SELECTING A COST-EFFECTIVE SCREENING MEASURE FOR THE
ASSESSMENT OF PRESCHOOL SOCIAL WITHDRAWAL

CHARLES R. GREENWOOD, HILL M. WALKER, NANCY M. TODD,
AND HYMAN HOPS

CENTER AT OREGON FOR RESEARCH IN THE BEHAVIORAL EDUCATION FOR
THE HANDICAPPED, UNIVERSITY OF OREGON

The utility of teacher judgment (rankings, ratings) and peer judgment (sociometric ratings) were studied as screening variables for preschool social withdrawal/responsiveness. Observational measures of preschoolers' interaction were used as the validation criterion based upon theoretical, empirical, and practical considerations related to development of a multipurpose behavioral assessment system. Results indicated that teacher rankings of students' verbal interaction frequencies (1) were most highly correlated to interaction rate in two preschool settings, (2) had test-retest reliability consistently above $Rho = .80$, and (3) were useful in that 77% of teachers using this procedure could identify their least socially responsive student within five rankings. The peer nomination sociometric rating was the least reliable measure at retest ($Rho = .35$) and showed consistently the lowest correlation with observational indices of interaction partner preferences, *e.g.*, reciprocal rate ($Rho = .29$ to $.39$). The tendency for relationships among screening measures and the criterion to increase at retest suggested a training effect on screening accuracy. The use of teacher rankings in combination with observational measures to confirm selection accuracy and for monitoring treatment progress is discussed as a cost-effective behavioral assessment procedure for preschool social withdrawal/responsiveness.

DESCRIPTORS: Social interaction, withdrawal, isolate behavior, screening procedures, preschool children

The problem of how children first come to the attention of behavior assessors, surprisingly, has been the least investigated aspect of behavioral assessment (Cone and Hawkins, 1977; Mash and Terdal, 1974, 1976). The majority of published behavioral research studies report data beginning at baseline. How, why, and through what means does a child's behavior first become viewed as problematic or potentially problematic? The lack of behavioral inter-

est in assessment at this level may be directly attributable to the questionable utility of referral or screening information for behavioral intervention because these measures do not directly pinpoint what should be treatment targets, suggest controlling stimuli, or identify maintaining consequences. Behavior analysts have been content to allow traditional procedures, operative in their settings, *e.g.*, teacher referral, psychological evaluations, case conferences, to handle the screening of clients. Also, assessments at first screening or referral exist at lower levels of measurement, *i.e.*, nominal, ordinal, interval (Guilford, 1954, p. 11-16). A teacher referral, for example, is measurement at the nominal level, wherein a student's performance is described, *e.g.*, shy or withdrawn, based strictly upon social judgment. In contrast to individual case referral, screening has been described as a process by which *all* children's behavior comes in contact with assessment for the purpose of

This work was performed pursuant to grant OEG07-07-7-7443 from the Bureau of Education of the Handicapped, U.S. Office of Education, HEW. The authors would like to acknowledge the contribution of Debra Hudson, Lloyd Maxfield, the observer team, and the cooperating preschool teachers for their invaluable assistance completing this research. Thanks are also due to Marleen Elliott and Debi Millsap for preparation of versions of this manuscript. Reprints can be obtained from Charles R. Greenwood, Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas, 1980 North Second Street, Kansas City, Kansas 66101.

locating individuals who may benefit from more detailed assessment. Screening measures have been described as methods of "wide band and narrow fidelity" (Cronbach, 1970) and are likely rankings or ratings at the ordinal or interval level of measurement. This implies that many characteristics may be assessed but that measurement accuracy or reliability may not be sufficient for use in other assessment roles, *e.g.*, monitoring progress.

Recently, however, interest in behavioral assessment and growth in the development of packaged behavior program technology have focused attention to the problem of initial contact assessment, particularly with socially withdrawn and socially aggressive children (Greenwood, Walker, and Hops, 1977; Walker, Greenwood, Hops, and Todd, 1979; Hops, Fleishman, Guild, Paine, Walker, and Greenwood, Note 1; Walker, Street, Garrett, Crossen, Hops, and Greenwood, Note 2). In these service packages, procedures are required for initial referral, case intake, and screening (Greenwood, Hops, Walker, Guild, Stokes, Young, Keleman, and Willardson, 1979; Hops, Walker, Fleishman, Nagoshi, Omura, Skindrud, and Taylor, 1978). Behavior ratings, for example, have been used following teacher referral, prior to the use of more time-consuming and costly behavioral observations in the natural setting (Hops *et al.*, Note 1; Walker *et al.*, Note 2).

The ideal screening assessment measure must have the following characteristics: (1) it can be administered to *all* students economically, (2) personnel can be easily trained to implement the measure in the natural setting, (3) the measure bears a substantial direct relationship to the behavior it predicts, (4) the instrument must be reliable, and (5) the instrument must be used in continuity with other assessment measures to complete problem definition, monitoring, and follow-up assessment functions essential to effective treatment (Greenwood *et al.*, 1977; Hops and Greenwood, *in press*).

The present study was designed to identify and evaluate screening measures for social with-

drawal and limited social responsiveness in preschool children that could be effectively used by teachers. This work was completed as a component of research developing a complete social assessment manual for preschool teachers (Greenwood, Todd, Walker, and Hops, Note 3.) Screening measures studied were indices of *teacher judgment* in the form of rankings and ratings and *peer judgment* using sociometric nomination. These measures were evaluated against direct observation of children's interactive behavior in preschool settings.

The specific purposes of this study were to (1) test the ease of administration of each measure, (2) assess the reliability of each measure, (3) determine the concurrent validity of the screening measures with actual levels of interaction in the classroom, and (4) determine the accuracy by which withdrawn students could be identified using these measures. This study was designed to improve upon prior studies of teacher and peer judgment methods by including (a) three measurement sources, within the same study, *i.e.*, peers, teacher, and direct observation (Greenwood *et al.*, 1977), (b) a large *N* based upon 299 preschool students and 20 classroom units, (c) observation data defined and recorded to ensure representation of reciprocity in child-child interactions (Strain and Shores, 1977), and (d) observational data collection in two distinct classroom structure/settings to assess the generality of the results.

METHOD

Subjects and Settings

The subjects were 299 preschoolers, (49% male and 51% female) 3 to 7 years of age, attending 20 private, parochial, and community preschools in and around the Eugene-Springfield, Oregon area (pop. 200,000). Over 95% of the children were from white, middle to lower class families. Across the total sample, the ages of 3, 4, 5, 6, and 7 accounted for 6%, 29%, 35%, 25%, and 5% of subjects, respectively.

Teacher Judgment Screening Measures

Three screening instruments were developed for use by the preschool teachers. These were two rankings and one rating scale.

Teacher verbal interaction frequency ranking. Teachers rank ordered their students on the basis of verbal interaction frequency during the day. Research has shown that normal preschoolers are highly verbal and frequently use grammatically correct language forms (Mueller, 1972; Mueller and Brenner, 1977; Greenwood, Walker, Todd, and Hops, Note 4). Low ranking on verbal interaction could be indicative of social and possibly language delay. The child assigned rank 1 was the child who talked with the other children the least number of times, while the child who was assigned the largest number was the child who talked with the other children the most number of times.

Teacher popularity ranking. Each child was also assigned a popularity rank by the teacher, *i.e.*, the child assigned rank number 1 was the child in the class who students most liked to play with. The child assigned the highest number was the student children least liked to play with. The scaling of the verbal interaction and popularity rankings was reversed to ensure a unique rank by each teacher.

Teacher social behavior rating scale. A 21-item social behavior rating scale developed by Greenwood *et al.* (Note 3) was completed for each student by the teacher. Each behavior descriptive item was marked on a 7-point bipolar rating differential. Two subscale scores (negative rating and positive rating) were generated reflecting the average teacher rating of the child's (a) problem social behaviors (9 negative items) and (b) social assets (12 positive items).

Peer Nomination Sociometric Screening Measure

The peer nomination sociometric reported by Marshall and McCandless (1957) and more recently by Gottman (1977a) was selected for

use in the present study. This was done specifically to replicate other researchers' use of the sociometric in the assessment of social isolation (Gottman, Gonso, and Schuler, 1976).

Using black and white polaroid pictures of classmates displayed on a tack board (48 cm by 72 cm), each child was privately asked three questions by the third author and requested to point to pictures of a child in compliance with each direction, which were: (1) "Can you point to your picture on this board?" (2) "What is this child's name?" (The question was repeated in relation to each of the child's classmates), and (3) "Point to the *three* children you like to play with best." After finishing an interview, the investigator redistributed the pictures on the board ensuring a fair test for the next child. All children in all 20 classrooms were tested in an identical manner. As in Gottman (1977a) and Tiktin and Hartup (1965), the number of total nominations received by each child served as the peer *acceptance* score for that child. Because human subjects review procedures at the University of Oregon did not permit the use of sociometric rejection questions, no rejection score was available in this study.

Behavioral Observations of Social Interaction—The Criterion Measure

A simplified behavioral observation system (Todd, Note 5) for recording the reciprocal interaction of preschool children was developed based upon an earlier system (Hops, Todd, Garrett, and Stokes, Note 6). The system was designed with future use by preschool teachers in mind. An interaction was defined as occurring each time children communicated or exchanged signals verbally, nonverbally, or physically. The system did not differentiate between topographic social responses exhibited within an interaction but only the total frequency with which each child originated or participated in interactions.

Interactions were comprised of initiations, response or no response, continuing responses, and termination. (a) An *initiation* was the first at-

tempt to communicate by an individual beginning the interaction. (b) A *response* was any observable signal or reply to the initiation. (c) *Continuing responses* were all additional interactive behavior exchanges following the original initiation and the response. (d) *Termination* occurred when both individuals stopped using any form of communication for longer than 5 sec. Only the originating initiation and response components of the interaction were recorded by observers because it was not economically possible to collect data on continuing responses for all students in a classroom.

When an initiation-response exchange occurred, identification (ID) numbers worn by participating children were recorded in coding boxes on observation forms: the initiator's ID number first, followed by the respondent's ID number. If the respondent showed no visible behavior in response to an initiation, a slash (/) was drawn through the respondent's ID number. Examination of each coding box yielded information concerning the following subject interaction variables: (a) the frequency of initiations to peers, (b) the frequency of interaction, *e.g.*, initiations responded to by peers, (c) the frequency of no response to an initiation, and (d) who interacted with whom.

General observation procedures involved assigning a minimum of one observer for every 10 children (*e.g.*, three observers were assigned to a class of 21). On the average, one observer coded the interactive behavior of seven children but never more than 10. When two or more observers coded simultaneously in one classroom, they divided the children into approximately equal groups by location—one group for each observer. Each observer was then responsible for coding all children's interactive behavior within these prearranged boundaries as it occurred. All children's interactions were coded within each 15-min observation period.

Coding complexity. A preliminary evaluation of the system's coding complexity was carried out (Greenwood, Walker, Todd, and Hops, Note 7). The question evaluated was "Does an

observer, responsible for coding interactions for 10 children, record the same interaction rates for one specific child as does an observer coding just that child?" On 13 separate occasions a calibrating observer recorded the behavior of one child while 10 were being observed. Observers knew that checks were being made but did not know which target child was being observed by the calibrating observer or that only one child was being recorded. The correlation between the two sets of data was .84. A correlated *t* test indicated no significant differences between the two observation methods ($\bar{X}_1 = 13.32$ interactions/min *vs.* $\bar{X}_2 = 13.61$ interactions/min). These results indicated that the system's coding complexity (1 observer/10 children) did not appreciably reduce coding accuracy over that normally produced when one child was being observed.

Observer training. Observer trainees read an observation manual (Todd, Note 5) and took unit mastery tests corresponding to the chapters in the manual. Videotapes were used to practice recognizing and coding interactions. Observers roleplayed preschoolers and trainees coded their interactions. Field training consisted of observers entering actual preschool classrooms and simultaneously coding children's behavior with an observer trainer coordinator. A criterion of 85% for sequential agreement (correct initiator-responder sequence) and 90% for interaction agreement (total interactions recorded) had to be achieved for three consecutive 15-min agreement checks for an observer to be considered reliable and ready for actual data collection.

Observers knew that the study was of children's social interaction but did not know of the research interest in withdrawn, unresponsive children or in teacher accuracy predicting children's interaction. Observer and teacher interactions beyond greetings were prohibited. Questions and information regarding observation problems or procedures with teachers were handled by an observer trainer coordinator.

Interobserver agreement. On a daily basis, simultaneous observations between observers

were scheduled between two, and sometimes three, observers. All checks lasted for 10 min, representing $\frac{3}{4}$ of the usual 15-min observation session. Quasi-random checks between observers prevented idiosyncratic differences arising between observer pairs from systematically influencing measurement (Johnson and Bolstad, 1973).

Percent agreement was computed by counting the number of adjacent initiator-responder entries that were in agreement. Agreements were divided by the maximum number of interactions coded by either observer and then multiplied by 100 to form a percentage score. The average percentage of agreement for number of interactions was 95%, $SD = 7\%$. Equivalent values for the correct interactive sequence (who initiated *vs.* who responded) was 96%, $SD = 9\%$. The overall Pearson correlation between observer pairs, using rate of interaction as the unit of analysis, was .99.

Test-retest reliability. A prior study of the test-retest reliability of the coding system was conducted using an odd-even day, split-half procedure over 20 days. Interaction rate means formed in this manner correlated .75. A *t* test indicated no differences between halves computed in this manner, $t(426) = .493$, $p > .10$. Odd-even mean rates were .666 *vs.* .662, respectively (Greenwood *et al.*, Note 7).

Observation scores. Four scores were computed for each child. These scores were: (a) the rate of interactions/min (rate), (b) the average proportion of peer's interactions in which the subject participated as the initiator or responder (shared interaction), (c) the average proportion of peer's interactions initiated to the subject (shared initiated interactions), and (d) the number of individual peers with whom the subject interacted (contacts).

The first score was a measure of interaction frequency and quantity. Scores b through d reflected *content* analogous to the nomination sociometric scores (Marshall and McCandless, 1957). High scores on these measures indicated that the subject tended to interact with many

peers (b and d) and/or the extent to which peers sought out the subject for interaction (c).

Design and Procedures

The teacher and peer judgment measures were completed twice for each student in each class. Following the first teacher-peer judgment phase, observation data were collected on alternate days for a maximum of 10, 15-min sessions during a free play period and during a more structured assigned task period (30 min total/day). During free play periods, the children were free of teacher-imposed structure, and opportunities for child-child interactions were at a maximum. In the assigned task setting, children were seated and engaged in teacher-directed tasks, *e.g.*, cut and paste.

At the completion of 10 observations in each classroom, the retest administration of the teacher and peer measures was scheduled for each student (1 month after the first test). The retest phase was a complete replication of procedures used in the first test. In both administrations, teachers were counterbalanced on the order in which they completed the two rankings and the social behavior rating.

RESULTS

Results are reported in three major areas: (a) reliability and stability within measures, (b) convergence of measures with the observational criterion and (c) accuracy of measures in accounting for interaction and in selecting withdrawn students. The choice of parametric or nonparametric correlational procedures was settled by computing both. Because there were no appreciable differences in these results using Pearson *r* *vs.* Spearman *Rho* (Siegel, 1956), *Rho* was used as, in the author's opinion, it best met the assumptions implied by the nature of rank and rating data.

Reliability and Stability within Measures

In order to control for mean differences due to classroom membership, test and retest mea-

Table 1
Spearman Intercorrelation Matrix for Observation Measures^a (N = 299)

Rate	FP	1.00							
Shared/Interaction	FP	.82*	1.00						
Shared/Inits	FP	.70*	.90*	1.00					
Contacts	FP	.41*	.51*	.52*	1.00				
Rate	AT	.58*	.49*	.44*	.52*	1.00			
Shared/Interaction	AT	.46*	.58*	.55*	.72*	.72*	1.00		
Shared/Inits	AT	.33*	.47*	.49*	.61*	.60*	.84*	1.00	
Contacts	AT	.43*	.57*	.53*	.46*	.26*	.33*	.27*	1.00
		Rate	Shared/	Shared/	Contacts	Rate	Shared/	Shared/	Contacts
		FP	Inter-	Inits	FP	AT	Inter-	Inits	AT
			action	FP			action	AT	
			FP				AT		

* $p < .001$

^aFP = free play; AT = assigned task; shared/inits = shared initiated interactions

tures were transformed to standard score form with a mean of 0 and standard deviation of ± 1 by class. Correlations were then completed over the entire sample of 20 classes and 299 children with complete data.

Observation criterion measures. Four interaction scores were computed from the observation data taken in free play and in assigned task structures. The relationships among these measures and structure/settings are reported in Table 1.

In free play the largest within-method correlations were among rate, shared interactions, and shared initiated interactions, ranging from .70 to .90. These findings replicated for data collected in the assigned task setting, as equivalent

correlations ranged .60 to .84. Peer contacts were consistently least related to the other observation scores in both settings (ranging from .26 in assigned task to .41 in free play).

Children's interaction scores tended to be consistent across settings as rate, shared interactions, shared initiated interactions, and contacts were positively correlated at .58, .58, .49, and .46, respectively. Thus high interactors in free play tended to be high interactors in assigned task and vice versa.

Teacher measures. Test-retest correlations for the frequency rank, popularity rank, rating positive, and rating negative at 1 month were .81, .79, .82, and .67 (see Table 2). The first three measures showed surprisingly high stability at

Table 2
Spearman Intercorrelation Matrix for Teacher Measures^a (N = 299)

Rank Freq	1	1.00							
Rank Pop	1	-.65*	1.00						
Rating/Pos	1	.79*	-.69*	1.00					
Rating/Neg	1	-.53*	.55*	-.70*	1.00				
Rank Freq	2	.81*	-.67*	.75*	-.47*	1.00			
Rank Pop	2	-.58*	.79*	-.60*	.50*	-.67*	1.00		
Rating/Pos	2	.73*	-.68*	.82*	-.59*	.79*	-.68*	1.00	
Rating/Neg	2	.50*	.57*	-.63*	.67*	-.53*	.57*	-.70*	1.00
		Rank	Rank	Rating/	Rating/	Rank	Rank	Rating/	Rating/
		Freq 1	Pop 1	Pos 1	Neg 1	Freq 2	Pop 2	Pos 2	Neg 2

* $p < .001$

^aRank Freq = teacher verbal interaction frequency ranking;

Rank Pop = teacher popularity ranking;

Rating Pos = teacher social behavior rating scale—items measuring social assets;

Rating Neg = teacher social behavior rating scale—items measuring problem social behavior.

retest, while the negative rating was only moderately stable at retest.

Intercorrelations among teacher measures at Test 1 indicated moderately large interrelationships (range .53 to .79, absolute magnitude). This also replicated at Test 2 (range .53 to .79, absolute magnitude). Teacher rating negative was least related to other teacher measures. Cross time intercorrelations of measures demonstrated a highly uniform pattern among measures at moderately large levels.

Peer sociometric measures. The children were very successful at finding their own picture at the sociometric picture board. They averaged 97%, $SD = 3\%$, and 98%, $SD = 3\%$ at Test 1 and Test 2, respectively, over classes. These data suggest a near perfect self-identification ability, and individual class means tended to confirm these findings.

The ability of each child to name all students from the picture board was a more difficult task. Tests 1 and 2 average class mean percentages indicated that 81%, $SD = 12\%$, and 87%, $SD = 9\%$ of the children could identify by name all children. This ranged from 62% to 100% by class. A correlated t test indicated that identification was significantly improved (by 6%) 4 wk later at retest. These figures demonstrated that within any class, children will not be able to name two to five peers on the sociometric board.

The test-retest correlation among sociometric acceptance scores was .35 ($df = 297$, $p < .001$). Acceptance scores were the frequency of choices received by peers for each subject in each class (Gottman, 1977a, 1977b). Correlations by class were highly variable, ranging from $-.28$ to $.71$. The average class coefficient was .36 ($N = 20$). A closer inspection of the stability of individual child choices on the sociometric was made by making a count of the first three choices repeating in the second test trial. Test-retest agreement ranged from 0.0 to 3.0 using this method. The average agreement was 1.47 ($N = 20$). Averages ranged from 1.02 to 2.17 for individual classes. Within this framework, playmates se-

lected one month later were rechosen about half the time.

Convergence of Teacher, Peer, and Observational Measures across Methods, at Retest, and across Settings

The interrelationship among measurement methods, settings, and test occasions are presented in Table 3.

These results indicate that the largest, most uniform relationships occurred between the teacher screening measures and the observation criterion. These relationships were uniformly replicated both at retest and across two preschool settings.¹ Correlations at retest were also consistently larger ($\bar{X}_{Rho}$ gain = .04) than at Test 1.

Peer sociometric acceptance scores did not highly predict the observational measures. Correlations ranged from .19 to .23 in free play, .19 to .29 in assigned task. Increases were noted (X_{Rho} absolute gain = .11) as with previous measures at Test 2 (see Table 3). The largest correlations among measures occurred between interaction rate and the teacher verbal frequency range (.63) and sociometric acceptance (.39). These correlations were, in fact, significantly different, $t(296) = 3.17$, $p < .01$. The largest correlation involving sociometric acceptance occurred in relation to the teacher popularity ranking, .43 to .57.

Accuracy of Sociometric Choice in Accounting for Interaction Partners

An analysis comparing children's sociometric choices with peers most frequently initiated to and interacted with was conducted (Marshall and McCandless, 1957). A variable was formed

¹To ensure the generality of these relationships, correlations were computed using age and sex factors as breakdown categories on the total sample. The correlations for the 3- and 7-year-olds were least consistent, most likely due to a small N in each case, i.e., 16 and 14 students each. Correlations for 4-, 5-, and 6-year-olds and males and females were nearly identical to values generated for the total sample. No trends emerged by age or sex.

Table 3
Cross Method Spearman *Rho*'s (*N* = 299) Observations

Teacher Ratings/Rankings ^a	Free Play				Assigned Task				Sociometric Acceptance	
	Shared		Shared		Shared		Shared		Socio- metric 1	Socio- metric 2
	Rate	Interaction	Initiated	Contacts	Rate	Interaction	Initiated	Contacts		
Rank Freq.	.57**	.47**	.40**	.30**	.50**	.35**	.28**	.29**	.27**	.41**
Rank Pop.	-.48**	-.41**	-.40**	-.31**	-.42**	-.38**	-.31**	-.21**	-.57**	-.43**
Rating Pos.	.48**	.40**	.33**	.29**	.44**	.33**	.27**	.26**	.27**	.42**
Rating Neg.	-.33**	-.26**	-.25**	-.18**	-.30**	-.19**	-.16*	-.12	-.23**	-.34**
Rank Freq.	.63**	.53**	.48**	.36**	.56**	.43**	.34**	.31**	.31**	.41**
Rank Pop.	-.50**	-.46**	-.45**	-.31**	-.42**	-.40**	-.38**	-.24**	-.45**	-.57**
Rating Pos.	.53**	.46**	.40**	.35**	.54**	.42**	.36**	.26**	.36**	.43**
Rating Neg.	-.35**	-.33**	-.34**	-.28**	-.38**	-.30**	-.30**	-.17*	-.31**	-.41**
Sociometric Acceptance										
Sociometric	.29**	.24**	.23**	.10	.29**	.19**	.20**	.06	—	.35**
Sociometric	.39**	.33**	.35**	.23**	.31**	.31**	.23**	.16*	.35**	—

**p* < .01

***p* < .001

^aRank Freq. = teacher verbal interaction frequency ranking;

Rank Pop. = teacher popularity ranking;

Rating Pos. = teacher social behavior rating scale—items measuring social assets;

Rating Neg. = teacher social behavior rating scale—items measuring problem social behaviors.

for each subject reflecting the proportion of their total initiated interactions spent with three most frequently initiated to peers. An equivalent proportion score was formed using the subjects' three sociometric choices. Finally, a third equivalent score was formed selecting three peers in the class at random. *T* tests were computed to test differences between methods of accounting for interaction. A nonsignificant finding would suggest high agreement between children's choices and the children actually sharing most interactions.

The three peers who shared most of each subjects' initiated interactions in free play actually accounted for 50 to 60% of the subjects' total initiated interactions. The equivalent interaction values for the three sociometric choices, 30%, and random peer selection, 15%, were significantly less ($p < .01$). These differences replicated using the retest sociometric choices, and also held in the assigned task setting (see Table 4).

A final analysis looked at sociometric selection accuracy by assessing agreement between these three methods, *i.e.*, sociometric choice, most initiated to, and random. Complete accuracy between any two methods would equal 3.00. Values less than 3.00 reflect the extent of disagreement. The probability of a sociometric choice also being a most initiated to peer was

only slightly greater than $\frac{1}{3}$, as 1.14 was the average class agreement score at Test 1 and 1.19 at Test 2 in free play, 1.12 and 1.13 were noted in assigned task. Comparing the random method to both the most initiated and the sociometric produced estimates substantially below $\frac{1}{3}$, near .5 to .6. These data supported the prior results indicating the sociometric method of selection was better than random selection in identifying most initiated to peers but at best accurate only $\frac{1}{3}$ the time.

Accuracy of Teacher Verbal Frequency Ranking

The last analysis was completed to assess the accuracy of teacher verbal frequency ranking in predicting low rate, withdrawn students. The results indicate that 23% of teachers could identify the lowest child in their class on their first rank (see Table 5). Given three ranks, 64% could identify their lowest child. By five ranks, 77% of teachers identified the lowest subject.² By eight ranks, 91% had identified their lowest target. One teacher required 11 ranks.

²These findings have been replicated in two samples, one year apart (Greenwood, Walker, Todd, and Hops, Note 9).

Table 4

The proportion of subject's initiated interactions with three peers using three peer selection procedures.

Selection Procedure	Freeplay Setting				Assigned Task Setting			
	Test 1		Test 2		Test 1		Test 2	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Random (3)	.160 *	.084	.154 *	.036	.148 *	.060	.155 *	.052
Sociometric (3)	.290 *	.074	.290 *	.061	.295 *	.177	.297 *	.144
Most Inter- acted (3)	.510	.069	.503	.061	.642	.116	.655	.111
		N = 20 Classes				N = 19 Classes		

**t* test, $p < .01$.

Table 5

Number of teacher ranks required to identify the lowest interacting subject in a class.

<i>No. of Ranks</i>	<i>Freq.</i>	<i>Cum. Freq.</i>	<i>Cum. %</i>
1	5	5	23
2	4	9	41
3	5	14	64
4	0	14	64
5	3	17	77
6	0	17	77
7	3	20	91
8	0	20	91
9+	2	22	100
N		22 ^a	

^aEach teacher contributed only one class.

DISCUSSION

This study was a component of research directed toward development of a cost-effective social behavior assessment system for preschool teachers (Greenwood *et al.*, Note 3). This system is a tool for the identification and remediation of social withdrawal and social unresponsiveness in young children (3 to 7 years). The system was developed to operate within a behavioral framework geared toward intervention by organizing a continuum of assessment functions. The system currently focuses on (1) screening of all children in a preschool class group, (2) identification of skill deficiency areas and determination of behavior change objectives, (3) frequent monitoring of intervention effects during a behavior change program, and (4) assessing effects at follow-up, after intervention (Cone and Hawkins, 1977; Greenwood *et al.*, 1977).

The selection of observation as the criterion measure for withdrawal and social responsivity within this system was based upon theoretical, empirical, and practical considerations. The first consideration was validity. A measure of classroom social interaction was required for the preschool classroom setting. The basic tenet of behavioral assessment is that people should be assessed in terms of what they do within the context in which they do it (Mash and Terdal, 1974). Observational measures when correctly implemented are the most natural measures of

children's interaction within the preschool environment (Walker, 1973). Certainly observational systems vary in their approach to the measurement of social interaction (Hops and Greenwood, *in press*; Strain, Cooke, and Apoloni, 1976) and they are not without their share of methodological problems (Johnson and Bolstad, 1973; Kent and Foster, 1977; Nelson, Kapust, and Dorsey, 1978; O'Leary and Kent, 1973).

A criterion measure was also required that could be practically taught to and used by teachers for daily monitoring of procedures to increase children's social functioning. Repeated assessment is a hallmark of behavioral assessment methods (Cone and Hawkins, 1977; Mash and Terdal, 1974). There has been substantial recent progress in the development of observational methods that can be effectively used by natural social agents in applied settings (see review by Hops and Greenwood, *in press*).

Research has also demonstrated the validity of the global interaction rate measure used in this study's observational code in discriminating comprehensive behavioral differences in children's interactive topography (Greenwood *et al.*, Note 4).³ Children selected as withdrawn (2 standard deviations below the mean interaction rate, Greenwood *et al.*, Note 4) verbalized, initiated to peers, responded to peer initiations, and were initiated to by peers significantly less often than average or high rate interactors. These children also spent significantly more time alone, observing peers but not joining in, removed from peer proximity, and alone with a task, than did higher rate interactors. Many of these specific behaviors have been related to other measures of social disability, *i.e.*, sociometrics and later life maladjustment (Cowen, Pederson, Babigren, Izzo and Trost, 1973; Rolf and Hasazi, 1977; Rolf, Sells, and Golden, 1972) and have been used to describe social

³Previous analysis of these data demonstrated significantly higher interaction rates in free play than in assigned task (Greenwood, Walker, Todd, and Hops, Note 4).

isolation in other research studies (Gottman, 1977a; 1977b; Gottman *et al.*, 1976). Unfortunately there have not been longitudinal research efforts including observational measures of interaction, *i.e.*, rate, in relationship to later life adjustment, as in retrospective studies using sociometrics (Cowen *et al.*, 1973). Research on the longitudinal predictive validity of behavioral measures of social interaction is an obvious gap in our understanding of the long-term effects of withdrawal in children.

The screening variables developed for evaluation herein were based upon similar considerations and their demonstrated use in the literature. Thus teacher judgment in the form of rankings and ratings was selected (Bolstad and Johnson, 1977, Marshall, 1957; McCandless and Marshall, 1957; Jones and Cobb, Note 8). Similarly the picture nomination sociometric for young children was selected for evaluation (Gottman, 1977a, 1977b; Gottman *et al.*, 1976; Marshall, 1957).

Results clearly indicate that the teacher verbal frequency ranking was most effective in locating unresponsive interactors. This was true across the entire distribution of interaction rates ($Rho = .57$) and within each teacher's classroom in selecting the lowest interactor (selection accuracy probability = .77 within 5 ranks). The teacher verbal frequency ranking appeared highly useful for screening large numbers of children. When used in combination with additional observational assessment and observational normative data to confirm the actual suitability of targeted subject for intervention, a cost-efficient system of subject selection appears possible. This is the current model used in the Social Assessment Manual for Preschool Level (Greenwood, Note 3; Hops and Greenwood, *in press*).

A criticism of these results is offered by Cone (*in press*). Cone described a method by content confound in assessment wherein the size of correlations between measures is affected by differences in both (1) the behaviors assessed, *e.g.*, verbal interaction frequency *vs.* total inter-

action rate, and (2) the methods used to make the assessment, *e.g.*, ranking *vs.* observation. In the current study, our objective in particular was to replicate other researcher's use of the sociometric (Gottman, 1977a, 1977b). Because both the sociometric and teacher popularity content items asked for the "most liked or best liked" (similar content) this may explain the .57 correlation obtained between these variables. The high .57 to .63 correlations between teacher verbal frequency ranking and interaction rate could also be similarly explained. Inversely, the correlation (.29) between observation rate and sociometric, where content most differed, was low. Extending this logic, if the sociometric had asked "Who do you most interact, talk, and play with during the day?" perhaps higher correlations would have been found between the sociometric report and the observation data in this study. However, this does not explain the .48 correlation between teacher popularity and interaction rate wherein content and method differed. The differences in young children's reports of "like" *vs.* "who they interact with" await more in-depth study.

These interpretations, however, do not limit the utility of the current findings. In fact teacher verbal frequency ranking accuracy has been replicated in two additional studies, each one year apart, further demonstrating the stability of these procedures (Greenwood, Walker, Todd, and Hops, Note 9). Additional research from Cone's perspective would increase content similarity across methods, likely strengthening the sensitivity of the current screening measures.

The sociometric acceptance and individual peer choices failed to meet acceptable test-retest reliability and selection accuracy requirements and were not considered further as a functional screening variable. The problems with reliability of sociometrics used with preschoolers is not a new finding (Blau and Rafferty, 1970; Dunnington, 1957; Horrocks and Buker, 1951; Marshall, 1957). Asher, Singleton, Tinsley, and Hymel (Note 10) have reported high reliabilities (.80 at one month) using a three-point

rating differential sociometric with preschool children. Rather than nominating three students as in this study, the child rates each child's picture by marking a positive, neutral, or negative face. Moreover, research has not supported strong relationships or high predictability between observation and preschool peer sociometric measure (Gottman, 1977a; Marshall, 1957; Marshall and McCandless, 1957).

A recent practical problem with sociometrics has been human subjects' committees hesitancy to approve their use in some instances, particularly rejection nominations. If this trend increases, use of the instrument in this fashion may become a moot point (Asher *et al.*, Note 10). The use of the nomination procedure to monitor change also requires careful consideration. Experimental-group-only studies will not yield consistent results and risk detecting pre-post changes due to measurement variation.

The best use of sociometric nomination with young children may prove to be in the area of social validation (Kazdin, 1977; Wolf, 1978). These measures can be used in combination with more objective measures to indicate change in variables not measured by the primary measures (Drabman, Spitalnik, and O'Leary, 1973; Drabman, Spitalnik, and Spitalnik, 1974) or with item refinement, to suggest satisfaction with peer relationships.

An interesting result across all measures was the increase in correlations at the retest phase. Since observers were naive with respect to the researcher interest and teachers were counterbalanced to control order effects of ranking and rating, it appeared that training of both teachers and students occurred with no direct effort on the part of the experimenters. Practice, knowledge of the items, and the opportunity to observe students more closely during the one month intervening time were not controlled and likely explain this effect for teachers. On the sociometric, the accuracy of children's picture-peer identification increased significantly by 6% indicating greater familiarity with the peer group. These factors, in combination with the

practice of nominating in Test 1, were likely responsible. Training in aspects of social competence and deviancy as described by Hops and Greenwood (*in press*) could be an effective way to raise screening accuracy. Assessment research should focus on this issue, particularly because it has been noted that when many teachers are asked to refer withdrawn children, they may in fact refer socially aggressive, negative children with whom peers simply refuse to interact (Hops and Greenwood, *in press*).

These findings supporting the accuracy and utility of teacher judgment are not unique in the literature. Teacher judgment research (Jones and Cobb, Note 8; Bolstad and Johnson, 1977) has demonstrated surprising accuracy particularly when teachers are asked to judge behaviorally specific items as opposed to the global constructs or attitudes traditionally assessed by these methods. Unfortunately, research is suggesting that single behavior specific ratings may have low reliability over time (Evans and Nelson, 1977). These problems reflect the new uses of behavioral assessment information that are beginning to be investigated.

REFERENCE NOTES

1. Hops, H., Fleischman, D. H., Guild, J., Paine, S., Walker, H. M., and Greenwood, C. R. *Program for establishing effective relationship skills (PEERS): Consultant manual*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, University of Oregon, Eugene, Oregon, 1978.
2. Walker, H. M., Street, A., Garrett, B., Crossen, J., Hops, H., and Greenwood, C. R. *Reprogramming environmental contingencies for effective social skills (RECESS): Consultant manual*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, University of Oregon, Eugene, Oregon, 1978.
3. Greenwood, C. R., Todd, N. M., Walker, H. M., and Hops, H. *Social assessment manual for preschool level (SAMPLE)*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, 1590 Willamette Street, 1978.
4. Greenwood, C. R., Walker, H. M., Todd, N. M., and Hops, H. *Description of withdrawn children's behavior in preschool settings. (Report No. 40)*. Eugene: Center at Oregon for Research

- in the Behavioral Education of the Handicapped, University of Oregon, Eugene, Oregon, 1978.
5. Todd, N. M. *The frequency interaction recording system (FIRS)*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, 1590 Willamette Street, 1976.
 6. Hops, H., Todd, N. M., Garrett, B., and Stokes, J. *The peer interaction recording system (PIRS)*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, 1590 Willamette Street, 1976.
 7. Greenwood, C. R., Walker, H. M., Todd, N. M., and Hops, H. *Preschool teachers' assessments of social interaction: Predictive success and normative data (Report No. 26)*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, 1590 Willamette Street, 1976.
 8. Jones, R. R. and Cobb, J. A. *Validity of behavioral scores derived from teachers' ratings vs. naturalistic observations*. Paper presented at the meeting of the Western Psychological Association, Anaheim, California, April 1973.
 9. Greenwood, C. R., Walker, H. M., Todd, N. M., and Hops, H. *Cost-effective prediction and screening variables for preschool social withdrawal (Report No. 39)*. Eugene: Center at Oregon for Research in the Behavioral Education of the Handicapped, 1590 Willamette Street, 1978.
 10. Asher, S. R., Singleton, L. C., Tinsley, B. R., and Hymel, S. *The reliability of a rating scale sociometric method with preschool children*. Unpublished manuscript, University of Illinois, Urbana-Champaign, 1977.
- Drabman, R., Spitalnik, R., and Spitalnik, K. Sociometric and disruptive behavior as a function of four types of token reinforcement programs. *Journal of Applied Behavior Analysis*, 1974, 7, 93-101.
- Dunnington, M. J. Behavioral differences of sociometric status groups in a nursery school. *Child Development*, 1957, 28, 103-111.
- Evans, I. M. and Nelson, R. O. Assessment of child behavior problems. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment*. New York: Wiley, 1977.
- Gottman, J. The effects of a modeling film on social isolation in preschool children: A methodological investigation. *Journal of Abnormal Child Psychology*, 1977, 5, 69-78. (a)
- Gottman, J. M. Toward a definition of social isolation in children. *Child Development*, 1977, 48, 513-517. (b)
- Gottman, J., Gonso, J., and Schuler, P. Teaching social skills to isolated children. *Journal of Abnormal Child Psychology*, 1976, 4, 179-197.
- Greenwood, C. R., Hops, H., Walker, H. M., Guild, J. J., Stokes, J., Young, K. R., Keleman, K. S., and Willardson, M. A standardized classroom behavior management program (PASS): Social validation and replication studies in Utah and Oregon. *Journal of Applied Behavior Analysis*, 1979, 12, 235-253.
- Greenwood, C. R., Walker, H. M., and Hops, H. Issues in social interaction/withdrawal assessment. *Exceptional Children*, 1977, 43, 490-499.
- Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- Hops, H. and Greenwood, C. R. Assessment of children's social skills. In E. J. Mash and L. G. Terdal (Eds), *Behavioral assessment of childhood disorders*. New York: Guilford Press, in press.
- Hops, H., Walker, H. M., Fleischman, D. H., Nagoshi, J. T., Omura, R. T., Skindrud, K., and Taylor, J. CLASS: A standardized in-class program for acting-out children. II. Field test evaluations. *Journal of Educational Psychology*, 1978, 70, 636-644.
- Horrocks, J. E. and Buker, M. E. A study of friendship fluctuation of preadolescents. *Journal of Genetic Psychology*, 1951, 78, 131-144.
- Johnson, S. M. and Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds), *Behavior change: Methodology, concepts, and practice*. Champaign, Ill.: Research Press, 1973. Pages 7-68.
- Kazdin, A. E. Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1977, 1, 427-452.
- Kent, R. N. and Foster, S. L. Direct observational procedures: Methodological issues in naturalistic

REFERENCES

- settings. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment*. New York: Wiley, 1977. Pages 279-328.
- Marshall, H. R. An evaluation of sociometric-social behavior research with preschool children. *Child Development*, 1957, **28**, 131-137.
- Marshall, H. R. and McCandless, B. R. A study in prediction of social behavior of preschool children. *Child Development*, 1957, **28**, 149-159.
- Mash, E. J. and Terdal, L. G. Behavior-therapy assessment: Diagnosis, design and evaluation. *Psychological Reports*, 1974, **35**, 587-601.
- Mash, E. J. and Terdal, L. G. *Behavior-therapy assessment: Diagnosis, design, and evaluation*. New York: Springer, 1976.
- McCandless, B. R. and Marshall, H. R. A picture sociometric technique for preschool children and its relation to teacher judgments of friendship. *Child Development*, 1957, **28**, 139-147.
- Mueller, E. The maintenance of verbal exchanges between young children. *Child Development*, 1972, **43**, 930-938.
- Mueller, E. and Brenner, J. The origins of social skills and interactions among play group toddlers. *Child Development*, 1977, **48**, 854-861.
- Nelson, R. O., Kapust, J. A., and Dorsey, B. L. Minimal reactivity of overt classroom observations on student and teacher behaviors. *Behavior Therapy*, 1978, **9**, 695-702.
- O'Leary, K. D. and Kent, R. Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds), *Behavior change: Methodology, concepts and practice*. Champaign, Ill.: Research Press, 1973. Pages 69-96.
- Rolf, J. E. and Hasazi, J. E. Identification of preschool children at risk and some guidelines for primary intervention. In G. W. Albee and J. M. Joffe (Eds), *Primary prevention of psychopathology (Vol. 1): The issues*. Hanover, N.H.: University Press of New England, 1977. Pages 121-152.
- Rolf, M., Sells, S. B., and Golden, M. M. *Social adjustment and personality development in children*. Minneapolis: University of Minnesota, 1972.
- Siegel, E. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- Strain, P. S., Cooke, T. P., and Apolloni, T. *Teaching exceptional children: Assessing and modifying social behavior*. New York: Academic, 1976.
- Strain, P. S. and Shores, R. E. Social reciprocity: A review of research and educational implications. *Exceptional Children*, 1977, **43**, 526-531.
- Tiktin, S. and Hartup, W. W. Sociometric status and the reinforcing effectiveness of children's peers. *Journal of Experimental Child Psychology*, 1965, **2**, 306-315.
- Walker, D. K. *Socio-emotional measures for preschool and kindergarten*. San Francisco: Jossey-Bass, 1973.
- Walker, H. M., Greenwood, C. R., Hops, H., and Todd, N. M. Differential effects of reinforcing topographic components of social interaction. *Behavior Modification*, 1979, **3**, 291-321.
- Wolf, M. M. Social validity. The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 1978, **11**, 203-214.

Received 18 September 1978.

(Final Acceptance 24 July 1979.)