# UNDERSTANDING DISPROPORTIONATE REPRESENTATION IN SPECIAL EDUCATION BY EXAMINING GROUP DIFFERENCES IN BEHAVIOR RATINGS

CHRISTINA D. PETERS, JOHN H. KRANZLER, JAMES ALGINA, STEPHEN W. SMITH, AND ANN P. DAUNIC

*University of Florida*

The aim of the current study was to examine mean-group differences on behavior rating scales and variables that may predict such differences. Sixty-five teachers completed the Clinical Assessment of Behavior–Teacher Form (CAB-T) for a sample of 982 students. Four outcome variables from the CAB-T were assessed. Hierarchical linear modeling was used to analyze variance components across three levels; examine mean-group differences across outcome variables for student gender, race/ethnicity, and free or reduced-price lunch status; and examine whether teacher variables predicted teacher-specific differences in ratings. Results revealed that a significant amount of variance was attributable to teacher- and school-level variables. Several mean-group differences emerged, and some teacher-specific differences in ratings across groups were predicted by teacher self-efficacy for behavior management and teacher age, but not for teacher race/ethnicity, gender, or years of experience. © 2014 Wiley Periodicals, Inc.

Many researchers have examined the topic of potential teacher bias against racial/ethnic minority students in the public school system. Initial research on teacher bias (e.g., Gottlieb, 1964) focused on how teacher expectations differed as a function of student race/ethnicity (e.g., Clifton, Perry, Parsonson, & Hryniuk, 1986; Dusek & Joseph, 1983). Recently, the topic of potential teacher bias has been an interest as it relates to the disproportionate representation of racial/ethnic minority students in some special education programs (e.g., Cullinan & Kauffman, 2005). Disproportionate representation is defined as "the extent to which membership in a given ethnic group affects the probability of being placed in a specific special education disability category" (Oswald, Coutinho, Best, & Singh, 1999, p. 198). An example for students identified with an emotional/behavioral disorder (EBD) is as follows: in 2002–2003, African American students represented 27% of students with EBD, but made up only 17% of public school students (Cullinan & Kauffman, 2005), whereas Hispanic children and youth made up 8% of students with EBD but 16% of public school students.

Ever since racial/ethnic disproportionality was first identified, education professionals have debated the meaning of these differences. The hypothesis that teacher bias is importantly related to disproportionality in special education has attracted by far the most attention over the years (e.g., Kauffman, 2001). Students who are referred by their teachers have a high probability of being identified for special education services (Algozzine, Christenson, & Yesseldyke, 1982). In many school systems, identification begins with teacher referral, and students who exhibit externalizing behaviors that disrupt the classroom are highly likely to be referred. Kauffman (2001) argued that, when teachers observe the behavior of students who are from a different cultural background, their own cultural background may bias their perceptions because standards and expectations regarding behavior are bound by culture. Given that the majority of teachers are Caucasian, cultural bias may disproportionately influence the referral rates of racial/ethnic minority students (Donovan & Cross, 2002).

If a student demonstrates inappropriate behavior and does not display adequate response to behavioral interventions, he or she may be referred for a psychological evaluation to consider eligibility for a special education program. Frequently, teachers are asked to complete behavior

---

Correspondence to: Christine D. Peters, 301 North Kingsway Road, Brandon, FL 33510. E-mail: DrChristinePeters@gmail.com

rating scales as part of the evaluation. Within recent years, many researchers have examined the completion of behavior rating scales in an attempt to assess the possibility of teacher bias (Cohen, DuRant, & Cook, 1988). Several studies indicate that African American students tend to be rated as exhibiting more problematic behavior than Caucasian students do. On the Conner's Teacher Rating Scale (Conners, 1973), for example, African American students were rated significantly worse than were Caucasian students on the conduct disorders subscale (Cohen et al., 1988). Similar results were found in studies by Epstein, March, Conners, and Jackson (1998) and Sbarra and Pianta (2001).

Some researchers have examined the interaction of teacher and student race/ethnicity on teachers' ratings of student behavior (Cullinan & Kauffman, 2005; Rayfield, 1997; Rong, 1996). Results of this research have been mixed, with some studies indicating the presence of teacher bias toward racial/ethnic minority students and other studies showing no difference in ratings between Caucasian and racial/ethnic minority teachers. For example, Rong (1996) examined the interaction of student and teacher gender and race/ethnicity on social behavior ratings on the Behavioral Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). After controlling for socioeconomic status (SES), Rong found that African American teachers produced similar ratings for African American and Caucasian students on all of the BASC subscales. Caucasian teachers, however, rated African American students significantly lower than Caucasian students on several of the prosocial subscales. Downey and Pribesh (2004) also found that African American students received less-favorable ratings than their Caucasian peers did. However, when teachers' race/ethnicity was taken into account statistically, they found that African American students were rated more favorably by African American teachers than were Caucasian students by Caucasian teachers.

Other studies have found that teacher race/ethnicity is not related to differences in behavior ratings across racial/ethnic groups of students. Pigott and Cowen (2000) found that African American students received less-favorable ratings than Caucasian students did, although teacher race/ethnicity was not related to these ratings and did not predict differences in behavior ratings across different racial/ethnic groups of students. Cullinan and Kauffman (2005) also found that teacher-rated differences between Caucasian and African American students were not predicted by an interaction of student and teacher race/ethnicity. Finally, Abidin and Robinson (2002) found no statistically significant differences among students' gender, race/ethnicity, SES, and age on either of the behavioral rating scales utilized in their study.

One possible explanation for the contradictory results of research on teacher bias may be related to the fact that in these initial studies, teachers did not rate the behavior of all students in their classrooms, but focused primarily on the behavior of children and youth with behavioral disorders. When teachers rate all students in their classes, it is possible to examine whether variance in mean-group differences (e.g., African American students being rated as behaving worse than Caucasian students do) is attributable to the teacher or the classroom. It is also possible to include predictor variables at the teacher and school level.

Mashburn, Hamre, Downer, and Pianta (2006) examined within-class and between-class sources of variance in teachers' ratings of students' social competence. In addition, they explored the effects of professional background, psychological factors, and characteristics of pre-kindergarten classrooms on teacher ratings. Using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), Mashburn et al. (2006) found that 15% to 33% of the variance in teacher ratings was related to differences between teachers or classrooms. In addition, they found that teachers' race/ethnicity was significantly related to three of the four student-level variables examined in the study. Specifically, Caucasian teachers rated their students as having more behavior problems and less competence than did African American teachers, on average (Mashburn et al., 2006). These results exemplify how the use of HLM techniques can provide more information about predictors of mean-group differences in ratings than can single-level regression equation models.

Further research on referral bias in which teachers rate all students in their classrooms is needed to further understand the nature and cause of differences in teacher perceptions of student behavior across racial/ethnic groups. Although some research suggests that teacher race/ethnicity is related to mean-group differences in behavior ratings (e.g., Downy & Pribesh, 2004; Rong, 1996), results of other research suggests the opposite (e.g., Cullinan & Kauffman, 2005). Whereas the previous research has added important information to the field, a limitation is that teacher variables have been limited to gender and race/ethnicity. One important variable that has yet to be considered in this context is teacher self-efficacy. Theories of teacher self-efficacy, which is defined as the belief that teachers can influence a student's learning and behavior despite the presence of risk factors, have been based on Bandura's (1977) theory of self-efficacy. According to Ashton and Webb (1986), teacher self-efficacy is importantly related to teachers' self-perceived role, attitudes toward their job, and interactions with students. Teacher self-efficacy may play an important role in how teachers perceive their students' behavior. Additional teacher variables that may play an important role are teacher age and years of experience, as they have been shown to be related to teachers' perception of self-efficacy (e.g., Chester & Beaudin, 1996).

The purpose of the current study was to build on the current literature regarding behavioral differences among students when rated by their teachers. We were interested in whether mean-group differences emerge when teachers complete a behavior rating scale for every student in their class and also whether any teacher variables predicted mean-group differences. HLM techniques were used to partition variance associated with teacher ratings of student behavior (internalizing, externalizing, competence, and social skills) at the student, teacher, and school levels among students of multiple racial/ethnic groups. At the student level, we included student gender, race/ethnicity, and SES, as these variables have been identified as some of the conventional predictors of disproportionality (Shifrer, Muller, & Callahan, 2011).

In addition, we examined the effects of teacher gender, race/ethnicity, age, years of experience, and self-efficacy as predictor variables for teacher-specific differences in ratings. Finally, we included school size, school percentage of minority students, and school percentage of students eligible for free or reduced-price lunch (FRL) to control for those variables at the school level, as differences in ratings have been found to be related to those variables in other studies (e.g., Caldas & Bankston, 1998; Koth, Bradshaw, & Leaf, 2008; Ryabov & Van Hook, 2007).

## METHODS

Data for this study were compiled from a large dataset collected by the Prevention Research Team (PRT). The PRT examined the effectiveness of a cognitive–behavioral intervention, Tools for Getting Along (see Daunic, Smith, Brank, & Penfield, 2006; Daunic et al., 2012), and collected pre- and post-test data on students and teachers from 2006 to 2009 in North Central Florida. Pre-test data were used in the current study to avoid any effects that Tools for Getting Along may have had on the students' behavior or teachers' ratings.

The PRT recruited schools from a pool of approximately 70 elementary schools in North Central Florida (suburban and rural districts) that contained a high percentage (60% to 95%) of students who were eligible for FRL. Overall, 18 schools and 140 fourth- and fifth-grade classrooms participated in the study. Student self-report data were collected for all students as part of classroom activities, although only data for students with returned informed consents were analyzed. The rate of return for informed consent was approximately 70% of students in participating classrooms. After approximately 3 months of school, each teacher completed pre-test measures on their students over a period of 2 to 3 weeks prior to beginning the intervention. Teacher self-report measures were completed within the same timeframe. The researchers randomized the order of students

for each measure and asked teachers to complete surveys in the order specified to avoid possible contamination effects of completing multiple surveys in the same order.

Participants for the current study consisted of 982 fourth- and fifth-grade students, 65 teachers, and 11 schools with completed datasets. Approximately 64% of students were Caucasian ($n = 628$), 21% were African American ($n = 209$), 10% were Hispanic ($n = 101$), 4% were Other ($n = 40$), and 0.4% were Asian ($n = 4$). Given the limited sample size for students identified as Other and Asian American, data for these groups were not included in the analyses. Approximately 51% of the students were girls ($n = 505$) and 49% were boys ($n = 477$). In addition, approximately 69% of students ($n = 676$) were eligible for FRL.

Approximately 86% of the teachers were Caucasian ($n = 56$), 11% were African American ($n = 7$), and 3% were Hispanic ($n = 2$). Seventy-seven percent of the teachers were women ($n = 50$) and 23% were men ($n = 15$). The age of teachers ranged from 23 to 69 years ($M = 41.2$, $SD = 11.3$). Teachers' years of experience ranged from 0 (new teachers) to 38 ($M = 9.3$, $SD = 8.2$). Regarding overall school variables, student enrollment in kindergarten-to-fifth grade elementary schools ranged from 411 to 1,108 ($M = 628.6$, $SD = 240.7$). The percent of racial/ethnically diverse students in the schools ranged from 5.6% to 86.5% ($M = 40.8\%$, $SD = 25.8\%$). Finally, the percentage of students eligible for FRL ranged from 59.3% to 88.2% ($M = 73.08$, $SD = 9.61$) across schools.

## Instruments

*Clinical Assessment of Behavior Teacher Form.* The Clinical Assessment of Behavior Teacher Form (CAB-T; Bracken & Keith, 2004) is an omnibus behavior rating scale. The CAB-T is nationally normed and can be used with students who are 5 through 18 years old. Consisting of 70 items, the CAB-T produces scores for two Clinical scales, the Internalizing Behaviors Scale (INT) and the Externalizing Behaviors Scale (EXT), and two Adaptive scales, the Social Skills Scale (SOC) and the Competence Scale (COM), as well as several additional subscales. For the purposes of the current study, the INT, EXT, SOC, and COM (a measure of adjustment and adaptive strength in an area that is closely related to cognitive and academic functioning scales) were used as outcome variables. According to Bracken and Keith (2004), scores on the CAB-T are highly reliable, and cluster interpretations are valid for students across racial/ethnic groups. The Adaptive and Clinical Scale scores were used as the dependent variables in this study.

*Teacher Efficacy in Classroom Management and Discipline Scale.* The Teacher Efficacy in Classroom Management and Discipline Scale (TECMD; Emmer & Hickman, 1991) was developed from a teacher efficacy scale created by Gibson and Dembo (1984). The TECMD contains three subscales. The classroom management and discipline factor assesses teachers' self-perception of their abilities to manage and discipline their students. The external influences factor assesses teachers' beliefs regarding the strength of external influences that are beyond the teacher's control. The personal efficacy factor assesses teachers' perceptions that they can make a difference in students' lives. Emmer and Hickman (1991) showed that their scale has adequate test–retest reliability, internal consistency, and construct validity.

## Data Collection

*Student-Level Data.* Information on students' gender, race/ethnicity, and FRL status (as a measure of family income) was obtained from each student's school district.

*Teacher-Level Data.* Information on teachers' gender, race/ethnicity, age, and years of experience was obtained from each teacher's school district. Additionally, teachers' self-efficacy was measured with the TECMD, and all three subscales were included in the analyses.

*School-Level Data.*    Although school-level data were included to assess variance components at the school level, school-level variables were not included as predictors of differences in ratings, given the difficulty of interpreting three-level interactions among student, teacher, and school variables. Information on school size, percentage of students receiving FRL, and percentage of racial/ethnic minority students in the school were obtained from each school district.

*HLM Analyses*

The use of HLM procedures allows for the estimation of variance and covariance components with nested data (i.e., determining what proportion of variance in teachers' ratings of student behaviors is attributable to the teachers and the schools) and for an analysis of interactions between variables at the different levels. The method of estimation utilized was full information maximum likelihood, with HLM6 software to conduct the analyses (Raudenbush, Bryk, & Congdon, 2004).

As a preliminary analysis, we examined a random effects analysis of variance (ANOVA) model (unconditional three-level model) to determine whether additional variance existed at the teacher (Level 2) and the school (Level 3) levels. If significant intraclass correlation coefficients were identified at Level 2 and Level 3, results would indicate that variables at these levels also contributed to teacher ratings of student behaviors. Therefore, the initial model included student-, teacher-, and school-level variables in an intercepts and slopes as outcomes (ISAO) model.

Based on the ISAO models, we examined the main effects of the Level-1 variables (i.e., student gender, family income, and race/ethnicity), the main effects of the Level-2 variables (i.e., teachers' race/ethnicity, gender, age, years of experience, and self-efficacy), and the main effects of the Level-3 variables (i.e., school size, average family income, and percentage of minority students). We also examined the interactions between the Level-1 and Level-2 variables, which can be interpreted as a Level-2 variable predicting differences among Level-1 variables. Interactions of students' gender and race/ethnicity with teachers' race/ethnicity, gender, age, years of experience, and self-efficacy were examined, controlling for students' family income, school size, average family income, and percentage of racial/ethnic minority students.

Last, we examined the unconditional model and ISAO model four times to evaluate the four outcome variables of the CAB-T (INT, EXT, SOC, and COM). Due to the complexity of the three-level model with multiple variables at each level, we are unable to provide specific equations in this article due to space limitations. We answered the following questions using this approach:

1. What proportion of variance in teacher ratings of student behaviors is associated with the teacher/classroom and the school and what are the intraclass correlations for the teacher/classroom and school levels?
2. What are the mean differences in teacher ratings of student behaviors across student gender, race/ethnicity, and family income?
3. What are the interaction effects of teacher gender, race/ethnicity, age, years of experience, and perceived self-efficacy on teacher-specific differences for racial/ethnic groups and gender groups in teacher-rated student behaviors?

<div align="center">Results</div>

*Random ANOVA Hierarchical Linear Models*

Table 1 displays the results from the three-level random ANOVA models (i.e., no predictors) for the four outcome variables in the study. The random ANOVA models were included to examine the proportion of variance (intraclass correlations) at the student, teacher, and school levels, respectively. For teacher ratings of students' behaviors, the majority of variance for each outcome variable was

Table 1
*Percentage of Variance at Student, Teacher/Classroom, and School Levels*

| Variable | Student Level | Teacher/Classroom Level | School Level |
|---|---|---|---|
| Internalizing | 56.7 | 34.6 | 8.7 |
| Externalizing | 70.7 | 22.6 | 6.6 |
| Social Skills | 73.5 | 22.2 | 4.2 |
| Competence | 79.3 | 20.0 | 0.8 |

attributable to the student level, followed by the teacher/classroom level. Relatively little variance, in contrast, was explained at the school level. However, all variance components at the school level were significant at $p \leq .01$, with the exception of teacher ratings of students' competence.

## ISAO Models

The IASO models were included to examine the main effects of the Level-1 variables (i.e., student gender, race/ethnicity, and family income), the main effects of the Level-2 variables (i.e., teachers' age, years of experience, self-efficacy, gender, and race/ethnicity), and the main effects of the Level-3 variables (i.e., school size, percentage minority students per school, and average family income per school). Additionally, interactions between some Level-1 and all Level-2 variables were examined for significance. Significant interactions can be interpreted as a Level-2 variable predicting a teacher-specific difference for a Level-1 variable. For each outcome variable, the main effects and interaction effects for Level-1 variables are discussed in the following sections. Space limitations restrict presenting the main effects of Level-2 and Level-3 variables, in addition to nonsignificant findings related to interaction variables. In all of the ISAO models, if the residual variance for a variable was nonsignificant, the variance was set to zero and the model was re-run. We chose this method for efficiency, given the size of the three-level models with multiple variables at each level.

*Internalizing.* Table 2 displays the coefficients for the Level-1 variables in the HLM, with the INT scale as the outcome variable. On average, Hispanics were rated as significantly less internalizing than were Caucasians, and males were rated as significantly less internalizing than were females. None of the coefficients for the cross-level interactions between teacher-level variables and racial/ethnic and gender group contrasts were statistically significant. Nonsignificant cross-level

Table 2
*Hierarchical Linear Model with Internalizing as Outcome: Coefficients for Student-Level Independent Variables*

| Fixed Effect | Coefficient | Standard Error | *t* | *df* | *p* |
|---|---|---|---|---|---|
| Intercept[a] | 45.07 | 0.98 | 45.95 | 7 | .00 |
| FRL Eligible Slope[b] | 1.45 | 0.73 | 1.98 | 64 | .05 |
| African American Slope[b] | 0.23 | 0.90 | 0.26 | 57 | .80 |
| Hispanic American Slope[b] | −3.03 | 1.19 | −2.54 | 57 | .01* |
| Male Slope[b] | −2.12 | .64 | −3.33 | 57 | .00* |

*Note.* FRL = Free and reduced-price lunch status.
[a]The residual parameter variance for the parameter has been set to zero. [b]This variable has been centered around its group mean.
*$p < .05$.

Table 3
*Hierarchical Linear Model with Externalizing as Outcome: Coefficients for Student-Level Independent Variables*

| Fixed Effect | Coefficient | Standard Error | t | df | p |
|---|---|---|---|---|---|
| Intercept[a] | 46.03 | .72 | 64.28 | 7 | .00* |
| FRL Eligible Slope[b] | 1.33 | .68 | 1.94 | 64 | .06 |
| African American Slope[b] | 3.00 | .84 | 3.55 | 57 | .00* |
| Hispanic American Slope[b] | − 2.13 | 1.12 | − 1.90 | 57 | .06 |
| Male Slope[b] | − 3.25 | .60 | − 5.43 | 57 | .00* |

*Note.* FRL = Free and reduced-price lunch status.
[a]The residual parameter variance for the parameter has been set to zero. [b]This variable has been centered around its group mean.
*p < .05.

interactions imply that the relationships between scores on the internalizing variable and teacher variables are similar for the groups of students compared.

*Externalizing.* Table 3 displays the coefficients for the Level 1 variables in the HLM, with the EXT scale as the outcome variable. African Americans were rated as significantly more externalizing than were Caucasians, and males were rated as significantly less externalizing than were females. Table 4 displays the coefficients for the cross-level interactions of the teacher-level variables and the difference between African American and Caucasian students. There is a significant interaction between teacher self-efficacy and behavior management. Thus, teacher self-efficacy for behavior management significantly predicted the teacher-specific difference between African American and Caucasian students. The size of the mean difference between African American and Caucasian students on the EXT scale observed for specific teachers was not predicted by any additional teacher-level variables.

Table 5 displays the coefficients for the cross-level interactions of teacher-level variables and the significant difference between male and female students. There was a significant interaction for teacher age, indicating that age significantly predicted the teacher-specific difference between male and female students. The size of the mean difference between male and female students on the EXT scale observed for specific teachers was not predicted by any additional teacher-level variables.

Table 4
*Hierarchical Linear Model with Externalizing as Outcome: Coefficients for Cross-Level Interaction of Teacher-Level Variables and Differences Between African American and Caucasian Students*

| Fixed Effect | Coefficient | Standard Error | t | Approx. df | p |
|---|---|---|---|---|---|
| Teacher Age[a] | .109 | .10 | .99 | 57 | .33 |
| Years of Experience[a] | − .01 | .15 | − .05 | 57 | .96 |
| Management[a] | − 4.42 | 2.18 | − 2.03 | 57 | .05* |
| External[a] | − 1.91 | 1.62 | − 1.17 | 57 | .25 |
| Personal[a] | 2.59 | 1.81 | 1.43 | 57 | .16 |
| Teacher Gender[a] | − 2.71 | 2.16 | − 1.26 | 57 | .22 |
| TeacherRace/Ethnicity[a] | .91 | 2.35 | .398 | 57 | .70 |

*Note.* FRL = Free and reduced-price lunch status.
[a]This variable has been centered around its group mean. The residual parameter variance for the parameter has been set to zero.
*p < .05.

Table 5

*Hierarchical Linear Model with Externalizing as Outcome: Coefficients for Cross-Level Interaction of Teacher-Level Variables and Differences Between Male and Female Students*

| Fixed Effect | Coefficient | Standard Error | $t$ | $df$ | $p$ |
|---|---|---|---|---|---|
| Teacher Age[a] | − .15 | .07 | − 2.15 | 57 | .04* |
| Years Experience[a] | .16 | .10 | 1.50 | 57 | .14 |
| Management[a] | 1.72 | 1.51 | 1.15 | 57 | .26 |
| External[a] | − .10 | 1.11 | − .09 | 57 | .93 |
| Personal[a] | − 2.17 | 1.35 | − 1.61 | 57 | .11 |
| Teacher Gender[a] | − .38 | 1.54 | − .25 | 57 | .81 |
| Teacher Race/Ethnicity[a] | − 2.51 | 1.87 | − 1.34 | 57 | .19 |

*Note*. FRL = Free and reduced-price lunch status.
[a]This variable has been centered around its group mean. The residual parameter variance for the parameter has been set to zero.
*$p < .05$

Table 6

*Hierarchical Linear Model with Social Skills as Outcome: Coefficients for Student-Level Independent Variables*

| Fixed Effect | Coefficient | Standard Error | $t$ | $df$ | $p$ |
|---|---|---|---|---|---|
| Intercept[a] | 52.50 | .73 | 72.33 | 7 | .00 |
| FRL Eligible Slope[b] | − .98 | .69 | − 1.43 | 64 | .16 |
| African American Slope[b] | − 2.89 | .86 | − 3.36 | 57 | .00* |
| Hispanic American Slope[b] | 1.38 | 1.13 | 1.22 | 57 | .23 |
| Male Slope[b] | 2.68 | .60 | 4.45 | 57 | .00 |

*Note*. FRL = Free and reduced-price lunch status.
[a]The residual parameter variance for the parameter has been set to zero. [b]This variable has been centered around its group mean.
*$p < .05$.

*Social Skills*.    Table 6 displays the coefficients for the Level 1 variables in the HLM, with the SOC scale as the outcome variable. African Americans received significantly lower social skills ratings than did their Caucasian peers, and males received significantly higher social skills ratings than did their female peers. Table 7 displays the coefficients for the cross-level interactions of teacher-level variables and the difference between African American and Caucasian students on the SOC scale. There was a significant interaction for teacher self-efficacy for behavior management. This finding suggests that teacher self-efficacy for behavior management significantly predicted the teacher-specific difference between African American and Caucasian students. Mean differences between African American and Caucasian students on the SOC scale observed for specific teachers were not predicted by any additional teacher-level variables. None of the coefficients for the cross-level interactions between teacher-level variables and the contrast of Hispanic and Caucasian students or between teacher-level variables and the contrast of male and female students on the SOC scale were significant.

*Competence*.    Table 8 displays the coefficients for the Level11 variables in the HLM, with the COM scale as the outcome variable. Students who were eligible for FRL received significantly lower competence ratings than did their peers who were not eligible for FRL. African American

Table 7

*Hierarchical Linear Model with Social Skills as Outcome: Coefficients for Cross-Level Interaction of Teacher-Level Variables and Differences Between African American and Caucasian Students*

| Fixed Effect | Coefficient | Standard Error | $t$ | $df$ | $p$ |
|---|---|---|---|---|---|
| Teacher Age[a] | − .15 | 0.10 | − 1.42 | 57 | .16 |
| Years Experience[a] | .06 | 0.16 | .38 | 57 | .70 |
| Management[a] | 4.61 | 2.22 | 2.08 | 57 | .04* |
| External[a] | 1.88 | 1.66 | 1.14 | 57 | .26 |
| Personal[a] | − 3.48 | 1.84 | 1.89 | 57 | .06 |
| Teacher Gender[a] | 2.19 | 2.209 | 1.00 | 57 | .32 |
| Teacher Race/Ethnicity[a] | − 1.80 | 2.41 | − .75 | 57 | .46 |

*Note.* FRL = Free and reduced-price lunch status.
[a]This variable has been centered around its group mean. The residual parameter variance for the parameter has been set to zero.
*$p < .05$.

Table 8

*Hierarchical Linear Model with Competence as Outcome: Coefficients for Student-Level Independent Variables*

| Fixed Effect | Coefficient | Standard Error | $t$ | $df$ | $p$ |
|---|---|---|---|---|---|
| Intercept[a] | 52.91 | .47 | 113.74 | 7 | .00 |
| FRL Eligible Slope[b] | − 1.55 | .73 | − 2.12 | 64 | .04* |
| African American Slope[b] | − 2.24 | .89 | − 2.52 | 57 | .02* |
| Hispanic American Slope[b] | 0.63 | 1.20 | .53 | 57 | .60 |
| Male Slope[b] | 2.48 | .63 | 3.92 | 57 | .00* |

*Note.* FRL = Free and reduced-price lunch status.
[a]The residual parameter variance for the parameter has been set to zero. [b]This variable has been centered around its group mean.
*$p < .05$.

students received significantly lower competence ratings than did Caucasians. Finally, males received significantly higher competence ratings than did females.

None of the coefficients for the cross-level interactions between teacher-level variables and the contrast of African American and Caucasian students on the COM scale were significant. The size of the mean difference between African American and Caucasian students on the COM scale observed for specific teachers was not predicted by any teacher-level variables. Additionally, none of the coefficients for the cross-level interactions between teacher-level variables and the contrast of Hispanic and Caucasian students on the COM scale were significant. Finally, none of the coefficients for the cross-level interactions between teacher-level variables and the contrast of male and female students on the COM scale were significant.

## DISCUSSION

This study examined whether mean-group differences exist among student demographic variables and whether teacher-variables predicted mean-group differences on an omnibus behavior rating scale. We were primarily interested in examining the variance components in teacher-rated behaviors at the student, teacher/classroom, and school levels, but we were also interested in determining whether there were mean-group differences in teacher-rated behaviors across student racial/ethnic groups. In addition, we examined whether teacher-specific differences in ratings could be predicted

by several teacher variables. By examining variance components, we wanted to understand how much the outcomes from a specific behavior rating scale were related to actual student behaviors and how much they could be attributed to other variables, such as teachers and schools.

Results indicated that a majority of variance across all four outcome variables was attributable to the student level. However, we found that a significant amount of variance in behavior ratings for all four outcome variables was attributable to the teacher/classroom level. Our results are consistent with Mashburn et al. (2006)'s findings that a substantial amount of variance in teacher-rated behavior is attributable to differences in teachers or classrooms.

Because differences in ratings on behavior scales between teachers or classrooms reflect a number of variables, we examined five potential variables (i.e., teacher age, years of experience, gender, race/ethnicity, and self-efficacy). Given that 20% to 35% of variance lies at the teacher/classroom level, it is important to identify potential contributors to this variability when decisions are made about students. These results support the best practices model to obtain information regarding student behaviors across multiple locations and multiple raters (Sattler & Hoge, 2006). Finally, results of this study indicated that a significant proportion of variance in teacher-rated behaviors was attributable to the school level, with the exception of the variance component for competence ratings. These findings suggest that variables at the school level contribute to teacher ratings of behavior, the actual behavior of students, or both. In this study, we controlled for school size, percentage of racial/ethnic minority students, and percentage of students eligible for FRL.

In addition to these results, we also examined whether mean-group differences in behavior ratings existed across gender, race/ethnicity, and family income. Several studies have shown that African American students tend to be rated as exhibiting more problematic behavior than their Caucasian peers (e.g., Downy & Pribesh, 2004). Additionally, research findings typically show that males receive higher ratings, indicating more problem behavior, than females, when controlling for race/ethnicity (e.g., Rayfield, 1997; Rong, 1996). It is important to note that we examined mean-group differences based on ratings for all students in classrooms while controlling for gender, family income, and FRL status. Results indicated significant differences between Hispanic and Caucasian students, in addition to significant differences between male and female students on the internalizing variable. That Hispanic students were rated as significantly less internalizing, on average, than were their Caucasian peers was surprising. In a study examining ratings for Hispanic and Caucasian students related to attention deficit/hyperactivity disorder, Dominguez and Shapiro (1998) found a consistent, yet nonsignificant trend across several measures and subscales indicating that Hispanic students received lower mean behavior scale ratings than did their Caucasian peers. They hypothesized that Hispanic families socialize children by instilling values related to obedience and rule following.

Conversely, Glover, Pumariega, Holzer, Wise, and Rodriguez (1999) conducted a study in which adolescents completed self-report measures. Two groups in Texas were compared: one consisting of 94% Hispanic and another consisting of Hispanics, Caucasians, and African Americans. The researchers found that the predominately Hispanic group evidenced significantly higher mean levels of anxiety and, further, that students born outside of the United States evidenced significantly higher mean levels of anxiety than did those born in the United States. They hypothesized that Hispanics, particularly females who are in the process of acculturating, may experience more feelings of internalized distress. An alternative possibility is that teachers are not observant of behavioral indicators of anxiety and depression in Hispanic students, which may contribute to underrepresentation of Hispanic youth identified with EBD (see Cullinan & Kauffman, 2005). Regarding externalizing behaviors, we found that African American students were rated as significantly more externalizing than were their Caucasian peers, on average, controlling for gender and FRL status, consistent with previous research (e.g., Epstein et al., 1998; Sbarra & Pianta, 2001).

It is also interesting to note that males were rated significantly less externalizing than were females, controlling for race/ethnicity and FRL status. We hypothesize that the observed gender differences may be related to the norming process of the CAB-T; females and males have their own norm groups, and scores are reflective of comparisons to gender-specific norms as opposed to coed norms. However, some recent studies have also found support for females exhibiting more externalizing behavior than their male peers (Moller-Leimkuhler & Yucel, 2009; Moretti, Catchpole, & Odgers, 2005). This is certainly an area for future research.

Our results suggested significant differences between African American students and Caucasian students as well as between males and females with regard to ratings of social skills. This finding related to race/ethnicity is similar to Rong's (1996) findings that African American students received lower social skills ratings than did their Caucasian peers. The finding related to gender is in contrast to previous research in that Mashburn et al. (2006) found that teachers rated females as displaying significantly better social skills. Again, the gender difference in our study may be a result of the norming/scoring process of the CAB-T.

Finally, our results suggested significant differences between students eligible for FRL and students who were not eligible for FRL, significant differences between African American students and Caucasian students, and significant differences between males and females on ratings of competence. The link between lower competence ratings and FRL status is no surprise. Barbarin et al. (2006) found that students from poor families displayed significantly lower receptive language and math scores than did their peers. Additionally, they found that students who were not from poor families scored higher on all pre-academic tasks when assessed at the beginning of pre-kindergarten.

We also examined predictors of teacher-specific differences in ratings. Although Hispanic students and males were rated as significantly less internalizing than were their peers, none of the teacher variables (age, years of experience, self-efficacy, race/ethnicity, or gender) significantly predicted the size of the teacher-specific differences for these groups. Considering that approximately 35% of variance in teacher-rated internalizing problems lay at the teacher/classroom level and approximately 9% at the school level, additional research is needed to explore potential predictors of teacher-specific mean differences in internalizing behavior ratings.

For the externalizing variable, African American students were rated as exhibiting significantly more externalizing behaviors. Most of the teacher variables did not significantly predict the teacher-specific differences for the race/ethnicity variables in this study. However, teacher self-efficacy related to classroom management and discipline significantly predicted the teacher-specific differences in ratings for Caucasian and African American students. As teachers' sense of self-efficacy regarding their behavior management and classroom discipline skills increased, the teacher-specific difference between African American students and Caucasian students decreased. One hypothesis is that teachers with higher self-efficacy in behavior management skills do in fact have better skills in the classroom and, therefore, their students exhibit fewer behavior problems. A second hypothesis is that teachers with higher self-efficacy skills relating to behavior management do not perceive behaviors exhibited by African American students as being as problematic as teachers with a lower sense of self-efficacy do. In either case, this outcome underscores the importance of pre-service and in-service training for teachers regarding classroom management and discipline.

Additionally, we found that females were rated as exhibiting significantly more externalizing behaviors than males and that teacher age predicted teacher-specific differences in externalizing behaviors across gender. Research focusing on the increasing rates of externalizing behavior among females is needed to determine whether this is in fact a new trend before suggestions can be made regarding in-service training for teachers regarding behavior management for females exhibiting

these behaviors. An alternative hypothesis for the gender differences in our study relates to the fact that males and females have separate norm groups when scoring the CAB-T.

Furthermore, African American students received significantly lower social skills ratings than did their Caucasian peers, and teacher self-efficacy in classroom management and discipline predicted the teacher-specific differences, similar to the finding that teacher self-efficacy predicted the teacher-specific differences for African American and Caucasian students on externalizing behavior. We found several mean-group differences for school competence: students eligible for FRL, African American students, and females received significantly lower competence ratings. No teacher variables significantly predicted any teacher-specific differences related to these variables, indicating that there are real group differences in school competence. However, as 19% of the variance in competence ratings was attributable to the teacher/classroom level, additional research examining possible predictors is warranted.

Finally, the only teacher variable that predicted a teacher-specific difference in social skills ratings across groups was self-efficacy in classroom management and discipline. This finding suggests that, as teacher self-efficacy for classroom management and discipline increases, the size of the teacher-specific difference in social skills ratings decreases. Similar results were found by Mashburn et al. (2006). Additional research on the relationship between teacher self-efficacy in classroom management and discipline and perceptions of students' social skills is needed.

*Limitations*

First, generalizability of these results is limited to the population of fourth- and fifth-grade students in suburban and rural districts from a large state in the Southeast. Further research is obviously needed with other age groups and from other areas of the country. Similar to many other studies, the teachers in the sample were predominately Caucasian females. Due to the limited number of racial/ethnic minority teachers, we analyzed the data comparing Caucasian teachers' ratings with non-Caucasian teachers' ratings. Our student sample was limited to Caucasian, African American, and Hispanic students due to difficulties obtaining a large enough subgroup for additional race/ethnicities. Additional research examining ratings of additional racial/ethnic groups is needed. It is also important to note that the results of this study are based on teacher rating scales and not direct observation of student behavior. Even with the finding that teacher self-efficacy predicted teacher-specific differences in externalizing ratings for African American and Caucasian students, questions still remain as to whether African American students exhibit fewer problem behaviors when placed with teachers with higher self-efficacy or if the teachers perceive problem behaviors as less problematic. Research comparing teacher ratings of student behaviors in an HLM model with observational data would provide much needed information to help answer the question of whether differences in behavior ratings are evidence of real group differences in behavior or a matter of teachers' perceptions.

## Conclusions

In summary, results of this study indicated that a significant amount of variance in teacher-rated behaviors lies at the teacher/classroom level and at the school level. When school psychologists examine student data, it is important to keep these results in mind. School psychologists are encouraged to follow best practices as related to evaluations and collect behavioral data from a variety of sources. This may include collecting multiple behavior rating scales in addition to collecting other data (e.g., classroom observations, frequency counts). Finally, educational teams are encouraged to carefully consider all data collected as part of multidisciplinary team evaluations for making special education eligibility decisions.

## References

Abidin, R. R., & Robinson, L. L. (2002). Stress, biases, or professionalism: What drives teachers' referral judgments of students with challenging behaviors? Journal of Emotional and Behavioral Disorders, 10, 204–212.

Algozzine, B., Christenson, S., & Yesseldyke, J. E. (1982). Probabilities associated with the referral to placement process. Teacher Education and Special Education, 5, 19–23.

Ashton, P. T., & Webb, R. B. (1986). Making a difference: Teachers' sense of efficacy and student achievement. White Plains, NY: Longman.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychology Review, 84, 191–215.

Barbarin, O., Bryant, D., McCandies, T., Burchinal, M., Early, D., Clifford, R., . . . Howes, C. (2006). Children enrolled in public pre-K: The relation of family life, neighborhood quality, and socioeconomic resources to early competence. American Journal of Orthopsychiatry, 76, 265–276.

Bracken, B. A., & Keith, L. K. (2004). Clinical assessment of behavior. Lutz, FL: Psychological Assessment Resources.

Caldas, S. J., & Bankston, C. (1998). The inequality of separation: Racial composition of schools and academic achievement. Educational Administration Quarterly, 34, 533–557.

Chester, M. D., & Beaudin, B. Q. (1996). Efficacy beliefs of newly hired teachers in urban schools. American Educational Research Journal, 33, 233–257.

Clifton, R. A., Perry, R. P., Parsonson, K., & Hryniuk, A. (1986). Effects of ethnicity and sex on teachers' expectations of junior high school students. Sociology of Education, 59, 58–67.

Cohen, M., DuRant, R. H., & Cook, C. (1988). The Conners teacher rating scale: Effects of age, sex, and race with special education children. Psychology in the Schools, 25, 195–202.

Conners, C. K. (1973). Rating scales for use in drug studies with children. Psychopharmacology Bulletin: Pharmacotherapy with children. Washington, DC: Government Printing Office.

Cullinan, D., & Kauffman, J. M. (2005). Do race of student and race of teacher influence ratings of emotional and behavioral problem characteristics of students with emotional disturbance? Behavioral Disorders, 30, 393–402.

Daunic, A. P., Smith, S. W., Garvin, C. W., Barber, B. R., Becker, M. K., Peters, C. D., . . . Naranjo, A. H. (2012). Reducing developmental risk for emotional/ behavioral problems: A randomized controlled trial examining the tools for getting along curriculum. Journal of School Psychology, 50, 149–166.

Daunic, A. P., Smith, S. W., Brank, E. M., & Penfield, R. D. (2006). Classroom based cognitive-behavioral intervention to prevent aggression: Efficacy and social validity. Journal of School Psychology. 44, 123–139.

Dominguez, R. R., & Shapiro, E. S. (1998). Teacher ratings of attention deficit hyperactivity disorder symptoms in Hispanic children. Journal of Psychopathology and Behavioral Assessment, 20, 275–293.

Donovan, M. S., & Cross, C. T. (Eds.) (2002). Minority students in special and gifted education. Washington, DC: National Academy Press.

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. Sociology of Education, 77, 267–282.

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. Journal of Educational Psychology, 75, 327–346.

Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners teacher rating scale. Journal of Abnormal Child Psychology, 26, 109–118.

Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. Educational and Psychological Measurement, 51, 755–765.

Glover, S. H., Pumariega, A. J., Holzer, C. E., Wise, B. K., & Rodriguez, M. (1999). Anxiety symptomatology in Mexican-American adolescents. Journal of Child and Family Studies, 8, 47–57.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76, 569–582.

Gottlieb, D. (1964). Teaching and students: The views of Negro and Caucasian teachers. Sociology of Education, 37, 345–353.

Kauffman, J. M. (2001). Characteristics of emotional and behavioral disorders of children and youth (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. Journal of Educational Psychology, 100, 96–104.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behavior. Journal of Psychoeducational Assessment, 24, 367–380.

Moller-Leimkuhler, A. M., & Yucel, M. (2009). Male depression in females? Journal of Affective Disorders, 121, 22–29.

Moretti, M. M., Catchpole, R. E. H., & Odgers, C. (2005). The dark side of girlhood: Risk factors and trajectories to aggression and violence. Canadian Child and Adolescent Psychiatry and Review, 14, 21–25.

Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. The Journal of Special Education, 32, 194–206.

Pigott, R. L., & Cowen, E. L. (2000). Teacher race, child race, racial congruence, and teacher ratings of children's school adjustment. Journal of School Psychology, 38, 177–196.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HLM6 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Rayfield, A. (1997). Concurrent validity of the Sutter-Eyeberg student behavior inventory with grade school children (unpublished doctoral dissertation). Gainesville, FL: University of Florida.

Reynolds, C. R., & Kamphaus, R. W. (1992). Behavior assessment scale for children: Manual. Circle Pines, MN: American Guidance Service.

Rong, X. L. (1996). Effects of race and gender on teachers' perception of the social behavior of elementary students. Urban Education, 31, 261–290.

Ryabov, I., & Van Hook, J. (2007). School segregation and academic achievement among Hispanic children. Social Science Research, 36, 767–788.

Sattler, J. M., & Hoge, R. D. (2006). Assessment of children: Behavioral, social, and clinical foundations (5th ed.). La Mesa, CA: Jerome Sattler, Publisher.

Sbarra, D. A., & Pianta, R. C. (2001). Teacher ratings of behavior among African American and Caucasian children during the first two years of school. Psychology in the Schools, 38, 229–238.

Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. Journal of Learning Disabilities, 44, 246–257.