The Pupil Evaluation Inventory

A Sociometric Technique for Assessing Children's Social Behavior[1]

**Eugene G. Pekarik, Ronald J. Prinz, Diane E. Liebert,
Sheldon Weintraub, and John M. Neale[2]**
State University of New York at Stony Brook

The Pupil Evaluation Inventory was developed to assess peer ratings of the behavior of male and female children in grades one through nine. Three homogeneous and stable factors emerged from a factor analysis: Aggression, Withdrawal and Likeability. High interrater agreement between male and female raters, stability of factorial structure, high internal consistency, and moderate concurrent validity were found across sex and grade levels. Test–retest reliability was high for a separate sample of third and sixth graders. There was a tendency for higher internal consistency and teacher–peer correlations for the Aggression factor.

Peer evaluations in a classroom setting are particularly valuable for assessing childhood adjustment. Interactions with peers constitute the most significant social arena a child encounters, outside of his or her family, and the classroom situation is representative of the work, competitive, and social demands with which the child will later have to cope. Furthermore, peer evaluations are obtained in the rich, nontest context of the child's real-life environment and are based on observations made over extended periods of time by multiple observers with whom the child has different personal relationships, and who consequently

view him from varying perspectives (Smith, 1967). On a more empirical level, peer evaluations have been demonstrated to be stable over time (Minturn & Lewis, 1968; Sells & Roff, 1967), across sex of raters, and over a wide age range (Minturn & Lewis, 1968), and have been shown to be minimally influenced by the tendency for raters to respond with prosocial evaluations (Norman, 1963; Walder, Abelson, Eron, Banta, & Laulicht, 1961; Wiggins & Winder, 1961). Peer evaluations have been validated successfully against parental, clinician, and teacher ratings, as well as behavioral observation (Bower, 1969; Wiggins & Winder, 1961; Winder & Wiggins, 1964) and have effectively predicted maladjustment (Roff & Sells, 1968; Rolf, 1972).

Peer evaluations also fare well when compared with other means of assessing childhood adjustment. Both parent ratings and self-ratings are heavily biased by a prosocial rating set (Eron, Walder, & Lefkowitz, 1971; Himelstein & Lubin, 1965). Furthermore, parent ratings are contaminated by parents' tendency to systematically distort reports of unpleasant events in their children's lives (Burton, 1970) and self-ratings have been found to have little relationship to overt behavior (Buss, 1967; Leibowitz, 1968). Teacher ratings are generally considered as relatively reliable and valid assessors of children's adjustment (Bardon & Bennett, 1972) but teacher ratings combined with peer ratings are a more valid index of disturbance than teacher ratings alone (Bower, 1967). In direct comparisons with teacher ratings, peer ratings were better at discriminating sons of schizophrenic mothers from their controls (Rolf, 1972) and predicting school success (Smith, 1967).

Most research on the relationship between peer evaluations and maladjustment has involved undifferentiated peer problems as measured by acceptance—rejection (Sells & Roff, 1967). Such an approach has the obvious limitation of ignoring specific behavior patterns associated with maladjustment. A potentially more fruitful approach was first developed by Mitchell (1956) who factor analyzed a peer rating scale (Havighurst, DeHaan, Dietrich, Hackamack, Johnson, & King, 1952) composed of items designed to assess specific behavior patterns and isolated the factors of "social acceptability," "aggressive maladjustment," and "social isolation." Subsequent factor analyses of other such peer evaluation instruments have found factors composed of items which describe behaviors associated with maladjustment—aggression (Walder et al., 1961; Wiggins & Winder, 1961), social withdrawal, depression, and dependency (Wiggins & Winder, 1961; Siegelman, 1966), social acceptability (Mitchell, 1956), likeability (Siegelman, 1966; Walder et al., 1961) and academic achievement (Smith, 1967).

Unfortunately, the available peer evaluation scales have been designed for use with restricted ranges of subject populations and/or limited categories of behaviors. For example, the most rigorously derived and tested peer evaluation scales are those developed by Walder et al. (1961) and Wiggins and Winder (1961); the Walder et al. instrument is an assessment of aggression only, and the

Wiggins and Winder scale is designed for use with preadolescent males only. A more useful instrument would assess a broad range of behaviors in both males and females of a wide age range. The development of the Pupil Evaluation Inventory to meet these needs was the major goal of this study. Secondarily, it was hoped that the instrument would tap behavioral characteristics which would predict later adjustment.

METHOD

Format and Selection of Items

An item by peer matrix was chosen as a format since it allows every child to be selected for each item; an additional advantage is that all children are rated item by item preventing a possible bias or set that may develop when all items are rated for one person at a time. The items appeared as rows down the left side of the page and the names of the children in the class were across the top of the page; six pages, approximately six items per page, were used. The rater checked each child he believed to be described by a particular item. In any classroom only one sex was rated by both boys and girls in that class as well as by the teacher. After rating the other children in the class, each child made a self-evaluation on the same form. In this way, peer, teacher, and self-ratings were obtained in only one 30-minute period.

The choice of items to be included in the instrument was guided by two considerations. First, most of the items were drawn from pools loading on discrete factors in previous studies of peer rating scales, thus tapping types of behaviors already demonstrated to be reliably identifiable. Second, items were selected on the basis of their reported association with psychopathology in the literature. A survey of the research literature on peer ratings disclosed several rating devices with acceptable reliability and validity: Lesser (1959); the Minturn and Lewis (1968) analysis of Walder et al. (1961); Mitchell (1956); the Siegelman (1966) analysis of Winder and Wiggins (1961); Walder et al. (1961); Winder and Wiggins (1961). Factor analyses of each of these instruments were inspected, and items with the highest loadings on each factor were compiled for a total of 13 factors and 80 items. From this list, 29 redundant items were eliminated; the remaining 51 items were rationally grouped into eight categories: classroom disruption, aggression, social isolation, neurotic behaviors, likeability, and intelligence. These items, along with 10 others rationally generated to cover behaviors not assessed by other rating scales but considered important indices of maladjustment, served as the first pilot scale.

The pilot instrument was presented to 45 third- and fourth-grade children; one class rated boys and the other girls. A factor analysis of the results led to a

reduction of the initial item pool to 39; some were not endorsed sufficiently and subjects reported that some items "meant the same thing." The revised form was then readministered to a second sample consisting of 74 second- and third-grade children and the results were again subjected to principal components analysis. Five principal components were obtained and labeled: (1) aggressive disruptiveness; (2) immature, nonaggressive disruptiveness; (3) social isolation; (4) oversensitive, unhappy; and (5) popularity and likeability. Four items which children had asked questions about and found difficult to understand were eliminated for the final format which then consisted of 35 items, the first being a filler item which was used to demonstrate how the items were to be answered. This final form was used for all classes, grades two through nine; a shorter version consisting of 17 items, some from each factor, was used for first graders who would have found the longer form too tiring.

Subjects

The Pupil Evaluation Inventory was administered to a large number of classes, grades one through nine. The classes were selected from many different schools which were predominantly white, suburban, lower- to upper-middle-class areas on Long Island, New York. For the purposes of data analysis, three children were selected from each class.[3] A sample of 181 males and 171 females was thus formed who had been rated by the remaining children of their classrooms and their teacher; there were approximately 4000 raters.

Instructions and Procedure

Students were told that a survey was being conducted to "find out what kinds of things boys and girls do at different ages and in different grades." They were asked to think about either the boys or the girls in their class and to put an X under the name of everyone who fit the description on the questionnaire. Students were told to cross out their own names and not to rate themselves. As the students completed their ratings, the teacher also rated either the boys or the girls on the same instrument.[4] Following the completion of the inventory, all

---

[3] This research was conducted as part of a large study on the adjustment of the children of psychiatric patients (Neale & Weintraub, 1975). In each classroom one of the three children had a parent in psychiatric treatment; one of the other two was matched to the target child on IQ and socioeconomic status and the remaining same-sex child was randomly selected.

[4] For grades one to three and others where the teacher indicated reading problems, the research assistants read each question aloud. A guide sheet was also used for the younger children so that only one item at a time was exposed. The research assistants checked to make sure that the children were following the instructions and helped those having difficulty.

students were asked to rate themselves by circling the items that were self-descriptive.

Test-Retest

Four additional classes (two third and two sixth grade) from a predominately white middle-class suburban school were tested and then retested 2 weeks later. One class of both boys and girls at each grade level rated boys and the other class rated girls. The same instructions and procedure for administering the questionnaire as described above were used for both the test and retest, with one additional comment on the retest: "We would like you to answer this questionnaire again." The children included in the rater group differed slightly from first to second testing because of some absences on each testing, but the number of raters was 90 on both tests with 53 students being rated.

RESULTS

An item by subject matrix was obtained which indicated the number of nominations each child received from his or her peers for each item. In order to allow comparability between classes of different sizes, these frequencies for each item were converted to percentage scores based on the total number of girl and boy raters in each class. These percentages were used in subsequent analyses.

Factor Analysis of the Items

The 35 items rated by male peers and the same items rated by female peers were each analyzed using a principal components analysis. Analyses were terminated by the occurrence of one of three conditions: if a component root fell below unity, if a component root did not increase the variance of the extracted roots by at least 10%, or if 100% of the variance was extracted. Varimax (orthogonal) rotations were performed an all extracted components.

Each analysis produced four factors accounting for approximately 65% of the variance. The rotated factor matrices are presented in Table I. The percent of variance accounted for by the first three factors was 37.8, 15.2, and 7.5 for male raters, and 37.4, 16.3, and 7.6 for female raters. The first factor consists of items which describe classroom disruption, physical aggression, and attention-seeking. The items in the second factor suggest social withdrawal, shyness, and oversensitivity. The third factor contains items of popularity and competence. Factors 1, 2, and 3 were labeled Aggression, Withdrawal, and Likeability, respectively. The fourth factor accounts for only a small proportion of the variance (less than 4%) and contains items which, upon inspection, do not suggest a clear label; consequently, factor four was excluded from subsequent analyses and discussion.

Table I. Rotated Principal Components Factor Analyses

Item	Factor (male peers)					Factor (female peers)				
	I	II	III	IV	h²	I	II	III	IV	h²
1. Those who are taller than most.	06	05	14	73	56	10	12	17	-36	18
2. Those who help others.	-13	-03	78	-01	63	-23	-08	79	-11	70
3. Those who can't sit still.	75	02	-06	-32	67	72	-07	-07	48	75
4. Those who try to get other people into trouble.	82	-04	-19	03	70	87	-03	-17	-05	78
5. Those who are too shy to make friends easily.	-19	75	-07	08	61	-31	72	-05	-04	62
6. Those whose feelings are too easily hurt.	10	58	08	-06	36	05	64	14	-07	44
7. Those who act stuck-up and think they are better than everyone else.	79	-07	02	18	66	78	-17	-09	-25	70
8. Those who play the clown and get others to laugh.	74	-11	17	-26	66	72	-13	16	41	73
9. Those who start a fight over nothing.	80	01	-01	17	67	83	04	-07	-19	73
10. Those who never seem to be having a good time.	-09	75	-21	11	62	-10	71	-13	-07	54
11. Those who are upset when called on to answer questions in class.	43	49	-19	06	46	25	60	-12	03	44
12. Those who tell other children what to do.	77	-13	11	19	66	77	-15	08	-15	65
13. Those who are usually chosen last to join in group activities.	13	70	-13	-12	54	14	70	-27	09	59
14. Those who are liked by everyone.	04	-16	77	11	63	03	-23	80	02	69
15. Those who always mess around and get into trouble.	86	-12	-14	-04	77	85	-16	-16	23	83
16. Those who make fun of people.	83	-09	-06	12	71	81	-16	-05	07	68
17. Those who have very few friends.	10	82	-10	09	70	-04	78	-16	01	64
18. Those who do strange things.	71	16	02	-18	56	64	12	-05	47	65

Item	Factor (male peers)					Factor (female peers)				
	I	II	III	IV	h²	I	II	III	IV	h²
19. Those who are your best friends.	06	-23	59	-11	56	-05	-12	79	-01	65
20. Those who bother people when they are trying to work.	85	05	-09	-09	74	83	-01	-14	24	76
21. Those who get mad when they don't get their way.	85	-01	-09	03	72	82	-02	-09	-12	70
22. Those who don't pay attention to the teacher.	77	10	-23	-10	66	76	07	-21	33	79
23. Those who are rude to the teacher.	85	-03	-17	-09	75	83	-04	-11	21	75
24. Those who are unhappy or sad.	-04	78	-11	06	62	-14	80	-08	-05	68
25. Those who are especially nice.	-29	-05	79	05	73	-29	-10	84	-01	80
26. Those who act like a baby.	68	28	-20	-14	60	66	35	-17	14	61
27. Those who are mean and cruel to other children.	86	02	-09	18	78	83	-02	-15	-20	74
28. Those who often *don't* want to play.	09	68	00	-12	48	03	69	01	-02	54
29. Those who give dirty looks.	82	09	-19	01	71	82	-01	-11	01	69
30. Those who want to show off in front of the class.	88	-07	01	-10	80	83	-17	-04	27	79
31. Those who say they can beat everybody up.	71	-11	02	48	76	78	-12	-10	-30	73
32. Those who aren't noticed much.	-20	81	-05	01	70	-27	74	-19	-03	67
33. Those who exaggerate and make up stories.	75	12	-01	14	60	72	10	-09	02	53
34. Those who complain, nothing makes them happy.	71	25	-07	17	59	65	26	-02	-15	61
35. Those who always seem to understand things.	-20	-05	73	16	60	-19	-11	75	-16	63
Sum of squares	13.0	4.9	3.2	1.3		12.6	5.1	3.7	1.5	
Percent of total variance	37.8	15.2	7.5	3.6		37.4	16.3	7.6	3.8	

Table II. Means and Standard Deviations for Peer Rating Factor Scores by Sex and Grade

	Score (mean ± SD)				
	Males		Females		
Factor	Grades 1–5 ($N = 107$)	Grades 6–9 ($N = 68$)	Grades 1–5 ($N = 88$)	Grades 6–9 ($N = 77$)	Total sample ($N = 340$)
Aggression (male raters)	23.7 ± 17.6	20.2 ± 16.6	19.2 ± 14.0	17.5 ± 14.4	20.4 ± 16.0
Aggression (female raters)	24.4 ± 18.1	20.5 ± 18.1	18.4 ± 14.1	14.8 ± 13.2	19.9 ± 16.5
Withdrawal (male raters)	17.6 ± 11.5	13.8 ± 11.7	16.0 ± 11.6	16.4 ± 13.8	16.2 ± 12.2
Withdrawal (female raters)	16.5 ± 11.0	14.4 ± 13.6	17.3 ± 11.3	14.6 ± 13.6	15.8 ± 12.3
Likeability (male raters)	37.8 ± 16.3	33.3 ± 12.7	33.1 ± 14.8	29.4 ± 15.3	33.8 ± 15.3
Likeability (female raters)	20.9 ± 18.1	15.0 ± 12.6	33.0 ± 18.2	30.3 ± 18.8	25.0 ± 18.7

Factor Scale Scores

Based on the rotated component matrix (excluding the fourth component), items with loadings having an absolute value greater than or equal to .40 were selected for construction of factor scale scores. Items with significant positive loadings were equally weighted; items with loadings with significant negative loadings were reflected and equally weighted. An individual's score on a factor was then represented as the mean of items (and reflected items) which had significant loadings on the given factor. Factor scale scores computed with zero–one weights typically correlate highly with those calculated by fractional weights (Gorsuch, 1974); consequently, the former were chosen for ease of scoring. Factor scale scores were computed for both female and male raters on each factor, producing a total of six factor scale scores per subject. Means, standard deviations, and sample sizes for the factor scale scores are presented in Table II by sex of child and grade level (1–5, 6–9).

Factor scale scores did not correlate highly with each other. For male and female raters the correlations of the Aggression and Withdrawal scores were close to zero (r = .08 and .01, respectively). The Likeability scale scores correlated modestly with Aggression (for male raters r = −.23 and for females r = −.32) and Withdrawal (male raters r = −.22, female raters r = −.26).

A 2 (Sex) × 2 (Grade Level; 1st–5th vs. 6th–9th) analysis of variance with unweighted means was performed on each of the six factor scale scores (three for male raters and three for female). Main effects for sex were found for Aggres-

sion with males rated higher than females by both male ($F = 4.12$, $df = 1/336$, $p < .05$) and female ($F = 10.70$, $df = 1/335$, $p < .002$) raters. Main effects for Sex were also found for Likeability; males were rated higher than females when rated by males ($F = 6.55$, $df = 1/336$, $p < .001$), and lower when rated by females ($F = 50.70$, $df = 1/335$, $p < .001$). Grade Level main effects were found for female-rated Aggression ($F = 4.40$, $df = 1/335$, $p < .05$) and for male ($F = 6.13$, $df = 1/336$, $p < .02$) and female-rated Likeability ($F = 5.00$, $df = 1/335$, $p < .05$); the sixth- through ninth-grade group was lower than the first- through fifth-grade group in all three cases. No significant interactions between Sex and Grade Level were found for any of the factor scale scores.

Internal Consistency for the Factor Scale Scores

For each factor, items with significant factor loadings were divided into two arbitrary halves. As a measure of internal consistency, the mean of half the significant items in a given factor was correlated with the mean of the other half of the significant items. The correlations were all above .70 across factors, raters, and grade and sex of children being rated, with the exception of Likeability for males in seventh through ninth grade (.588 for male raters and .682 for female raters). Correlations for Aggression are mostly over .90, a significantly higher value than those for the Withdrawal and Likeability factors. Fewer items in the Withdrawal and Likeability factors may account for this difference.

Interrater Agreement

Interrater agreement between males and females was computed by correlating each male-rated factor scale score with each corresponding female-rated factor scale score. The interrater correlations, presented in Table III, were computed for three grade groupings (1–3, 4–6, 7–9) within each sex of child being rated. All interrater correlations are greater than .60 with the exception of

Table III. Factor Score Agreement Between Male and Female Peers

	Aggression	Withdrawal	Likeability	N
Males				
Grades 1–3	.820	.619	.709	64
Grades 4–6	.921	.780	.621	79
Grades 7–9	.897	.739	.250	38
Females				
Grades 1–3	.753	.611	.751	46
Grades 4–6	.777	.793	.640	66
Grades 7–9	.847	.854	.653	59

Table IV. Test–Retest Correlations for Factors by Sex of Rater

	Factor		
	I	II	III
Sex of rater	(Agression)	(Withdrawal)	(Likeability)
Male	.864	.893	.813
Female	.953	.913	.818

Likeability rated for seventh- through ninth-grade males. No Aggression correlation falls below .75, with the highest (.92) found for fourth- through sixth-grade males. The Withdrawal correlations range from .61 to .85, and the Likeability correlations range from .25 to .75 with the second lowest correlation being .62.

Interrater correlations for items were also computed. Median agreement between male and female peers on items for the total sample of children was .607 (interquartile range .545 to .697). Further, there was adequate consistency across both sex and grade groupings. Males were rated with a median agreement of .668 (interquartile range .551 to .729), and females with a median agreement of .577 (interquartile range .526 to .650). First through fifth graders were rated with a median agreement of .609 (interquartile range .524 to .704) and sixth through ninth graders were rated with a median agreement of .622 (interquartile range .548 to .722).

Test–Retest

Factor and item test–retest correlations were obtained for the test–retest sample of two third- and two sixth-grade classes. The correlations for the three factors for male and female raters are found in Table IV. All the factor test–retest correlations are greater than .80. For the items rated by males, the median test–retest correlation was .711 (interquartile range .596 to .778); the median for female raters was .760 (interquartile range .676 to .827).

Concurrent Validity

As a measure of concurrent validity, teacher and self-ratings were each correlated with peer ratings (see Table V and VI). The teacher–peer and self–peer correlations are based on factor scale scores (for Aggression, Withdrawal, and Likeability) derived from the factor analysis of the peer items.

All of the teacher–peer correlations are positive and reliable ($p < .025$), ranging from .278 to .729 with a median of .566. All of the self–peer correla-

Table V. Correlations Between Teacher and Peer Rating Factors

	Aggression		Withdrawal		Likeability		
	Male raters	Female raters	Male raters	Female raters	Male raters	Female raters	N
Males							
Grades 1–5	.586	.673	.344	.287	.604	.523	105
Grades 6–9	.729	.669	.549	.550	.541	.278	65
Females							
Grades 1–5	.612	.609	.510	.505	.607	.492	85
Grades 6–9	.667	.625	.582	.628	.524	.488	75
Total sample	.633	.640	.465	.465	.562	.453	330

Table VI. Correlations Between Self- and Peer Rating Factors

	Aggression		Withdrawal		Likeability		N
	Male raters	Female raters	Male raters	Female raters	Male raters	Female raters	
Males							
Grades 1–5	.390	.425	.399	.402	.536	.346	101
Grades 6–9	.459	.522	.199	.183	.386	.317	61
Females							
Grades 1–5	.453	.473	.282	.467	.213	.113	83
Grades 6–9	.464	.546	.380	.590	.096	.088	67
Total sample	.433	.476	.300	.404	.362	.212	312

tions are positive and most are reliable ($p < .025$) with the exception of the Withdrawal factor rated for males in grades six through nine and the Likeability factor rated for females. The self–peer correlations range from .088 to .590 with a median of .394. The Aggression factor produced higher correlations than the other two for both teacher–peer and self–peer relationships.

For each of the three factors, the teacher–peer correlations are higher ($p < .01$) than the self–peer correlations. The median teacher–peer correlations for the Aggression, Withdrawal, and Likeability factors were .646, .529, and .523, respectively, as compared with the median self–peer correlations of .461, .389, and .265. Within each factor, there are some differences between ratings by male peers and ratings by female peers. For Likeability, the ratings by male peers correlate higher ($p < .01$) with both teacher (median .472) and self-ratings (median .599) than do the ratings by female peers (medians .490 and .215). For Aggression and Withdrawal, self-ratings for the total sample correlated significantly higher ($p < .01$) with ratings by female peers (.476 and .404) than with ratings by male peers (.433 and .300). With regard to grade group, there are two particularly clear trends for the Withdrawal factor: For the higher grade group (6–9), teacher–peer correlations are higher (median .566 vs. median .424, $p < .01$) and the self–peer correlations for males are lower (.191 vs. .400, $p < .01$).

DISCUSSION

Each of the three components of the Pupil Evaluation Inventory — Aggression, Withdrawal, and Likeability — appears to be pure, homogeneous, and stable. All of the items composing these factors have substantial and unique loadings on the factors they compose; only one item of the 35 male and 35 female items has a loading greater than .40 on more than one factor. Temporal reliability is high across all factors for the subsample tested. There is a tendency for higher internal consistency and concurrent validity for the Aggression factor. This may be due to the greater number of aggression items.

The correspondence between teacher and peer ratings is considered evidence for concurrent validity of the Pupil Evaluation Inventory. These modest correlations are consistent with the findings of other investigators who have reported relationships between teacher and peer ratings (Lesser, 1959; Walder et al., 1961; Wiggins & Winder, 1961). The lower correlations between self- and peer ratings compared to teacher–peer ratings replicates other findings (e.g., Eron et al., 1971) and was expected given that self-ratings are a generally poor method of assessment, heavily influenced by a tendency to respond with pro-social ratings (Eron et al., 1971; Himelstein & Lubin, 1965). The higher self–

peer and teacher—peer correlations found for the Aggression factor probably reflects that factor's greater reliability.

Our major goal was to develop a peer rating scale appropriate for use with both males and females of a wide age range. The utility of the Pupil Evaluation Inventory for assessing behavior of both sexes is supported by substantial inter-rater agreement between males and females and similarities in factorial structure, temporal reliability, internal consistency and teacher—peer correlations between males and females. The instrument is also suitable for a wide age range; internal consistency, sex of rater agreement, and teacher—peer correlations are little affected by grade level. The major exception is the Likeability factor at the highest grade level. Here there was low agreement ($r = .250$) between males and females rating males, indicating different criteria for likeability in evaluating males, but not females, at this age. The correlation of teacher ratings with those of females rating males at the 7—9 grade level also was lower than other correlations. Thus females' criteria for rating males at this age differs from teachers as well as their male peers.

In conclusion, the results support the potential utility of the Pupil Evaluation Inventory for research on social adjustment with both males and females ranging from grades one through nine. The Aggression factor is generally more reliable than the Withdrawal and Likeability factors, but all factors are homogeneous and temporally stable and have concurrent validity as measured by correlations with teacher rating.

REFERENCES

Bardon, J. I., & Bennett, V. D. Helping children in school. In B. B. Wolman (Ed.), *Manual of child psychopathology*, New York: McGraw-Hill, 1972.

Bower, E. M. The school psychologist's role in their identification and adjustment of socio-emotionally handicapped children. In J. I. Magary (Ed.), *School psychological services in theory and practice: A handbook.* Englewood Cliffs, New Jersey: Prentice-Hall, 1967.

Bower, E. M. *Early identification of emotionally handicapped children in school.* Springfield, Illinois: Charles C Thomas, 1969.

Burton, R. V. Validity of retrospective reports assessed by the multitrait—multimethod analyses. *Developmental Psychology Monographs*, 1970, (3, Pt. 2).

Buss, A. H. Personal communication, 1967, cited in L. D. Eron, L. O. Walder, and M. M. Lefkowitz, *Learning of aggression in children.* Boston: Little, Brown & Co., 1971.

Eron, L. D., Walder, L. O., & Lefkowitz, M. M. *Learning of agression in children.* Boston: Little, Brown & Co., 1971.

Gorsuch, R. L. *Factor analysis.* Philadelphia: W. B. Saunders Co., 1974.

Havighurst, R. J., DeHaan, R. F., Dietrich, W. J., Hackamack, H., Johnson, L., & King, R. D. A commininity youth development program. *Supplementary Education Monograph*, 1952, No. 75.

Himelstein, P., & Lubin, R. Attempted validation of the self-disclosure inventory by the peer nomination technique. *Journal of Psychology*, 1965, *61*, 13-16.

Leibowitz, G. Comparison of self-report and behavioral techniques of assessing aggression. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 21-25.

Lesser, G. S. The relationships between various forms of aggression and popularity among lower class children. *Journal of Educational Psychology*, 1959, *50*, 20-25.

Minturn, M., & Lewis, M. Age differences in peer ratings of socially desirable and socially undesirable behavior. *Psychology Reports*, 1968, *23*, 783-791.

Mitchell, J. V. The factor analysis of a "guess-who" questionnaire assigned to identify significant patterns in children. *Journal of Personality*, 1956, *24*, 376-386.

Neale, J. M., & Weintraub, S. Children vulnerable to psychopathology: The Stony Brook high-risk project. *Journal of Abnormal Child Psychology*, 1975, *3*, 95-113.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, *66*, 574-583.

Roff, M., & Sells, S. Juvenile delinquency in relation to peer acceptance—rejection and socioeconomic status. *Psychology in the Schools*, 1968, *5*, 3-18.

Rolf, J. E. The academic and social competence of school children vulnerable to schizophrenia and other behavior pathologies. *Journal of Abnormal Psychology*, 1972, *80*, 225-243.

Sells, S. B., & Roff, M. *Peer acceptance—rejection and personality development.* Final report of a five year research program, including studies of the origins of acceptance—rejection, effects on early delinquency and drop-out, and developmental dynamics. U. S. Office of Education, January, 1967.

Siegelman, M. Psychometric properties of the Winder and Wiggins peer nomination inventory. *Journal of Psychology*, 1966, *64*, 143-149.

Smith, G. Usefulness of peer ratings of personality in educational research. *Educational and Psychological Measurement*, 1967, *24*, 967-984.

Walder, L. O., Abelson, R., Eron, L. D., Banta, T. J., & Laulicht, J. H. Development of a peer-rating measure of aggression. *Psychological Reports*, 1961, *9*, 497-556 (monograph supplement 4-49).

Wiggins, J. S., & Winder, C. L. The peer nomination inventory: An empirically derived sociometric measure of adjustment in preadolescent boys. *Psychological Reports*, 1961, *9*, 643-677.

Winder, C. L., & Wiggins, J. S. Social reputation and behavior: A further validation of the Peer Nomination Inventory. *Journal of Social Psychology*, 1964, *68*, 681-684.