THE TWO SCALES

Judge	Compliance with Routine		Response to Authority	
	r	PEr	r	PEr
A	.48	±.15	.59	±.13
B	.30	±.18	.07	±.19
C	.44	±.16	.45	±.16

Comparison of Observation with Eight Ratings. Because the correlations between observational data and the ratings of the three judges were low, five more raters were secured. In general, mean scores of several ratings is more reliable

than the rating of a single judge. The mean of the eight judgments was therefore checked against the observational scores.*

Correlations for the mean scores of the eight judges with the observational data on "Compliance with Routine" are .48 ±.15 and with "Response to Authority" .30 ±.18. Results show the addition of raters did not raise the agreement between the two measures. Since as previously mentioned, the judgments of the five added judges were made approximately eight weeks after the observational data were secured. It is probable in this period of time the children had changed in their behavior to some extent.

Hughes (4) pointed out few studies have been attempted to verify the results of rating with actual behavior of the individual. There are not many published reports on validity of rating since Hughes made this comment. The following are some studies the results of which are nearly comparable to those secured in the present investigation.

Between experimental scores and rating scores Marston (9) obtained correlations of .52 ±.09 and .54 ±.10 for boys and girls respectively.

Reynolds (12) found his experimental negativistic scores correlated with teachers' rating scores .35 ±.12 for

---

* All children were independently rated by eight judges with the exception of child No. 5 who dropped from school shortly after the close of the observational period.

one group and .26 ±.06 for the other.

Newcomb (10) checked daily record scores against ratings. The mean correlation for Group I is .489 with a range from .23 to .73. For Group II the mean correlation is .409 with a range from .02 to .68.

Berne (1) reported an average correlation of .76 ±.03 between her rating and experimental scores.

Jack (5) established the validity of her experiment by its correlation of .81 ±.01 (rank-difference) with a composite of three teachers' ratings.

Jersild and Markey (6) secured a correlation of .53 between composite teachers' ratings and observation data.

The discrepancy between rating and observational scores the writer believes may be due to one or more of the following factors:

1. The behavior record of a child was by no means a complete picture of his reaction to a particular teacher. The fact that the writer tried to make her following inconspicuous caused her to miss part of the teacher-child contacts. As each teacher was observed on alternate days, a part of the child's reactions during the observational period was left out.

2. The teachers' opinion of a child might be influenced by what they learned from his parents of his compliance at home although they were told to base their judgment on his behavior in school.

3. One rating is insufficient to compare with observation.

4. The observational record fails to give credit to those children who went from one routine to another without any request from the teacher.

Introspective Reports of Raters. With the hope to throw some light on the lack of high agreement between rating and observational scores, a rough analysis was made to study the basis on which the judges made their estimates. After the judges had rated the children they were asked some questions concerning their experience in rating. (See Appendix for complete form).

Their opinions are summarized as follows:

1. The rating scales fail to provide for degrees of compliance. Some statements in both scales were true of the children occasionally, but were not absolutely true all the time.

2. Some of the terms are vague in meaning i.e., "usually", "a long time", "often", are difficult to evaluate.

3. To what degree may a characteristic be present and considered typical? i.e., How many times does a child need to "pretend not to hear" to be given this rating?

4. It is difficult to evaluate such feelings of a child as "proud", "absorbed in own thoughts", "pretended absorption", etc.

5. Some children were not easy to rate because of

their inconsistency of behavior.

6. The judges tried to base their estimates on specific instances. One judge reported she attempted to recall many specific reactions of a child but a few stood out more clearly than the others.

7. They suggested the rating scales could be improved by having a range of the variables and having a definition of the terms.

Relationship between Intelligence and Compliance. To determine the relationship between intelligence and compliance mental test results in terms of percentile rank were compared with average rating scores given by eight judges. On the scale "Compliance with Routine" a correlation (by rank-order) of -.16 ±.19 was obtained; on the other scale "Response to Authority" the correlation was -.13 ±.03. This is in agreement with the report of Bridges (2) who found correlations of -2 and 0.

Between observational data and intelligence scores the writer secured a correlation of .36 ±.17.

SUMMARY

Twelve nursery school children were observed daily for a period of six weeks on their compliance with teachers. Three judges rated them on two scales of Merrill-Palmer Personality Rating, "Compliance with Routine" and "Response to Authority", while the observation was in process. Five more qualified judges were added after the completion of the observation. Each judge showed self-consistency on the rating of the two scales. Correlations between rating and observational scores ranged from .07 ±.19 to .59 ±.13 with a mean of .39 ±.16. The positive correlations, though low, suggest the possibility of assigning numerical values to children's compliance as revealed in observation of their behavior in natural nursery school situations. The present investigation with so few cases is suggestive only. Much further research is needed to determine the factors that cause disagreement between ratings and observation of personality traits.

CONCLUSIONS

1. All the judges except F showed high agreement with herself on the ratings of the two scales.

2. The observational method of measuring personality of children yielded differences in children's compliance with routine in school and also compliance with teachers.

3. It was found possible to assign quantitative measures to children's compliance.

4. Low correlations were found between the observational method and the rating scale method of measuring personality.

5. The results of the present study show no relationship seems to exist between compliance and intelligence.

****
**

REFERENCES

1. Berne, Esther Van Cleave. An experimental investigation of social behavior patterns in young children. Univ. Iowa Stud., Stud. in Child Welfare, 1930, 4, No. 2, 61.

2. Bridges, K. M. Banham. The social and emotional development of the preschool child. London: Kegan Paul, Trench, Trubner, 1931, X, 277.

3. Conrad, Herbert S. The validity of personality ratings of preschool children. Ped. Sem. and J. Genet. Psy., 1932, 41, 267-293.

4. Hughes, W. H. A rating scale for individual attitudes and interests. Journal of Ed. Method, 1923, 3, 56.

5. Jack, Lois M. An experimental study of ascendant behavior in preschool children. Univ. Iowa Stud., Stud. in Child Welfare, 1934, 9, No. 3, 7-65.

6. Jersild, Arthur T. and Markey, Frances V. Conflicts between preschool children. Teach. Coll., Columbia Univ., Child Develop. Monog. No. 21, 1935, IX, 181.

7. Koos, Leonard V. and Kefauver, Grayson N. Guidance in secondary schools. The McMillan Company, New York, 1937, XI, 640.

8. Kornhauser, A. W. Reliability of average ratings. J. of Personnel Research, 5, 1926, 309-317.

9. Marston, Leslie R. The emotions of young children. Univ. Iowa Stud., Stud. in Child Welfare, 1925, 3, No. 3, 99.

10. Newcomb, Theodore M. The consistency of certain extrovert-introvert behavior patterns in 51 problem boys. Teach. Coll., Columbia Univ., Cont. to Educ., 1929, No. 382, 123.

11. Olson, Willard C. Measures of character and personality through conduct and information. Rev. of Ed. Research, June 1935, Vol. 5, 3, 274.

12. Reynolds, Martha May Negativism of pre-school children. An observational and experimental study. Teach. Coll., Columbia Univ., Cont. to Educ., 1928, No. 288, VIII, 126.

13. Roberts, Katherine E. and Ball, Rachel S. A study of personality in young children by means of a series of rating scales. J. of Genetic Psy., 1 March, 1938, 79-149.

14. Ruch, G. M. and Stoddard, George D. Tests and measurements in high school instruction. World Book Co., New York, 1927, 313.

15. Shen, Eugene The reliability coefficient of personal ratings. J. of Ed. Psy., 16, April, 1925, 236.

16. Stutsman, Rachel Mental measurement of preschool children. World Book Co., New York, 1931, X, 368.

17. Symonds, Percival M. Diagnosing personality and conduct. Appleton-Century Co., New York, 1931, XVI, 602.

18. Volker, Paul F. The function of ideals and attitudes in social education. Teach. Coll. Contributions to Educ., No. 112, 1921.

19. Webb, Edward Character and intelligence. British J. of Psy. Monog. Supplement, No. 3, 1915, 1, 27.

20. Weiss, LaBerta A. An experimental investigation of certain factors involved in the preschool child's compliance with commands. Univ. of Iowa Stud., Stud. in Child Welfare, Vol. 9, No. 3, 1934, 129-157.

******
****
**

APPENDIX

OBSERVATION BLANK

Child - Allen

Place - Roosevelt School

Teacher - B

Date - 8/29/38 Mon.

Time	Stimulus		Response		Score				
	Situation	Teacher's behavior and request	Child's behavior	Verbal response	5	4	3	2	1
9:50	Allen forgot to put away cubes. Went to play with blocks.	"I think you have something to put away first."	Went back to cubes but did not put them away.			*			
		"You put these (cubes) away."	Stood still.				*		
		"You put these on the shelf."	He did.			*			
10:10	Washing time. Allen wanted to go out without washing.	"We are not quite ready to go out yet."	Came back to wash.		*				
10:25	Allen was climbing on slide. Wanted to kick off ladder.	"Leave the ladder there."	Left it there.		*				
					Total				
					2	2	1	0	0

40

Merrill-Palmer School

PERSONALITY RATING

Date _____

Name _____ Rated by _____

Directions for Rating:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones.

( ) Adjusts immediately to the daily routine.
( ) Objects violently to routine activities.
( ) Always goes through the daily procedure willingly.
( ) Usually goes through the daily procedure willingly.
( ) Has to be constantly urged to carry out routine activities.
( ) Takes a long time to adjust to the daily routine.
( ) Tries to prevent other children from carrying out the routine activities.
( ) Quiet enjoyment from routine activities.
( ) Accepts the routine as a matter-of-course.
( ) Responds readily to direction in the day's routine.
( ) Likes to assist the adult in routine tasks.
( ) Cooperates or not in routine, according to his mood.
( ) Acts silly at the lunch table.
( ) Refuses many foods.
( ) Often cries during nap period.
( ) Objects to being examined by the nurse.
( ) Talks and laughs with adjacent children during rest period or nap.
( ) Routine proceeds as usual in the presence of visitors.
( ) Presence of visitors upsets his routine.
( ) Presence of a specific child upsets his routine.
( ) Emotional outburst upon leaving parents.
( ) Business like and systematic in endeavoring to carry out routine activities.
( ) Dawdles over routine activities.
( ) Carries out routine tasks in a haphazard manner.
( ) Makes a routine of his play activities.
( ) Always cooperates in trying to keep the schoolrooms neat and clean.
( ) Usually cooperates in trying to keep the schoolrooms neat and clean.
( ) Usually put things away carefully.

( ) Occasionally puts things away.
( ) Seldom cooperates in trying to keep the schoolrooms neat and clean.
( ) Never puts things away.

Merrill-Palmer School

PERSONALITY RATING

Date _____

Name _____ Rated by _____

Directions for Rating:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones.

RESPONSE TO AUTHORITY

( ) Attempts to change conversation from suggested activity to other channels.
( ) Adds cooperative additions to the suggestion.
( ) Resistant even against suggestion.
( ) Planned evasion.
( ) Proud of his cooperation.
( ) Lags in following suggestion.
( ) Responds without undue delay to authority.
( ) Cries if has to submit to authority.
( ) Runs away if called.
( ) Comes quickly if called.
( ) Thinks of immediate arguments as to why he shouldn't do suggested activity.
( ) Contemplates legitimate suggestion long time before doing it.
( ) Resists when required to do something new.
( ) Says "No" but does suggested activity.
( ) Says "Yes" but doesn't do suggested activity.
( ) Frowns, shrugs shoulders, pouts or stamps foot when suggestion made.
( ) Resistant only when in a particular mood.
( ) Pretends not to hear.
( ) So absorbed in own thoughts that doesn't comprehend.
( ) Defiant against authority.
( ) Accepts any command without question.
( ) Experimentation with new authority to see how far he can go.
( ) Rebels physically: temper-tantrum, hitting, kicking, etc.
( ) Pretended absorption to evade suggestion.
( ) Cooperative and responsible.
( ) Child tries to get task done by the person who suggests it.
( ) Follows suggestion only while teacher is in sight.

( ) Resists if suggestion is not about the things he has planned himself.

Name_____

Date_____

As a part of my thesis problem, I am studying two of the Merrill-Palmer rating scales. Some time ago you checked these scales for several children. I am interested in knowing your reaction to these scales. Please include not only material on the following questions, but also give in detail any other facts that will describe your reactions to these scales.

Tseng-Min Sun

When you checked the Merrill-Palmer Personality Rating Scale:

1. Did you have any difficulty in rating children? If so, what are the difficulties?

a.
b.
c.
d.
e.

2. Did you find it equally difficult to rate all the children? Why?

3. In rating a certain child, did you base your judgment upon your general impression of him or upon certain specific instances? Explain.

4. What would you suggest to improve the rating scale?