

Racial Bias in Behavioral Assessment of Children's Social Skills

**Verdi R. Lethermon,¹ Donald A. Williamson,^{1,2}
Sarah C. Moody,³ and Patricia Wozniak¹**

Accepted: May 26, 1986

The social skills of 20 second- and sixth-grade students were assessed by 20 trained raters using the Social Skills Test for Children (SST-C). Rater and child characteristics were examined to determine whether differences in social skills ratings were due to the race of the rater or the race of the children being rated or due to the interactive effects of these characteristics, which would suggest racial bias in the ratings procedure. The results showed that the race of the rater did affect some behavioral observations. Black raters gave higher scores than white raters on four behavioral categories: response latency, appropriate assertion, effective assertion, and smiling. White raters gave higher scores for head position and gestures. The results of this study replicated earlier findings of significant differences in social skills ratings due to the race and age of the child being rated. The results also showed modest racial bias effects in that black and white raters scored black and white children differentially on two behavioral categories: overall skill ratings and smiling. These results suggested that most behavioral categories of the SST-C were not systematically affected by racial bias. However, the most subjective rating, overall skill, did evidence racial bias effects. This finding is consistent with previous data showing that subjective ratings may be most affected by racial bias.

KEY WORDS: Social Skills Test for Children; behavioral assessment; social skills.

¹Department of Psychology, Louisiana State University, Baton Rouge, Louisiana 70803.

²To whom correspondence should be addressed.

³University of New Orleans, New Orleans, Louisiana 70122.

INTRODUCTION

Most of the work concerning behavioral observation has shown that trained raters/observers can reliably observe a variety of behaviors (Ciminero, Calhoun, & Adams, 1977; Hersen & Bellack, 1981; Kazdin, 1981; Lethermon *et al.*, 1984). Trained observers have been used to rate classroom behavior, time-sample behavior checklists among psychiatric patients, and social skills during role play. Using clearly defined behaviors, trained raters have been found to overcome potential sources of scorer unreliability (e.g., observer drift and familiarity of subjects) (Kent, O'Leary, Diament, & Dietz, 1974; Redfield & Paul, 1976) and have generally obtained accurate observation when proper training procedures have been used (House, 1980; Kazdin, 1981). There are, however, a number of problems with most behavioral observation procedures. They are generally unstandardized (Goldfried, 1979; Nelson, 1983) and few procedures have demonstrated satisfactory reliability, validity, or other psychometric characteristics (Cone, 1981). Furthermore, questions of their utility as measures of treatment outcome have been raised (Jacobsen, 1985) and data suggesting racial bias have been reported (Turner, Beidel, Hersen, & Bellack, 1984).

Williamson, Moody, Granberry, Lethermon, and Blouin (1983) reported the development of a standardized role-play test of children's social skills called the Social Skills Test for Children (SST-C). In this study it was demonstrated that the SST-C had a satisfactory criterion-related validity when response profiles of the SST-C were correlated with criterion measures using canonical correlations. In a second report on the SST-C, Granberry, Williamson, Moody, Lethermon, and Michaels (1983) found that there were significant differences in children's role-played social behavior due to age and race of the child. In particular, black children were found to receive lower ratings of social skills on some measures than did white children. In a third study, Lethermon *et al.* (1984) found that *untrained* observers evidenced clear signs of racial bias in their ratings of children's role-played behavior. Therefore, one explanation for the racial differences observed by Granberry *et al.* (1983) may be that the low ratings received by black children may have been due to racial bias of the predominantly white raters. This explanation is especially plausible given the recent report by Turner *et al.* (1984) that white and black raters differentially scored the social behavior of white and black adult schizophrenics in a manner suggestive of racial bias. This study utilized a role play test of social skills for adults (Behavioral Assertiveness Test—Revised) and behaviors were rated using the 9-point rating scales developed by Trower, Bryant, and Argyle (1978).

The present study was designed to test directly the hypothesis of potential racial bias in the scoring of the SST-C by controlling for the race of *trained*

raters during training as well as in their actual observations of children's behavior.

METHOD

Raters

Twenty undergraduate students from Louisiana State University assessed the social behavior of 20 children using the SST-C. The mean age for these college students was 18.86 years. Of the 20 college students, 10 were black and 10 were white. All of the raters were female in order to control for sex of the rater. The raters were trained in five groups of four raters each (counter-balanced for race). This training program involved educational presentations regarding behavioral assessment and practice sessions for rating the SST-C. Feedback on accuracy was provided for each practice session. The racial compositions for each group were as follows:

Group I: 0 blacks; 4 whites

Group II: 1 black; 3 whites

Group III: 2 blacks; 2 whites

Group IV: 3 blacks; 1 white

Group V: 4 blacks; 0 whites

The raters were trained until the interrater reliability for all behavioral categories of the SST-C was above 80% for three consecutive pilot subjects.

Children

The social behavior of 10 second- and 10 sixth-grade students from Louisiana State University Laboratory School and Southern University Laboratory School was assessed by the trained raters. Five of the second graders were white and five were black. Similarly, five of the sixth graders were white and five were black. Of the 20 children assessed, 10 were girls and 10 were boys.

Social Skills Test for Children

The social behavior of the children was assessed using the SST-C. The SST-C consists of 30 role-play scenes for elementary-aged children. The 30 role-play scenes are divided into five categories: giving help, accepting help, giving praise, accepting praise, and assertiveness. There are six samples from each category and each of the 30 scenes is randomly administered.

The behavioral responses of the 20 children had been videotaped earlier using a narrator (age 24 years) to describe the social situations and a prompter

(age 26 years) to provide a verbal prompt to the children. The raters scored these videotapes of the children's social behavior in terms of 13 behavioral categories (i.e., response latency, desired content, appropriate assertion, effective assertion, smiles, gestures, intonation, overall skill, eye contact, speech duration, head position, body posture, and speech fluency). Eight of the behavioral categories (desired content, appropriate assertion, effective assertion, smiles, gestures, head position, body posture, and speech fluency) were scored in terms of occurrence or nonoccurrence. Response latency, eye contact, and speech duration were timed (in seconds). Intonation and overall skill were rated using Likert scales. For each child, these data were summarized by calculating the percentage of occurrence across the 30 role-play scenes for each of the behaviors scored on an occurrence-nonoccurrence basis by averaging the ratings or timed behaviors (across the 30 scenes). Consequently, each behavior was summarized as either a percentage score or a mean (duration or rating) for each behavioral category by each rater.

Procedure

Following training and the establishment of interrater agreement for all behavioral categories greater than or equal to 80%, each rater individually assessed the social behavior of 20 children using the SST-C. These assessments were conducted in quiet, isolated rooms. The children were randomly scored by each rater. Raters were provided rating forms and stopwatches.

RESULTS

Data were analyzed using a 2 (Race of Rater) \times 2 (Race of Child) \times 2 (Age of Child) \times 2 (Sex of Child) multivariate analysis of variance (MANOVA) using the 13 dependent variables of the SST-C. The results indicated significant main effects for race of rater [$F(13,780) = 2.94, p < .004$], race of subject [$F(13,780) = 13.71, p < .0001$], and age of subject [$F(13,780) = 8.97, p < .0001$] and a significant race of rater \times race of subject interaction effect [$F(13,780) = 1.7, p < .05$]. Thereafter, separate 2 (Race of Rater) \times 2 (Race of Subject) \times 2 (Age of Subject) post hoc ANOVAs were performed for each dependent variable across all raters and children.

The primary purpose of the present study was to examine the presence or absence of racial bias (between raters and children being rated) in the social skills ratings using the SST-C. This effect would be shown by a significant interaction between race of rater and race of subject, where the pattern of results showed white raters scoring white children higher than blacks and/or black raters scoring black children higher than whites. Table I summarizes the data pertaining to this interaction effect.

Table I. Summary of ANOVA Results Reflecting the Interaction Effect for Race of Rater × Race of Subject for the Dependent Variables^a

Dependent variable	F	Probability level	Race of rater		Race of children
			Black	White	
Response latency	.28	.60	1.30 sec .98 sec	1.18 sec .90 sec	Black White
Desired content	.44	.51	77% 77%	77% 77%	Black White
Appropriate assertion	.36	.55	14% 17%	14% 14%	Black White
Effective assertion	.63	.43	13% 18%	10% 10%	Black White
Smiles	5.45	.02*	20% ^a 10% ^b	13% ^b 13% ^b	Black White
Gestures	.03	.86	4% 3%	6% 6%	Black White
Intonation	.07	.79	1.03 1.03	1.00 1.01	Black White
Overall skill	12.34	.0005*	1.94 ^a 1.98 ^a	1.88 ^a 2.04 ^b	Black White
Eye contact	1.50	.22	.67 sec .74 sec	.61 sec .75 sec	Black White
Speech duration	.75	.39	1.14 sec 1.11 sec	1.12 sec 1.14 sec	Black White
Head position	.13	.72	98% 98%	99% 100%	Black White
Body posture	.59	.44	98% 99%	98% 100%	Black White
Fluency	.61	.43	97% 99%	97% 100%	Black White

^adf = 1,799. For significant variables, means which differ are shown by different superscripts.
*Means differ significantly ($p < .05$).

Table II. Summary of ANOVA Results Reflecting the Main Effect of Race of Rater for the Dependent Variables^a

Dependent variable	<i>F</i>	Probability level	Mean score for raters	
			Blacks	Whites
Response latency	5.83	.02*	1.14 sec	1.04 sec
Desired content	0.08	.78	77%	77%
Appropriate assertion	42.19	.0001*	16%	14%
Effective assertion	7.08	.008*	16%	12%
Smiles	8.64	.003*	15%	11%
Gestures	4.70	.03*	4%	6%
Intonation	1.05	.30	1.03	1.01
Overall skill	0.03	.85	1.96	1.96
Eye contact	1.05	.30	0.71 sec	.68 sec
Speech duration	0.02	.90	1.13 sec	1.13 sec
Head position	4.03	.04*	98%	99%
Body posture	0.69	.40	99%	99%
Fluency	0.76	.38	98%	99%

^adf = 1,799.*Means differ significantly ($p < .05$).

The data in Table I show significant interaction effects for 2 of the 13 behavioral categories of the SST-C (i.e., "smiles" and "overall skill"). Means for these responses showed that white and black raters scored white and black children differently on these two behavioral categories. Using the Newman Keuls method of individual treatment comparisons to investigate these dif-

Table III. Summary of ANOVA Results Reflecting the Main Effect of Race of Subject for the Dependent Variables^a

Dependent variable	<i>F</i>	Probability level	Mean score for subjects	
			Blacks	Whites
Response latency	52.40	.0001*	1.24 sec	.94 sec
Desired content	0.52	.46	77%	77%
Appropriate assertion	42.19	.0001*	14%	17%
Effective assertion	6.56	.01*	12%	16%
Smiles	36.33	.0001*	17%	10%
Gestures	4.82	.0284*	5%	4%
Intonation	0.05	.82	1.02	1.02
Overall skill	33.02	.0001*	1.91	2.01
Eye contact	14.00	.002*	.64 sec	.75 sec
Speech duration	1.01	.3144	1.13 sec	1.13 sec
Head position	0.60	.4406	98%	99%
Body posture	3.72	.05*	98%	99%
Fluency	7.80	.005*	97%	99%

^adf = 1,799.*Means differ significantly ($p < .05$).

ferences further, it was found that for the dependent variable smiles, black raters scored black children higher than they scored white children and higher than white raters scored either black or white children. For the dependent variable overall skill, white raters scored white children higher than they scored black children.

Table II summarizes the means for each response of the SST-C as a function of the race of the rater. Means marked with asterisk superscripts indicate that the group differed significantly ($p < .05$) for that behavioral category. Data in Table II show that black raters gave all children, regardless of race, significantly higher scores than white raters on response latency, appropriate assertion, effective assertion, and smiles.

Table III summarizes the SST-C response data as a function of the race of the child. These data show that black and white children were scored differentially on 9 of the 13 behavioral categories: response latency, appropriate assertion, affective assertion, overall skill, eye contact, body posture, fluency, and gestures.

The age of the child being rated did not affect the social skills rating in a consistent manner. As can be seen from the results in Table IV, second-graders received higher ratings than did sixth-graders on smiles, gestures, intonation, and speech duration, whereas sixth-graders received higher scores than did second-graders on appropriate assertion, effective assertion, overall skill, eye contact, and body posture. There were no significant sex differences found between the boys and the girls being rated.

Table IV. Summary of ANOVA Results Reflecting the Main Effect of Grade of Subject for the Dependent Variables^a

Dependent variable	<i>F</i>	Probability level	Mean score for subjects	
			Second	Sixth
Response latency	2.31	.1286	1.11 sec	1.07 sec
Desired content	1.35	.2450	77%	77%
Appropriate assertion	19.75	.0001*	15%	17%
Effective assertion	6.22	.0128*	12%	16%
Smiles	6.41	.0015*	14%	12%
Gestures	13.78	.0002*	6%	3%
Intonation	6.21	.0129*	1.04	.99
Overall skill	8.07	.0046*	1.94	1.99
Eye contact	4.26	.0393*	.67 sec	.72 sec
Speech duration	38.46	.0001*	1.20 sec	1.04 sec
Head position	2.73	.0987	98%	99%
Body posture	5.36	.0209*	98%	99%
Fluency	2.68	.1020	98%	99%

^adf = 1,799.

*Means differ significantly ($p < .05$).

DISCUSSION

This paper examined the influence of racial bias (between race of rater and race of child being rated) on assessments of social behavior using the SST-C, as well as the influence of individual characteristics of both raters and children being assessed. From the previous research that has been conducted on the SST-C, one can conclude that both rater characteristics (Lethermon *et al.*, 1984) and child characteristics (Granberry *et al.*, 1983) must be considered when using the SST-C. The results of the present study tend to support these previous findings. The results of this study suggest that black and white raters do evidence subtle, yet reliable, differences in scoring objective and subjective behavioral categories of the SST-C. White raters scored some categories higher than black raters, and vice versa. These data are consistent with the findings of Turner *et al.* (1984) with adults.

More importantly, however, the present findings suggest modest racial bias effects (between race of rater and race of children being rated) in the scoring of the SST-C. Racial bias did not consistently affect the majority of the behavioral categories of the SST-C, in that only 2 of the 13 behavioral categories of the SST-C were affected. For one variable, smiling, black raters were shown to give higher ratings to black subjects. For the overall skill rating, the opposite bias was found, with white raters giving white children higher ratings. These results suggest that, in general, the SST-C is affected only to a modest degree by racial bias when trained raters are used. The most important indication of racial bias in these results is the finding that the overall skill rating was affected by racial bias. This finding is of special significance because many investigators have relied upon this overall skill rating as one of the more clinically significant variables in role-play assessment. This finding is congruent with the findings of Lethermon *et al.* (1984) that racial bias primarily affects very subjective ratings. The overall skill rating is the most subjective of all the behavioral codes of the SST-C. It is of interest to note that in a study of sex bias in behavioral observation, Horn and Haynes (1981) found that only global subjective ratings were influenced by the sex of the rater. Also, Turner *et al.* (1984) used subjective ratings in their investigation, which yielded racial bias effects for two behaviors, speech duration and compliance with requests.

In conclusion, the present study supports previous research which suggests that both rater and child characteristics must be considered when assessing social skills using the SST-C. Additionally, this study points to the absence of consistent and widespread racial *bias* effects when using the SST-C with trained raters. These data in combination with those of Lethermon *et al.* (1984), Horn and Haynes (1981), and Turner *et al.* (1984) suggest that the race and sex of the rater must be controlled, especially when subjective ratings

are used. Furthermore, since the race of the rater was found to systematically affect 9 of the 13 behavioral categories of the SST-C, it is very important that equal numbers of black and white trained raters be used when comparing treatment conditions or making pre-post comparisons. Otherwise effects may be found which are a function of racial characteristics of the raters rather than of the treatment conditions.

REFERENCES

- Ciminero, A. R., Calhoun, K. S., & Adams, H. E. (1977). *Handbook of behavioral assessment*. New York: Wiley & Sons.
- Cone, J. D. (1981). Psychometric considerations. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd Ed.), New York: Pergamon Press.
- Goldfried, M. R. (1979). Behavioral assessment: Where do we go from here? *Behavioral Assessment, 1*, 19-22.
- Granberry, S. W., Williamson, D. A., Moody, S. C., Lethermon, V. R., & Michaels, M. (1983, November). *Role-play assessment of children's social skills: A standardized procedure*. Presented at the annual meeting of the Association for the Advancement of Behavior Therapy, Los Angeles.
- Horn, W. F., & Haynes, S. N. (1981). An investigation of sex bias in behavioral observations and ratings. *Behavioral Assessment, 3*, 173-183.
- House, A. E. (1980). Detecting bias in observational data. *Behavioral Assessment, 2*, 29-31.
- Jacobson, N. S. (1985). The role of observational measures in behavior therapy outcome research. *Behavioral Assessment, 7*, 297-308.
- Kent, R. N., O'Leary, K. D., Diament, C., & Dietz, A. (1974). Expectation biases in observational evaluation of therapy change. *Journal of Consulting and Clinical Psychology, 42*, 774-780.
- Lethermon, V. R., Williamson, D. A., Moody, S. C., Granberry, S. W., Lemanek, K. L., & Bodiford, C. (1984). Factors affecting the social validity of role-play assessment of children's social skills. *Journal of Behavioral Assessment, 6*, 231-245.
- Nelson, R. O. (1983). Behavioral assessment: Past, present, and future. *Behavioral Assessment, 5*, 195-206.
- Redfield, J., & Paul, G. L. (1976). Bias in behavioral observation as a function of observer familiarity with subjects and typicality of behavior. *Journal of Consulting and Clinical Psychology, 44*, 156.
- Trower, P., Bryant, B., & Argyle, M. (1978). *Social skill and mental health*. Pittsburgh, PA: University of Pittsburgh Press.
- Turner, S. M., Beidel, D. C., Hersen, M., & Bellack, A. S. (1984). Effects of race on ratings of social skill. *Journal of Consulting and Clinical Psychology, 2*, 474-475.
- Williamson, D. A., Moody, S. C., Granberry, S. W., Lethermon, V. R., & Blouin, D. C. (1983). Criterion-related validity of a role-play test of children's social skills. *Behavior Therapy, 14*, 466-481.