The Effects of Race and Problem Type on Teachers’ Assessments of Student Behavior

Doris F. Chang
Harvard Medical School

Stanley Sue
University of California, Davis

The authors tested the hypothesis that race bias in teacher evaluations may be problem specific by examining the extent to which ratings of child behavior were influenced by the interaction between the race of the child and the type of presenting behavior. Teachers (N = 197) were presented with three vignettes (overcontrolled, undercontrolled, and “normal”), systematically paired with a photograph of a male child (African American, Asian American, or Caucasian). Respondents rated the seriousness, referability, and typicality of the behavior; the child’s family life; academic ability and performance; and causal dimensions. Although results did not reveal a bias in the ratings of African American students, data suggest that stereotypes remain embedded in teachers’ interpretive framework for Asian Americans, particularly regarding expectations of overcontrolled traits.

As early as the 1960s, psychologists and educators examined the influence of racial stereotypes on ethnic minority students, most notably African Americans. Beginning with Kenneth Clark’s (1961) study of urban classrooms, researchers have generally found that teachers tend to (a) rate African American students less favorably on such measures as personality and behavior, motivation to learn, and classroom performance, (b) hold lower academic expectations for African American students, and (c) treat African American students less favorably than Caucasian students in the classroom (Adams, 1978; Cooper, Baron, & Lowe, 1975; Henderson & Long, 1973; Keller, 1986; Murray, 1996; Partenio & Taylor, 1985; Plewis, 1997). Some have attributed the poorer academic achievement of some African American groups to student characteristics such as race and dialect, which may bias teachers’ ratings and expectations of future performance (DeMeis & Turner, 1978). More recently, controversial evidence of differential treatment of African American students has also been described in instructional practices, the administration of corporal punishment, and referral for special education programs (Contreras & Lee, 1990; McFadden, Marsh, Price, & Hwang, 1992; Serwatka, Deering, & Grant, 1995; Shaw & Braden, 1990).

In contrast to the predominantly negative stereotypes associated with African American students, Asian American students have been lauded as a “model minority,” one that is hardworking, respectful to teachers, and demonstrates superior academic abilities, especially in mathematics and science (Allis, 1991; Kitano & Sue, 1973). Although empirical evidence does support the perception of some Asian American groups as academic overachievers (see Hsia & Peng, 1998, for a review), perceptions that all Asian American groups are overachieving and well-behaved may obscure the difficulties faced by recent immigrants, refugees, and other at-risk Asian American youth (Kim, 1983; Lee, 1994). These so-called “positive” stereotypes have other negative consequences as well. Research has indicated that teachers perceive Asian Americans as unassertive, unexpressive, and lacking in leadership skills (Bannai & Cohen, 1985), as well as more passive, quieter, and less interpersonally effective than Caucasian students (Kim, 1983; Sung, 1987). These stereotypes may explain why teachers have been found to call on Asian American students less often than on Caucasian students (Schneider & Lee, 1990) and to expect less classroom involvement from them.

Beyond their broader social impact on achievement opportunities for Asian and African Americans, the contrasting stereotypes of these two groups may also significantly influence teachers’ appraisals of behavior problems and the likelihood that troubled students receive appropriate services. In general, research representing a wide range of methodologies has revealed that teachers tend to rate African American children higher on undercontrolled behaviors and overall behavior problems than Caucasian children (Epstein, March, Conners, & Jackson, 1998; Kelly, Bullock, & Dykes, 1977; Reid et al., 1998). In contrast, the few studies that have examined behavior ratings of Asian American students have suggested that teachers view them as being less hyperactive and having fewer overall behavioral problems when compared with their Caucasian peers (Chang, Morrissey, & Koplewicz, 1995; Loo & Rappaport, 1998; Spring, Blunden, Greenberg, & Yellin, 1977).
Weisz and Weiss (1991) suggested that interpretation of these differences in behavior ratings must consider the child’s behavior as well as the appraiser’s perception of the behavior. In fact, some epidemiologic studies have shown that the frequency and type of child problems that occur in a given population vary significantly by factors such as ethnicity, culture, and social context (Shen & Wang, 1995; Weisz, Sigman, Weiss, & Mosk, 1993). However, this epidemiologic information must be complemented by research on adults’ definitions of “child psychopathology” and perceptions of what types of behaviors require clinical intervention (Weisz & Weiss, 1991). The present study attempted to examine this assessment process among teachers because of the significant influence they exert on parental decisions to seek help for their disturbed children (Rich, 1977; Rosenthal & Jacobson, 1968; Weisz et al., 1988). In addition, schools themselves are a primary site for identification and intervention of child behavioral problems (Amatea, 1989; Edens, 1998).

Applying an experimental model for assessing bias, we examined whether teacher ratings of undercontrolled, overcontrolled, and “normal” behaviors varied according to the race of the child. More specifically, we addressed the question of whether variations in teachers’ perceptions of what constitutes normal, or typical, behavior for African American, Asian American, and Caucasian students may result in different appraisals of the same behavior and produce different solutions for handling that behavior. Those perceptions may be based on racial stereotypes or direct experience with different ethnic groups, which may then provide a specific standard for evaluating the meaning and significance of an individual group member’s behavior.

Sonuga-Barke, Minocha, Taylor, and Sandberg (1993) found that teachers do tend to apply different standards for interpreting behavior in racially diverse groups of children. There is also evidence that teachers hold different definitions of positive adjustment according to race (Kim, 1983). Race differences in ratings of hyperactive behaviors in schoolchildren led Spring et al. (1977) to conclude that separate norms should be computed for African American and Asian American children. These “shifting standards” for evaluating racially diverse students may contribute to differences in subjective ratings of behavior, even when objective ratings are comparable (Biernat & Manis, 1994; Kahneinan & Miller, 1986).

One limitation of the current research is that with few exceptions the majority of studies have focused on the assessment of undercontrolled child behavior problems in the classroom. Because behavior problems are typically conceptualized within educational settings as “acting-out” behaviors that interfere with classroom instruction, teachers may not recognize certain overcontrolled behaviors as a sign of poor adjustment or psychological distress (Kim, 1983). The finding that overcontrolled problems are perceived by teachers to be less serious than undercontrolled problems may also contribute to the conspicuously low rate of referrals of Asian American children despite increased risk for psychological distress in some groups (Walker, Bettes, & Ceci, 1984; Weisz et al., 1988). Including both under- and overcontrolled behaviors in studies of teacher ratings of child behavior may reveal ways in which race and problem type can interact to influence the appraisal process.

We hypothesized that teachers’ ratings of Asian American and African American children would reflect popular racial stereotypes of these groups. Drawing on previous research findings, we further hypothesized that race-based expectancies, combined with a tendency to minimize overcontrolled behaviors in general, would lead respondents to (a) view undercontrolled behaviors in an African American child as more typical and more problematic than those same behaviors in Caucasian or Asian American children and (b) view overcontrolled behaviors as more typical of Asian American children but less of a concern compared with those same behaviors in the other two groups. Ratings of severity and need for referral were used to assess teachers’ perceptions of the behavior and level of concern, whereas perceptions of typicality of behavior, quality of family life, and academic performance were used to evaluate other aspects of a child’s presentation. Finally, teachers’ causal attributions of the child’s behavior were evaluated to explore differences along the two dimensions of race and type of problem.

Method

Sample

Data for this study were collected from 197 teachers (163 women, 34 men) representing over 160 schools in Southern California. All subjects were recruited from summer continuing education courses offered through the University of California. Only 4 teachers who were approached declined to participate (response rate 98.0%). Subjects were informed that their responses would be kept anonymous and that their participation was voluntary. Compensation was provided in the form of lunch subsidies (value $3) and extra credit to be applied to their final exams. The subjects’ mean age was 33.2 years (SD = 8.6) with ages ranging from 21 to 60 years old. The ethnic background of the teachers was as follows: 74.1% Caucasian, 10.6% Hispanic or Latino, 3.6% African American, 2.5% Asian/Pacific Islander, 2.0% other, and 7.1% mixed race/multietnic. The majority (86.3%) had received bachelor’s degrees, although 0.5% had only an associate’s degree and 13.2% had received a graduate degree. The average amount of teaching experience was 4.4 years (SD = 5.1). Nearly three quarters of the sample had experience teaching elementary school children. They referred an average of four students a year for problems related to behavior (59.9%), learning difficulties (58.4%), and other reasons (16.8%). The teachers reported having made referrals to learning specialists (33.5%), school principals (28.9%), psychologists (25.4%), medical doctors (5.6%), and school review committees (11.2%).

Materials

A survey was developed to assess the behavior of three hypothetical male children (Caucasian, African American, and Asian American) in the fourth grade. The basic survey included a stock photograph of a male child, a vignette describing the child’s behavior and basic demographic information, and questions about the child’s behavior including seriousness, need for referral, typicality of the behavior, perceptions of family life, academic performance and ability, and causal attributions.

Vignette construction. Three brief vignettes were written to represent overcontrolled, undercontrolled, and normal school behavior. The behaviors and traits used to characterize the problem syndromes were derived from the behavior problem scales of the Child Behavior Checklist—Teacher’s Report Form (TRF; Achenbach, 1991) for boys aged 6 to 11. The TRF is a checklist designed to obtain teachers’ reports of behavior problems, school performance, and adaptive functioning. The Anxious and Aggressive Problem scales were selected to represent the broad-band overcontrolled/internalizing and undercontrolled/externalizing syndromes respectively, because principal-components analysis indicated they had the highest loadings on these second-order Varimax factors (Edelbrock & Achenbach, 1984).
Characteristics described in the anxious/overcontrolled vignette included being anxious to please and afraid of making mistakes, feeling the need to be very neat and tidy, being too neat in clinging to adults, and being shy and timid. Characteristics described in the aggressive/undercontrolled vignette included being disobedient and easily frustrated, disrupting the class, talking out of turn, demanding attention, sulking and fidgeting. The normal vignette included characteristics such as fidgeting only occasionally, generally following the rules, demonstrating normal play, having some friends, showing some interest in school, excelling in one subject, and exhibiting some distractibility in the classroom.

Pilot studies of the vignettes were conducted to evaluate the effectiveness of the manipulation when the race of the child was unspecified. A convenience sample of undergraduates in psychology was presented with the three vignettes and asked to evaluate the seriousness of the child’s problem using a 9-point Likert-type scale. Results were used to inform subsequent revisions, and pilot tests were conducted until results showed that (a) the two problem conditions were judged to be equally serious and (b) the normal vignette was judged to be significantly less serious than the two problem conditions.

Photograph selection. Three individual photographs of a Caucasian, African American, and Asian American fourth-grade boy were selected to accompany the vignettes. The use of a single stimulus to represent a racial category has been previously observed in Stevens (1981) and others in cases where gender and level of attractiveness were held constant. Because research has shown that attractive children are evaluated more favorably by teachers (Adams, 1978; Ritts, Patterson, & Tubbs, 1992), a pilot study was conducted on 30 randomly selected photographs (10 of each race). A convenience sample of 35 ethnically diverse undergraduates were presented with 10 photographs of either Caucasian, African American, or Asian American boys and asked to assign a rating of physical attractiveness on a scale from 1 (very unattractive) to 5 (very attractive). In addition, the subjects were asked to identify the most and the least physically attractive child within each racial grouping. Photographs consistently ranked as either the most or least physically attractive within each racial group were eliminated. Mean rankings were calculated and of those rated “average” on physical attractiveness, 3 photographs (1 from each racial group) were selected for use in the study.

Assessment of behavior. Accompanying each set of stimuli were a series of questions that asked respondents to evaluate the child’s behavior along six dimensions: (a) severity of the behavior problem (1 item, “not severe at all” to “very severe”), (b) the likelihood that the respondent would refer the child for different services or interventions (8 items, “would not refer” to “very likely to refer”), (c) perceived quality of the child’s family life (1 item, “very poor quality” to “very good quality”), (d) perceptions regarding academic performance and ability (2 items, “very poor” to “very good”), (e) typicality of the child’s behavior (2 items, “not very typical” to “very typical”), (f) perceived causes of behavior and causal attributions (12 items as described below). Ratings were based on a Likert-type 9-point scale, with the exception of a single open-ended question, “What do you think are the primary causes of the child’s behavior?” Using an inductive approach, we content analyzed responses to this question to derive 13 causal categories. When more than one cause was suggested, multiple codes were assigned. All coding was conducted by Doris F. Chang.

Causal attributions of the behavior problems presented in the vignettes were further assessed using the Revised Causal Dimension Scale (CDS-II; McAuley, Duncan, & Ritts, 1992). The 12-item CDS-II is based on Weiner’s Attributional Theory (Weiner, 1985, 1990), which suggests that biases in judgment may be a result of different causal beliefs associated with characteristics of the individual being judged. Specifically, the CDS-II evaluates four causal dimensions: Locus of Causality, Stability, Personal Control, and External Control. Locus of Causality refers to whether the cause resides within or is external to the actor, whereas the Stability dimension assesses whether the cause is invariant or changeable over time.

The control dimensions comprise degree of personal control and degree of external control related to the cause of the behavior. Using data from four studies, McAuley et al. (1992) found average internal consistency coefficients that were within the acceptable range for all four subscales: Locus of Causality, 0.67; Stability, 0.67; Personal Control, 0.79; External Control, 0.82. A confirmatory factor analysis indicated that the hypothesized four-factor model provided an excellent fit to the data, $\chi^2(48, N = 380) = 96.85$, $p < .001$, goodness of fit index = .96 (McAuley et al., 1992).

**Design and Analysis Plan**

Because judgment tasks assume relative comparisons, prior studies of race bias in behavioral evaluations have typically used a single-factor within-subjects design in which different levels of race are examined. However, the present study hypothesized an interaction previously unexamined, namely that race bias may manifest itself differently for different types of problems. Traditionally, this interaction would have been evaluated in one of two ways. A fully crossed 3 (race) $\times$ 3 (problem type) within-subjects design would result in nine treatment conditions for each subject. However, the repetition associated with such a design would likely expose the research question and result in data of questionable validity. A second possibility is a between-subjects design in which each subject is presented with one of the nine possible treatment conditions. Such a design would guard against subject fatigue and obscure the research question, but it could only aggregate judgment tendencies across the sample as a whole.

A consideration of these factors resulted in the use of an experimental design that combines the advantages of the two approaches while minimizing their disadvantages. As described by Winer, Brown, and Michels (1991), a $3 \times 3$ design with a confounded interaction allows analysis of the Race $\times$ Problem Type interaction, while limiting the number of treatment conditions per subject. Each subject received three treatments that encompassed all levels of both factors but not the nine combinations that would be necessary to fully estimate the interaction. The complete design was run in sets of six blocks composed of two subjects each. Over the six blocks, the design formed two replications (three blocks each) of all combinations of the two factors. Data analysis involved configuring the data by race, problem type, block, and replication to yield all the terms necessary to conduct a series of two-way analyses of variance on a total of 591 cases. Analyses based on fewer than 591 cases were the result of missing data for those dependent variables.

To minimize the risk of chance findings, we applied the Bonferroni adjustment for multiple comparisons as follows. For the three key dependent variables, severity of behavior, likelihood of referral to parent, and likelihood of referral to a child specialist, the threshold for significance of the initial tests of main effects and interaction was set at $p < .015$ ($p < .05$ divided by 3, roughly). Similarly, for family quality, academics, and the two variables referring to typicality of behavior, the significance level was also set at $p < .01$ for these initial tests. Finally, the four causal attribution variables were assessed at the level of $p < .01$ to guard against overall error inflation.

Where significant main effects or interactions were identified, the Bonferroni principle was also used in post hoc analyses, using an error rate of $p < .003$ ($p < .01$ divided by 3). Significant Race $\times$ Problem Type interactions were examined by first testing the main effect of each independent variable within each level of the other. Significant simple main effect tests at $p < .0016$ ($p < .01$ divided by six tests) were followed by pairwise comparisons applying a significance level of $p < .0005$ ($p < .0016$ divided by 3, roughly).

**Results**

**Teacher Characteristic**

Univariate analyses were conducted to examine the role of subjects’ gender, referral practices, and teaching experience on...
their behavior ratings. (Unequal sample sizes did not permit an analysis of ratings by teacher ethnicity.) Results indicated that women were more likely than men to recommend a consultation with the child’s parent, \( F(1, 590) = 6.94, p < .01 \). Women were also more likely to view the causes of problems as reflecting an aspect of the child versus an aspect of the situation (Locus of Causality), \( F(1, 585) = 12.20, p < .001 \). Ratings on other dependent variables did not differ by gender.

The average number of referrals made each year was transformed into a categorical variable on the basis of a median split (0–3 referrals versus 4+ referrals per year). No significant effects of referral were found for any of the dependent variables. A similar methodology was applied to examine the effects of teaching experience (<1–3 years versus 4+ years) on behavior ratings; however, no significant effects were observed.

**Relationships Between the Dependent Variables**

Correlations between the dependent variables ranged from .00 to .79 and are presented in Table 1.

**Seriousness of Behavior and Likelihood of Referral**

A significant main effect of problem type was found for severity of the behavior problem, \( F(2, 590) = 291.27, p < .0001 \), likelihood of parental contact, \( F(2, 590) = 182.72, p < .0001 \), and likelihood of referral to a child specialist, \( F(2, 586) = 177.00, p < .0001 \). The undercontrolled condition was rated significantly more serious and more likely to result in consultation and referral than the overcontrolled or normal conditions (all \( p < .001 \)). There were no main effects of race, nor was the Race × Problem Type interaction significant for any of these variables.

**Typicality of Behavior**

A significant main effect of problem type, \( F(2, 368) = 46.19, p < .0001 \), and a Race × Problem Type interaction, \( F(4, 368) = 7.37, p < .0001 \), was found for ratings of the typicality of the presenting behavior compared with other boys of the same race. As expected, post hoc comparisons revealed significantly higher typicality scores for the normal condition compared with the undercontrolled or overcontrolled problem conditions (\( p < .0001, M = 6.10 \) vs. 4.08 and 4.34, respectively). There were no significant differences for typicality between the two problem conditions.

Examination of the Race × Problem Type interaction (see Figure 1) revealed striking differences with regard to expectations of typicality for the Asian American boy compared with his African American and Caucasian counterparts. Simple main effect tests revealed that race was significant at each level of problem type (all \( p < .0001 \)). Pairwise comparisons within levels of problem condition indicated that (a) the normal Asian American was rated significantly less typical of his race compared with the normal Caucasian (\( p < .0005 \)); (b) the undercontrolled Asian American was judged significantly less typical than the undercontrolled Caucasian and African American (both \( p < .0005 \)); and (c) the overcontrolled Asian American was viewed as significantly more typical than the overcontrolled Caucasian and African American cases (both \( p < .0001 \)). With regard to problem type, simple main effect tests were also significant for each level of race (all \( p < .0001 \)). Pairwise comparisons within levels of race indicated that (a) the Asian American boy in the undercontrolled condition was judged to be significantly less typical for his race than those in the normal or overcontrolled conditions (both \( p < .0001 \)); (b) the African American boy in the two problem conditions was seen as significantly less typical than the one in the normal condition (both \( p < .0005 \)); and (c) the Caucasian boy in the two problem conditions was viewed as significantly less typical than the one in the normal condition (both \( p < .0001 \)), and the overcontrolled Caucasian was rated significantly less typical than the undercontrolled Caucasian (\( p < .0005 \)).

Respondents were also asked to consider how typical the stimulus child was compared with male children in general, regardless of race. Results indicated only a significant main effect of problem type, \( F(2, 590) = 297.86, p < .0001 \). As expected, behaviors described in the normal condition (\( M = 7.05 \)) were considered significantly more typical of general male behavior than those described in the undercontrolled or overcontrolled conditions (both \( p < .0001, M = 4.65 \) and 3.67, respectively). The behaviors described in the undercontrolled and normal condition were rated significantly more typical of boys in general than those described in the overcontrolled condition (both \( p < .0001 \)).

**Table 1**

*Correlations Among Dependent Variables and Significance Level*

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Serious</td>
<td></td>
<td>.705**</td>
<td>.785**</td>
<td>-.460**</td>
<td>-.580**</td>
<td>-.467**</td>
<td>-.448**</td>
<td>-.342**</td>
<td>.026</td>
<td>.082*</td>
<td>-.319**</td>
<td>.157**</td>
</tr>
<tr>
<td>2. Parent referral</td>
<td></td>
<td></td>
<td>.643**</td>
<td>-.406**</td>
<td>-.484**</td>
<td>-.451**</td>
<td>-.390**</td>
<td>-.282**</td>
<td>.040</td>
<td>.102**</td>
<td>-.217**</td>
<td>.047</td>
</tr>
<tr>
<td>3. Specialist referral</td>
<td></td>
<td></td>
<td></td>
<td>-.438**</td>
<td>-.492**</td>
<td>-.414**</td>
<td>-.459**</td>
<td>-.319**</td>
<td>.092*</td>
<td>.016</td>
<td>-.306**</td>
<td>.125**</td>
</tr>
<tr>
<td>4. Typicality of race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.519**</td>
<td>.401**</td>
<td>.341**</td>
<td>.202**</td>
<td>.013</td>
<td>-.167**</td>
<td>.209**</td>
<td>-.042</td>
</tr>
<tr>
<td>5. Typicality of boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.431**</td>
<td>.228**</td>
<td>.187**</td>
<td>.093*</td>
<td>-.089*</td>
<td>.337**</td>
<td>-.167**</td>
</tr>
<tr>
<td>6. Family quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.484**</td>
<td>.424**</td>
<td>.023</td>
<td>-.147**</td>
<td>.209**</td>
<td>-.034</td>
</tr>
<tr>
<td>7. Academic performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.683**</td>
<td>-.094</td>
<td>.095</td>
<td>.158**</td>
<td>.029</td>
</tr>
<tr>
<td>8. Academic ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.069</td>
<td>.082</td>
<td>.120**</td>
<td>-.037</td>
</tr>
<tr>
<td>9. Locus of causality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.289**</td>
<td>.247**</td>
<td>.054</td>
</tr>
<tr>
<td>10. External control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.000</td>
<td>-.056</td>
</tr>
<tr>
<td>11. Personal control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.185**</td>
</tr>
</tbody>
</table>

* \( p < .05 \). ** \( p < .01 \).
Family and Academic Life

For quality of family life, there was also only a significant main effect of problem type, $F(2, 363) = 38.23, p < .0001$. Normal boys had higher scores on quality of family life than those in the undercontrolled or overcontrolled conditions ($ps < .0001, M = 6.76$ vs. 5.04 and 5.42, respectively).

A significant main effect of problem type was also found for predictions of academic performance, $F(2, 386) = 200.18, p < .0001$, and perceptions of academic ability, $F(2, 570) = 151.23, p < .0001$. Boys exhibiting normal or overcontrolled behavior were expected to perform significantly better than those characterized as undercontrolled ($ps < .0001, M = 6.35$ and 6.43 vs. 3.15, respectively) and to possess greater innate academic ability than those in the undercontrolled condition ($ps < .0001, M = 7.06$ and 6.98 vs. 5.05, respectively). There were no main effects of race, nor was the Race $\times$ Problem Type interaction significant for any of these variables.

Causal Attributions

A significant main effect of problem type was found for all four causal attribution variables, namely Locus of Causality, $F(2, 585) = 4.68, p < .01$, Stability, $F(2, 584) = 7.71, p < .001$, External Control, $F(2, 585) = 13.08, p < .0001$, and Personal Control, $F(2, 586) = 47.46, p < .0001$. Respondents rated the causes of overcontrolled behaviors as more stable than those attributed to the normal condition ($p < .001, M = 4.89$ vs. 4.37) and more controllable by others than causes attributed to normal behaviors ($p < .0001, M = 6.78$ and 6.09, respectively). The causes of normal behaviors were judged to be significantly more controllable by the child than those of the under- or overcontrolled conditions ($p < .0001, M = 6.47$ vs. 5.29 and 5.43, respectively). Significant effects were not found for race or for the Race $\times$ Problem Type interaction for any of the causal attribution variables.

Discussion

The purpose of the study was to examine how race and problem type influence teachers’ judgments of what constitutes normal and abnormal child behavior in an educational setting. This investigation is important because it is the first to include three racial groups (African American, Asian American, and Caucasian) that vary considerably in public stereotypes and three problem conditions (undercontrolled, overcontrolled, and normal behavior). Because teachers’ judgments and perceptions influence the attention and aid rendered to students, this study specifically examined the possible effects of these child characteristics on perceptions of severity, referability, and typicality of behavior; perceptions of cause; and positive characteristics such as academic potential. With respect to those evaluations, several interesting results emerged, providing partial support for our hypotheses. First, teachers did not rate the behavior of the undercontrolled African American boy as significantly more typical, more serious, or in greater need of referral compared with the other ethnic groups in the same problem condition. Second, although teachers did rate the over-
controlled Asian American boy as significantly more typical, they
did not view those behaviors as significantly more serious or in
need of referral than the African American or Caucasian in the
same condition. Third, although in many respects evaluations for
the African American case were highly similar to those for the
Caucasian case, those for the Asian American revealed differences
consistent with stereotypes of that group. Teachers perceived that
for Asian Americans, overcontrolled behaviors were significantly
more typical of the race than undercontrolled behaviors, whereas
the reverse was true in the Caucasian case. The undercontrolled
Asian American student was judged to be significantly less typical
of his race compared with the undercontrolled Caucasian or Afri-
can American student, whereas the exact reverse was true for the
overcontrolled condition: The overcontrolled Asian American case
was seen as significantly more typical of the race compared with
the other two groups. Finally, it is important to note that racial
differences as well as the Race × Problem Type interactions failed
to be significant in the majority of evaluations.

Why did evaluations of African Americans yield so few differ-
ences whereas those for Asian Americans largely support the
public stereotypes for this minority group? One possibility is that
stereotyping members of a racial group, especially with negative
characteristics, is socially undesirable. Characterizing African
Americans as being disobedient and disruptive may be more
undesirable than characterizing Asian Americans as being anxi-
ous to please and timid. Indeed, overcontrolled behavior was associ-
ated with significantly higher academic performance and innate
ability than undercontrolled behavior. Another possibility is that
teachers in a large, urban environment such as Los Angeles are becom-
ing more objective and less biased in their evaluations of an
increasingly multicultural student body. According to the 2000
Census, Los Angeles County had 2,667,976 children under the age
of 18, of whom 57.5% were Hispanic or Latino, 19.8% non-
Hispanic White, 9.9% African American, 9.3% Asian/Pacific Is-
lander, 0.2% American Indian/Alaska Native, 0.3% other races,
and 2.9% multiracial (U.S. Bureau of the Census, 2000).

In particular, the similarity in ratings for the African American
and Caucasian boys across variables in this study suggest that
respondents were less likely to rely on stereotypes of these groups.
However, teachers may have less experience and familiarity with
Asian Americans, whose numbers have increased substantially
only in the past 2 decades. Consequently, they may rely more often
on stereotypes. Finally, although the experimental design used in
this study enabled us to examine the Race × Problem Type
interaction while limiting the number of treatment conditions per
subject, smaller effects of race may have been obscured as the
result of combining within- and between-subject race effects.
Although this is a possibility, evaluations for Asian Americans
were largely consistent with racial stereotypes of this group.

In general, racial stereotyping was not pronounced and was
confined primarily to Asian Americans. It can be argued that
judgments about Asian Americans reflect knowledge of group
averages rather than stereotypes. Indeed, much has been written
about the influence of Confucianism and Buddhism on many
Asian cultures and, in particular, their emphasis on respect for
authority, self-control, and interconnectedness (e.g., Ho, 1981;
Bond & Hwang, 1986). To qualify as a stereotype, a characteriza-
tion must be overgeneralized or inappropriately applied to mem-
bers of a group, thereby overshadowing individual differences that
may exist. In the worst case, the image of Asian Americans as
typically overcontrolled may make teachers less sensitive to the
real problems that can occur among Asian American children or to
the effects of overcontrolled behaviors (extreme shyness, other-
directedness, etc.) on classroom adjustment.

With respect to over- and undercontrolled behavior problems,
the findings of this study converge with prior research suggesting
that children with overcontrolled problems, regardless of race,
are often more visible in the classroom setting (Rich, 1977; Walker
et al., 1984). The professional task of teachers and the need to
maintain a proper learning environment appears to result in a
greater recognition of and concern for undercontrolled problems,
due to their disruptive and distracting nature. As seen in this study,
overcontrolled problems such as anxiety, depression, and social
withdrawal often fail to attract the same level of concern, are
perceived as less serious, and are less likely to be referred for more
systematic assessment and treatment. In contrast, research indi-
cates that psychologists see both under- and overcontrolled prob-
lems as equally serious in children (Weisz et al., 1988; Solso &
Koechel, 1969). Although differences in perceived referability of
these two classes of behavior problems may be understood within
the goals and demands of the academic setting, the result is that
untold numbers of schoolchildren with overcontrolled behavior
disorders likely go undetected and untreated each year.

Limitations of this study include the use of undergraduates to
establish validity for the vignettes, the use of only three vignettes
and three photographs to represent the constructs of race and
problem type, restriction of this study to a single gender, and the
large number of related variables assessed. These factors raise the
possibility that subjects may have been responding to specific
details in the vignettes or in the photographs that may not be
generalizable to other overcontrolled or undercontrolled behavior
presentations or to other racial phenotypes. Nevertheless, the find-
ings of this study provide compelling preliminary evidence that
race and type of problem do interact to affect judgments of normal
and abnormal behavior.

Future studies should seek to demonstrate the ecological validity
of these findings by examining the Race × Problem Type inter-
action in a natural setting with a more ethnically diverse sample of
teachers. Such a study would also allow a test of the effects of race
and ethnic match on behavioral evaluations. Additional research is
needed to explore the multiple interactions between race, gender,
and behavior and how these factors collectively influence teachers’
appraisals of student behavior. Finally, more research on teachers’
evaluations of understudied ethnic minority groups such as Asian
Americans, Latinos, and Native Americans is also needed to
clarify the social, cultural, and ecological aspects of behavior
disorders as they relate to teacher expectancies, attributions, and
intervention efforts.

References
1991 profile. Burlington, VT: University of Vermont Department of
Psychiatry.
Adams, G. (1978). Racial membership and physical attractiveness effects


Bond, M. H., & Hwang, K. K. (1986). The social psychology of Chinese people. In M. H. Bond (Ed.), *The psychology of the Chinese people* (pp. 213–266). Hong Kong, Taiwan: Oxford University Press.


Weiner, B. (1990). Searching for the roots of applied attribution theory. In...
S. Graham & V.S. Folkes (Eds.), Attribution theory: Applications to achievement, mental health, and interpersonal conflict (pp. 1–13). Hillsdale, NJ: Erlbaum.


Received February 23, 2001
Revision received March 14, 2002
Accepted March 27, 2002