ABSTRACT

| | |
|---|---|
| Title of Document: | INFLUENCE OF STUDENT PROBLEM BEHAVIOR AND TEACHER TOLERANCE ON STUDENT GRADES |
| | Katherine J. Bruckman, Master of Arts, 2010 |
| Directed By: | Professor, Gary Gottfredson, Counseling and Personnel Services |

The current study examines the influence of student problem behavior (as rated by teachers), teacher disposition to tolerate problem behavior, and interaction effects on student grade point average, reading grades, and math grades. The sample includes 3rd through 5th grade students (*N* = 12,993) and their classroom teachers (*N* = 562) from 45 schools. Multilevel models, with students nested within classrooms, test the influence of student problem behavior and teacher tolerance on student grades. Results imply that problem behavior negatively influences grades for students at each grade level, controlling for standardized academic achievement and other student and classroom-level covariates. Results also indicate that low teacher tolerance predicts higher current student grades in some analyses. Finally, tests for interactions of teacher tolerance with student problem behavior indicate that 5th grade students rated as having extreme problem behaviors receive lower grades in classrooms with more tolerant teachers, and higher grades in classrooms with less tolerant teachers.

# INFLUENCE OF STUDENT PROBLEM BEHAVIOR AND TEACHER TOLERANCE ON STUDENT GRADES

By

Katherine J. Bruckman

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2010

Advisory Committee:
Professor Gary Gottfredson, Chair
Professor Sylvia Rosenfield
Associate Professor William Strein

Acknowledgements

Katherine J. Bruckman, Counseling and Personnel Services, University of Maryland.

This research was supported in part by Institute of Education Sciences (IES) grant #R305F050051. Opinions expressed do not necessarily reflect those of the sponsor. I am grateful for the contributions of the IC Teams evaluation research group at the University of Maryland (Gary Gottfredson, Sylvia Rosenfield, Phuong Vu, Jill Berger, Megan Vaganek, Deborah Nelson, Todd Gravois, and Eva Yiu). I especially wish to thank Gary Gottfredson for his support, advice, and manuscript editing. In addition, I want to thank Jennifer Cassata, Jessica Koehler, Kathy Aux, Amy Silverman and the Prince William County schools for their help in collecting the data used for this study.

Correspondence concerning this article should be addressed to Katherine Bruckman, Department of Counseling and Personnel Services, University of Maryland, College Park, MD 20742. E-mail: katejb@umd.edu.

Table of Contents

List of Tables

List of Figures

# Influence of Student Problem Behavior and Teacher Tolerance on Student Grades

Measures of student achievement can show what students know and learn, and can indicate the effectiveness of instruction. Norm-referenced standardized assessments provide one measure of student achievement. These tests are useful for comparing students to same age peers across their state or the nation and can demonstrate students' mastery of the curriculum. Scores that students receive on standardized tests, however, do not necessarily reflect their performance in their classroom (Johnson et al., 2005).

Report card grades, unlike standardized test scores, reflect students' classroom performance and are the primary way that schools and parents communicate about student achievement in the classroom (Friedman & Frisbie, 1995). Because grades are potentially more closely aligned with *classroom* curriculum, they may respond more rapidly than standardized test scores to changes in teaching practices or student learning. Despite the importance of grades for parents and students, current research presents an incomplete account of the factors that determine these grades. What student and teacher characteristics influence the grades that students receive?

## Teacher Grading Practices

Although grades are considered a measure of student achievement, evidence suggests that student behavior and teacher perceptions and beliefs also contribute to student grades. Teachers use judgment when assigning grades and consider many factors (not just achievement) when grading students (Brookhart, 1993). Many school

systems provide teachers with rubrics to guide grading practices, but teachers often do not follow these guidelines (Strein & Meshbesher, 2006).

Research has found that teachers use a "hodgepodge" of factors when grading students, but that academic achievement is generally the largest factor (McMillan, Myran & Workman, 2002; Bursuck et al., 1996; Cross & Frary, 1996; Brookhart, 1994; Strein & Meshbesher, 2006). Generally, researchers investigating teacher grading practices ask teachers to complete questionnaires designed to measure the degree to which they incorporate different factors into their grading (Bursuck et al., 1996; Cross & Frary, 1996; McMillan et al., 2002), or ask teachers to respond to hypothetical grading scenarios (Brookhart, 1993; Brookhart, 1994).

A recent study by McMillan and colleagues (2002) examined grading practices in a large sample of teachers and schools, and will be discussed here in detail both because of its relevance to the current study and because it represents a current trend in research on grading practices. The researchers examined upper elementary (grades 3-5) teachers' assessment and grading practices. The sample included 901 teachers in 105 schools in seven metropolitan Virginia school districts. A teacher self-report questionnaire measured the degree to which teachers use different assessment and grading practices (McMillan et al., 2002). Teachers rated the extent to which they incorporated different factors into their grading practices using a 6-point scale ranging from *not at all* to *completely*. Descriptive analyses summarized teacher responses, and multiple regression and paired *t* tests to examined relations between variables. Teachers reported that student disruptive behavior contributed *very little* to their grading practices. Academic performance and learning

goals mastered contributed *quite a bit* to *extensively* to their grading practices. Other variables, such as student effort, work habits, participation and/or attention, contributed *very little* to *quite a bit*, and had large standard deviations, indicating greater teacher variability in the use of these factors.

Grading practices are often examined by asking teachers either to describe how they grade students or to answer questionnaires about their practices. When asked to self-report their practices, teachers report that academic performance contributes the most to their grading practices (e.g. McMillan et al., 2002). Sometimes, questionnaire studies do not even ask teachers to indicate their use of student behavior when assigning grades (Bursuck, et al., 1996). Although these descriptive studies of grading practices may show teacher perceptions of grading practices, they do not objectively indicate the factors which influence student grades. Teachers may not self-report that they consider problem behavior when grading students, but extensive research shows that problem behavior and grades are related (for review, see Hinshaw, 1992 and Gottfredson, 1981).

**Student Behavior and Student Academic Achievement**

Research consistently finds a relation between student problem behavior and student achievement (Bubb, McCartney & Willett, 2007; Crosby & French, 2002; Gottfredson, 1981; Hinshaw, 1992; Johnson, McGue & Ianoco, 2006). Prior research, however, has not measured these constructs in consistent ways. Studies often measured achievement with group-administered classroom tests, rather than with grades (Bubb et al., 2007; Crosby & French, 2002; Hinshaw, 1992) and measured problem behavior with parent ratings or delinquency records rather than with

classroom behavior measures (Gottfredson, 1981; Johnson et al., 2006). Although some studies do report correlations between classroom problem behavior and grades, these results are not usually the focus of the research (e.g. Gottfredson & Gottfredson, 1999).

Despite extensive behavior-achievement research, the specific relationship between *classroom problem behavior* and student *grades* has not often been pulled apart. Research which examines the influence of classroom problem behavior on student grades can contribute to the literature, and provide results with relevance for students, parents, and classroom teachers.

What accounts for the relation between problem behavior and achievement? The relation may be stronger when achievement is measured with teacher ratings (such as grades) than when measured with standardized tests (Alexander, Entwisle, & Dauber, 1993). This suggests that teachers may take behavior into account when rating student academic performance on the report card. This possibility is plausible. Other research examining the relation between student variables (ethnicity, gender, socioeconomic status, etc.), teacher ratings of achievement, and standardized test scores suggests that the relation between student variables and grades may be stronger than their relation with performance on standardized tests (Beswick, Willms & Sloat, 2005; Stone, 1994).

For example, Beswick and colleagues (2005) investigated the discrepancy between teacher ratings and standardized measures of literacy, and examined whether other variables such as student behavior, family characteristics, or SES could explain the discrepancy. Nine Canadian schools participated in the study, including 205

kindergarteners and 12 teachers. The *Teacher Rating Scale(TRS)—Literacy* (Flynn, 1997) served as a teacher rating of literacy, and the *Word Reading* subtest from the *Wechsler Individual Achievement Test (WIAT)—Second Edition* (Psychological Corporation, 2002) provided a standardized measure. Teachers completed the Conners' Teacher Rating Scale (Conners, 2001), a rating of student behavior, for each student, and schools provided student demographic information including gender, retention, age, and family background. Correlations between behavior scales and teacher-rated literacy were greater than those between the behavior scales and the standardized literacy scores. Researchers also calculated difference scores between the two literacy measures by subtracting standardized raw scores on the TRS from scores on the WIAT. Results showed a discrepancy between the two literacy measures. In addition, behavior ratings significantly predicted the discrepancy, above and beyond the influence of child gender, parental education, and mother's work. The researchers concluded that teacher ratings seem affected by child and family characteristics and that child gender and behavior were most influential.

Although the study points to child factors, other than achievement, which influence teacher ratings of academic skills, some considerations limit causal inferences which can be made. One serious limitation is the reliance on discrepancy scores, which are unreliable. It would make more sense to look for predictors of the two different measures of literacy, and compare the regression coefficients, or to use hierarchical multiple regression, rather than to look for predictors of the discrepancy scores for the two literacy measures.

**Teacher Tolerance of Problem Behavior**

Research has found that students with behavioral problems tend to have lower academic achievement (as measured both by standardized achievement measures and grades) than those without behavior problems (e.g. Crosby & French, 2002; Gottfredson, 1981). This relation may be based in part on the influence of student problem behavior on teacher perceptions and ratings of academic achievement.

Teachers may be distressed or frustrated with "problem" students, and these feelings may influence the grades they assign these students. Research suggests that teachers in general are most distressed by externalizing behavior, compared to internalizing behavior or other problem behavior (Liljequist & Renk, 2007; Safran & Safran, 1984) and also that individual teachers differ in the degree to which they tolerate these problem behaviors in their classroom (Algozzine & Curran, 1979). Teacher tolerance could moderate the association between problem behavior and ratings of academic achievement found in other research. Specifically, the negative association between problem behavior and achievement may be weaker in classrooms taught by more tolerant teachers.

The "teacher tolerance phenomenon" was first investigated in the years following the Education for the Handicapped Act (EHA; 1975), which guaranteed all children age 5-21 the right to public education. Algozzine (1983, 2003a, 2003b; Algozzine, Mercer & Countermine, 1977) began his research in the 1970's at the time when learning and behavior disordered students were first mainstreamed into regular education classrooms (Safran & Safran, 1984) This was a significant change for

classroom teachers because “disordered” children had previously been educated in self-contained classrooms or separate schools.

Researchers have hypothesized that teacher tolerance for problem behavior may influence their ratings of “disturbing” children. Algozzine and Curran (1979) examined the relation between teacher tolerance of disturbing behavior and teacher-rated "interaction potential" for a hypothetical student. The researchers defined “interaction potential” as the hypothetical student’s chance of success in a regular education classroom. Teachers rated the “interaction potential” for a hypothetical “problem” student who was described in a brief written case study and also completed the Disturbing Behavior Checklist—I (Algozzine, 2003b), which measured teacher tolerance for social defiance and social facility. The Disturbing Behavior Checklist has demonstrated high reliability (alpha=.93) in validation studies, and lists 55 behaviors intended to represent common childhood behavioral problems (Algozzine, 2003b). The rater is asked to indicate the relative “disturbingness” of those behaviors using a five point Likert-type scale from *not very disturbing* to *very disturbing*.

Results for the 44 first- through sixth-grade general education teachers who participated showed a main effect for teacher tolerance; tolerant teachers gave better ratings to the hypothetical student. The study had several important limitations. The dependent variable was a teacher rating for a *hypothetical* child, sample size was small ($N$=44), and teacher tolerance was measured by self-report, rather than with objective measures of teacher behavioral responses to students.

In a more recent study, Liljequist and Renk (2007) examined factors related to teacher ratings of student emotional and behavioral problems. Teachers were asked to select a student from their classroom and then completed two rating scales for the student. One rating scale measured the student's emotional and behavioral functioning and asked the teacher to rate how often different behaviors were true of the student. The second rating scale asked teachers to rate the degree to which they were bothered by different types of student behavior measured with the first scale. Overall, teachers reported greater distress for student externalizing behavior problems than internalizing problems. Teachers distressed by externalizing problem behaviors also tended to be distressed by internalizing behaviors, while teachers who were less distressed by one type of behavior also tended to be less distressed by the other. Although this study did not focus specifically on teacher tolerance, the results suggest that some teachers may be more tolerant of problem behaviors than others, and that they may be more tolerant of internalizing problems, and less tolerant of externalizing ones.

Teacher tolerance of problem behavior seems to be a teacher characteristic that influences teacher perceptions of students. However, studies have not examined the effect of teacher tolerance on the way that teachers grade students in their classrooms. Prior research shows a relation between student problem behavior and academic achievement (e.g. Crosby & French, 2002). It is possible that teacher tolerance could moderate this relation such that the relation is weaker in classroom with more tolerant teachers.

## Limitations of Prior Research

There are some general limitations of the prior research examining grading practices, problem behavior, and teacher tolerance. Although past research has examined the effects of student behavior on achievement and investigated teacher tolerance of problem behavior, it has not examined the combined effect of these two factors (problem behavior and teacher tolerance) on grades. In addition, prior studies did not simultaneously model the individual student-level and the classroom-level effects on grades. Student grades are likely a function of both student characteristics and of characteristics that are common to all students in a classroom

## Purpose

The purpose of the current study is to examine the proportion of variance in grades that is explained by teacher-rated problem behavior, controlling for standardized achievement scores and other variables. The second purpose is to examine the effect of teacher tolerance on student grades and the extent to which teacher tolerance moderates the link between grades and problem behavior.

## Research Questions

1. Do problem behavior ratings significantly predict student grades beyond the influence of measured academic achievement?
2. Does teacher tolerance of problem behavior predict student grades?
3. Does teacher tolerance of problem behavior moderate the relation between student behavior ratings and student grades such that the relation is weaker in classrooms with tolerant teachers?

# Method

## Participants

Teachers and students for this study were drawn from the participants in large-scale experimental evaluation of Instructional Consultation Teams (IC Teams, Rosenfield & Gottfredson, 2004). The experimental study had four data collection waves: Pre-intervention baseline (2005-06), Year 1 intervention (2006-07), Year 2 intervention (2007-08), and Year 3 intervention (2008-2009). Data for the current study came from the 2005-06, 2006-07, and 2007-08 school years. All outcomes came from the 2007-08 year. A study which examined effects of the IC Teams intervention did not find significant effects on student grades for the 2007-08 year (Bruckman et al., 2009). Participants for the current study come from 45 schools in Prince William County, VA and consist of students in grades 3 through 5 ($N$ = 12, 993), and their classroom teachers ($N$ = 562). Table 1 details participant characteristics.

## Procedures

Data for this study were collected through two methods. Student and teacher demographic information, student grades, and student standardized test scores were provided by Prince William County Public Schools. Student behavioral information was collected using a Teacher Report on Student Behavior (TRSB) survey. The TRSB was administered via Prince William County Schools intranet in February of three consecutive years (2006-08) and was managed and monitored by school personnel. School officials requested that classroom teachers complete a TRSB report for each student in their classroom.

## Measures

The externalizing behavior scale from the TRSB survey provided teacher ratings of student problem behavior. School system archival files provided student report card grades, standardized achievement test scores, and demographic information. A measure of teacher tolerance was created for this study. The method for creating this new measure is described below. Table 2 summarizes descriptive statistics and reliabilities for student- and classroom-level variables, all of which are shown in standardized score form ($M = 0$, $SD = 1$).

## Student-Level Variables

**Student grades (outcome year).** Third through fifth grade students receive grades ranging from "F" (failure) to "A+" (outstanding). Teachers in Prince William County assign grades according to a detailed rubric. This rubric outlines levels of performance (academic and effort) that correspond to each grade (see Appendices A and B). The rubric instructs teachers to consider the student's "achievement in subject," "class performance," and "independence in work" when assigning grades. The "achievement in subject" category includes mastery of academic material and class objectives. The "class performance" category includes participation, effort, neatness, timely submission of work, and originality of thinking and expression. The "independence in work" category includes self-direction, completion of independent work in addition to required assignments, timely completion of work, and need for encouragement to complete tasks. Although the rubric does instruct teachers to incorporate student "independent" behaviors into their grading, it does not instruct

them to incorporate conduct or behavior problems. Students receive a single omnibus grade for each subject area.

Report cards include grades for math, reading, science, social studies and writing across four marking quarters, and a final year-end grade for each subject area. Student grades in the outcome year (2007-08) were converted to numeric values using the following conversion: A=4, B+=3.4, B=3, C+=2.4, C=2, D+=1.4, D=1, F=0. Numeric grades were then averaged across subjects and marking quarters to calculate an overall grade point average (GPA) for each student. Final reading and math grades served as additional subject-specific report card grade outcomes.

**Problem behavior.** The externalizing behavior scale from the TRSB survey measured student problem behavior. The externalizing behavior scale consists of 8 items measuring the degree to which students are able to regulate their behavior, emotions and interactions with other people. This scale is adapted for the present research from the TOCA-R (Werthamer-Larsson, Kellam & Wheeler, 1991). Sample items are "Takes others' property without permission," Is physically aggressive or in fights with others," and "Defies teachers or other school personnel." Items are rated on a four point Likert-type scale (0=Never/Almost never, 1=Sometimes, 2=Often and 3=Very Often).

**Standardized achievement scores**. Standards of Learning (SOL) scores for the 2007-08 school year provided a measure standardized academic achievement. The SOL assesses student achievement in reading, math, science, history and writing, and reports standard scores for each test subject. Scores on the different subject tests were averaged to create a composite SOL achievement score for each student. In

addition, standard scores for the reading and math subject tests provided subject-specific achievement measures. The SOL composite variable served as a covariate when student GPA was the outcome, SOL reading served as the covariate when reading grades was the outcome, and SOL math served as the covariate when math grades was the outcome.

**Student demographic characteristics**. Preliminary analyses examined student ethnicity, English speakers of other languages (ESOL) status, and free and/or reduced meals (FARM) status as potential covariates for predicting student grades. All covariates were significantly correlated with the outcomes (student GPA, reading grades and math grades) and significantly correlated with each other. To avoid potential problems with multicolinearity, FARM was selected as the single final covariate since it consistently had the highest correlation with outcomes. The FARM variable is coded 1 for students who receive free and/or reduced meals, and coded 0 for those who do not.

**Classroom-Level Variables**

**Teacher tolerance**. Previous research measured teacher tolerance using self-report measures (e.g. Algozzine & Curran, 1979; Safran & Safran, 1984). Teachers reported the degree to which different problem behaviors were disturbing to them. The self-report method has limitations for measuring teacher tolerance, as noted earlier

Another way to measure teacher tolerance of problem behavior is to examine teachers' ratings of problem behaviors for the students in their classrooms. Teacher ratings of student behavior, when summed to the classroom level, provide a measure

of how teachers tend to perceive student behavior in their classrooms. Teachers who show a pattern of rating their students more harshly (i.e. as having more serious behavioral problems) than previous teachers of the same students could be less tolerant of problem behaviors. Similarly, teachers who rate their students less harshly (i.e., as having fewer behavioral problems) than the students' previous teachers might be more tolerant of problem behaviors.

A teacher tolerance measure was created to indicate the way that teachers tend to rate the behavior of their students, as compared the ratings those students received from previous teachers. This teacher tolerance measure represents the degree to which teachers perceive their students' behavior as being more or less problematic than predicted based on the behavior ratings that the same students received from previous teachers.

The first step in creating the teacher tolerance variable was to use regression to predict student problem behavior ratings in the outcome year (2007-08) from the ratings they received from previous teachers in the 2005-06 and 2006-07 school years. Behavior ratings from 05-06 and 06-07 (two different teachers) were averaged together, and this mean was entered as the sole predictor of behavior ratings in the outcome year. The regression yielded a residual for each student. These residuals indicate whether the student's outcome year behavior rating is above or below the level which would be predicted based on the ratings they received from previous teachers. Next, the student-level residuals for 2007-08 were aggregated to the classroom level, so that so that *each teacher* had an aggregated student residual score. This teacher aggregate of residuals provides the measure of teacher tolerance.[1]

**Classroom demographic characteristics**. Preliminary analyses examined class proportion FARM, proportion ESOL, and proportion advantaged ethnicity (Caucasian/White) as potential covariates at the classroom level. Because these three variables are highly correlated at the classroom-level, they would likely introduce multicolinearity if included together in the multilevel models. All three covariates significantly predicted student grades when entered individually into regression models. When they were entered together, however, only proportion FARM and proportion advantaged ethnicity remained significant predictors. In the end, FARM was selected as the covariate, and advantaged ethnicity was disregarded because FARM was most highly correlated with outcomes. The FARM covariate was calculated by aggregating student FARM status (0 = not receiving FARM, 1 = receiving FARM) to the classroom level to get a classroom proportion receiving FARM.

**Standardized achievement and FARM composite.** A classroom mean academic achievement covariate was calculated by aggregating the individual student SOL composite to the classroom level. Initially, analyses included mean classroom academic achievement and proportion FARM as separate covariates at the classroom level. However, the correlation between the SOL and FARM classroom covariates was high ($r$ = -.64), and examination of standardized coefficients indicated that multicolinearity might be a problem. To reduce the threat of multicolinearity, the mean class academic achievement and proportion FARM covariates were combined into a composite. This composite was created by standardizing the covariates at the classroom level, adding the standardized scores together to combine the two

variables, and then standardizing again. The composite variable was called "SOL and FARM composite."

**Prior problem behavior**. The classroom prior problem behavior variable is an aggregated variable for each classroom that measures the mean prior behavior of all the students in the classroom (as rated by their prior teachers). The variable was created by averaging student prior problem behavior ratings for the 2005-06 school year with ratings from the 2006-07. This average of ratings from 05-06 and 06-07 was then aggregated to the outcome year (2007-08) classroom level. For students missing a problem behavior rating from either 2005-06 or 2006-07, the single complete rating was used. Students who did not have prior behavior ratings for either year did not contribute data to the classroom-level prior problem behavior covariate.

**Data Analysis**

Two-level hierarchical linear models (Raudenbush & Bryck, 2002), with students nested within classrooms, tested the effects of student problem behavior and teacher tolerance on grades and the effect of teacher tolerance on the association between problem behavior and grades.

The manner in which the level-1 variables were centered differed depending on the analysis, and is explained below. For simplicity of exposition, equation 1 simply shows the level-1 covariates, and not how they were centered. The equations shown below are written for the GPA outcome; equations for the reading and math grades outcomes were nearly identical.

The level-1 model is:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + \beta_{3j}(X_{3ij}) + + r_{ij} \quad (1)$$

where $Y_{ij}$ is the GPA for the $i^{th}$ student in the $j^{th}$ classroom,

$\beta_{0j}$ is the intercept or the average covariate-adjusted student GPA in the $j^{th}$ classroom if the covariates at level-1 are grand centered,

$\beta_{1j}$ is the slope for the regression of student GPA on the covariate (student SOL achievement) in the $j^{th}$ classroom,

$\beta_{2j}$ is the slope for the regression of student GPA on the covariate (student problem behavior) in the $j^{th}$ classroom,

$\beta_{3j}$ is the slope for the regression of student GPA on the covariate (student FARM status) in the $j^{th}$ classroom,

$X_{1ij}$ is student SOL academic achievement (outcome year) for student $i$ in classroom $j$,

$X_{2ij}$ is student problem behavior (outcome year) for student $i$ in classroom $j$,

$X_{3ij}$ is student FARM status for student $i$ in classroom $j$, and

$r_{ij}$ is residual error for student $i$ in classroom $j$.

The level-2 model consists of 4 equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \gamma_{03}W_{3j} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + \gamma_{13}W_{3j} + u_{1j} \quad (3)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}W_{1j} + \gamma_{22}W_{2j} + \gamma_{23}W_{3j} + u_{2j} \quad (4)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}W_{1j} + \gamma_{32}W_{2j} + \gamma_{33}W_{3j} + u_{3j} \quad (5)$$

Where $\beta_{0j}$ is the average covariate-adjusted student GPA in the $j^{th}$ classroom,

$\gamma_{00}$ is the grand mean student GPA for all classrooms,

$\gamma_{01}$ through $\gamma_{03}$ are the increment to the average student GPA for $W_1$ to $W_3$,

$\beta_{1j}$ is the slope in the partial regression of current GPA on current SOL standardized achievement in the $j^{th}$ classroom,

$\gamma_{10}$ is the average slope in the partial regression of current GPA on current SOL standardized achievement for all classrooms,

$\gamma_{11}$ through $\gamma_{13}$ are the increment to the slope in the partial regression of current GPA on current SOL standardized achievement for a unit change in $W_1$ to $W_3$,

$\beta_{2j}$ is the slope in the partial regression of current GPA on current student problem behavior in the $j^{th}$ classroom,

$\gamma_{20}$ is the average slope in the partial regression of current GPA on current student problem behavior for all classrooms,

$\gamma_{21}$ through $\gamma_{23}$ are the increment to the slope in the partial regression of current GPA on current student problem behavior for a unit change in $W_1$ to $W_3$,

$\beta_{3j}$ is the slope in the partial regression of current GPA on student FARM status in the $j^{th}$ classroom,

$\gamma_{30}$ is the average slope in the partial regression of current GPA on student FARM status for all classrooms,

$\gamma_{31}$ through $\gamma_{33}$ are the increment to the slope in the partial regression of current GPA on student FARM status for a unit change in $W_1$ to $W_3$,

$W_1$ is teacher tolerance,

$W_2$ is the classroom mean of prior student problem behavior,

$W_3$ is the classroom-level composite of classroom SOL achievement and classroom proportion FARM, and

$u_{1j}$ to $u_{3j}$ are the residual errors for classroom $j$.

Analyses proceeded in stages. I built my level-1 model by testing problem behavior ratings, standardized achievement scores, and FARM as potential covariates

and kept only those that significantly predicted student grades. Next, I tested the homogeneity of student-level slopes. For covariates with significantly varying slopes, the slopes were left free to vary in future analyses. In the next step, I examined the relation between student grades and the following potential classroom-level covariates: classroom prior problem behavior, mean classroom SOL achievement and proportion FARM. Classroom-level variables that significantly predicted student grades were retained in the model. After building the student and classroom-level models, I tested my three research questions. All analyses were conducted separately for each grade level.

For the main analysis, all variables were standardized, resulting in standardized coefficients. The standardized coefficients provide effect estimates for the predictors and are appropriate for comparing effects across the GPA, reading, and math grades outcomes *within grade levels*. Supplementary analyses included predictor and outcomes variables in raw metric to calculate unstandardized coefficients, which are recommended for making comparisons *across grade levels* (i.e. comparing problem behavior coefficents for 3$^{rd}$ grade to those for 4$^{th}$ grade; Richards, 1982). Statistical models were the same for both the main and supplemental analyses, except that the outcome, problem behavior, and teacher tolerance variables were entered in raw metric for the latter.

**Analysis for Question 1: Do problem behavior ratings significantly predict student grades, beyond the influence of student academic achievement?** One set of models tested Question 1. In the Level 1 model, the problem behavior predictor was group-mean centered, while the FARM and SOL covariates were

grand-mean centered. The Level 2 model was simplified to include only classroom-level control variables (and not teacher tolerance). All Level 2 variables were grand-mean centered.

Decisions for centering Level 1 predictors were based on recommendations by Enders and Tofighi (2007). According to these researchers, group-mean centering "may be the most appropriate form of centering in situations in which the primary substantive interest involves a Level 1 (i.e., person level) predictor" because it provides an "unbiased" estimate of the Level 1 relationship (p. 128). Using classroom-mean centering provides an estimate of the pooled within-classroom regression of problem behavior on grades.[2]

**Analysis for Question 2: Does teacher tolerance predict student grades?** A separate set of models tested Question 2. All Level 1 and Level 2 predictors were grand-mean centered for analyses answering Question 2, resulting in models similar to those described above. Grand-mean centering at level-1 is required so that $\beta_{0j}$ is covariate adjusted when testing the effect of the level-2 variable (teacher tolerance) on the outcome variables (student grades).

**Analysis for Question 3: Does teacher tolerance moderate the relation between student behavior ratings and student grades?** A final set of models tested Question 3. These models examined the cross-level interaction of teacher tolerance with student behavior ratings on grades. At Level 1, the problem behavior variable was group-mean centered in order to test the cross-level interaction. Other Level 1 covariates were grand-mean centered, and all Level 2 variables were grand mean centered.

# Results

## Correlations

Bivariate correlations between student-level and classroom-level variables were similar across the grade levels. For simplicity of presentation, Table 3 shows these correlations for grades 3, 4 and 5 combined. All student-level correlations were significant at the .05 level. The correlations between student-level variables ranged from 0.08 (problem behavior and FARM status) to 0.84 (reading grades and overall GPA). Correlations between student-level covariates used in any of the statistical models were all weaker than -0.35. The correlation between student problem behavior and report card outcomes was -0.27 for GPA, -0.26 for reading grades, and -0.23 for math grades.

At the classroom level, the correlation between teacher tolerance and the SOL and FARM composite variable was -0.09, and the correlation between teacher tolerance and prior problem behavior was also -0.09. These correlations were both significant at the .05 level. The correlation between prior problem behavior and the SOL and FARM composite variable, however, was not significant ($r$ = 0.04).

## Proportion of Variance Explained

The total variance in the student grade outcomes (GPA, reading and math) can be partitioned into within-classroom and between-classroom components. Variance within classrooms describes individual differences in grades among students in classrooms, and variance between classrooms describes how average grades differ between classrooms. The intraclass correlation coefficient measures the proportion of variance in grades that lies between classrooms. Variance results obtained by

estimating the unconditional models indicate that, on average for all grade levels, 18% of the variance in GPA, 17% of the variance in reading grades, and 16% of the variance in math grades lies between classrooms. These ICC values indicate that there is adequate between classroom variance in student grades to explain with multilevel models.

Results from the fully unconditional and the final models were used to estimate the proportion of variance explained within and between classrooms (Table 4). These proportions of variance explained were calculated separately by outcome and grade level for the three sets of models which tested the three research questions. Because variance results were nearly identical across models, those results will be summarized here. For a detailed breakdown showing proportions of variance explained according to outcome, grade level and model, see Table 4. The final models explained, on average, 68% of the within classroom and 15% of the between classroom variance in GPA, 45% of the within classroom and 28% of the between classroom variance in reading grades, and 56% of the within classroom and 9% of the between classroom variance in math grades. Proportion within classroom variance explained was nearly equal across grade levels for all outcomes and for all three models. However, proportion variance explained between classrooms varied across grade levels, and was lowest for all outcomes for $5^{th}$ grade.

**Question 1: Do problem behavior ratings significantly predict student grades beyond the influence of measured academic achievement?**

One set of models tested the relation between problem behavior and student grades, controlling for student academic achievement and free and/or reduced meals

(FARM) status at the student-level, and classroom mean prior problem behavior, classroom mean standardized achievement and classroom proportion FARM. These results are shown in Tables 5, 6, and 7 in the columns labeled "Model 2, Testing Question 1." Results indicate that students rated as having problem behaviors receive significantly lower grades across all grade levels. Coefficients are all significant ($p <$ .001) and range from -0.11 to -0.25, controlling for current year student standardized achievement scores, student FARM status, classroom mean prior problem behavior, classroom mean standardized achievement, and classroom proportion FARM.

Some trends emerged after review of the results. The relation between student problem behavior and grades (controlling for all other variables) was highest for reading grades, lower for GPA, and lowest for math grades. This trend appeared within each grade level. In addition, Table 8 shows a comparison of coefficients across grade levels and indicates that the relation between problem behavior and reading grades increases as grade level increases. This trend did not hold for the GPA and math grades outcomes.

**Question 2: Does teacher tolerance of problem behavior predict student grades?**

An additional set of models tested the influence of teacher tolerance on student GPA, reading and math grades, controlling for all other variables. These results are shown separately for grades 3, 4, and 5 in tables 5, 6, and 7 respectively (in columns labeled "models testing question 2). Although not all of the teacher tolerance coefficients reach statistical significance, they are all positive, indicating that students with less tolerant teachers tend to have higher grades. When interpreting teacher tolerance coefficients, it is important to remember that the tolerance variable is coded

such that teachers with low scores are more tolerant, while those with high scores are less tolerant.

Results indicate that for 5th grade (see Table 7), teacher tolerance significantly influences reading grades (coef. = 0.07, $p < .01$) and math grades (coef. = 0.06, $p < .05$), and approaches significance for GPA (coef. = 0.06, $p < .10$). For 4th grade, teacher tolerance significantly influences math grades (coef. = 0.08, $p < .05$), approaches significance for GPA (coef. = .06, $p < .10$), and is nonsignificant for reading grades. Teacher tolerance does not significantly predict any grades outcomes for 3rd grade students. For some outcomes, in some grade levels, students receive higher grades when their teachers have higher scores on the teacher tolerance variable.

Coefficients for the teacher tolerance variable show the increase in grades that accompanies a one standard deviation increase in the teacher tolerance variable. For example, the coefficient for predicting grade 5 reading grades from teacher tolerance indicates that for a one standard deviation increase in the teacher tolerance variable, there is a .07 standard deviation increase in reading grades (standardized), controlling for all other variables.

Comparisons of teacher tolerance coefficients across grade levels do not indicate a clear pattern (see Table 8). Coefficients for all outcomes are lowest in 3rd grade, but in 4th and 5th grade, the pattern is less clear. Although coefficients are highest for GPA and math grades outcomes in 4th grade, they are highest for reading grades in 5th grade.

**Question 3. Does teacher tolerance of problem behavior moderate the relation between student behavior ratings and student grades?**

A final set of models tested whether teacher tolerance moderates the relation between student problem behavior and student grades. Specifically, these models tested the cross-level interaction of teacher tolerance (classroom-level variable) with problem behavior (student-level variable). These results are displayed for grades 3, 4, and 5 in tables 5, 6, and 7 respectively, in the columns labeled "Model 4, Testing Question 3." Although interaction effects were tested for all grade levels and outcomes, results will only be discussed and interpreted for grade levels and outcomes for which teacher tolerance significantly influenced grades (grade 4 math grades, grade 5 reading grades, and grade 5 math grades). Interaction effects were not significant for the 4$^{th}$ grade math outcome. However, the interaction effects were significant for the 5$^{th}$ grade reading and math outcomes.

Results indicate that for 5$^{th}$ grade, teacher tolerance significantly moderates the relation between problem behavior and reading grades (coef. =0.05, $p < .01$) and the relation between problem behavior and math grades (coef. =0.04, $p < .01$). The coefficients imply that the relation between grades and problem behavior is more negative (stronger) in classrooms with more tolerant teachers, and less negative (weaker) with less tolerant teachers. Said another way, students rated as having extreme problem behaviors receive lower grades in classrooms with more tolerant teachers, and higher grades in classrooms with less tolerant teachers. These interaction effects are further described in Figures 1 and 2.

## Discussion

The current study reports results from multilevel analyses testing the effects of student problem behavior and teacher tolerance on grades, and the effect of teacher tolerance on the association between problem behavior and grades, for grades 3 through 5.

**Do problem behavior ratings significantly predict student grades beyond the influence of measured academic achievement?**

The current study contributes to our understanding of the factors that contribute to student report card grades, and shows that the relation between grades and problem behavior does not occur solely because problem behavior students learn less or have lower academic achievement. Controlling for concurrent standardized achievement scores and for additional student- and classroom-level covariates, problem behavior consistently and significantly predicts student grades. Results from this study are consistent with past research, which also consistently finds a relation between student problem behavior and student achievement measures (Bubb, McCartney & Willett, 2007; Crosby & French, 2002; Gottfredson, 1981; Hinshaw, 1992; Johnson, McGue & Ianoco, 2006). What distinguishes this study from previous research is its use of report card grades as the measure of academic achievement and concurrent standardized achievement scores as a covariate. In addition, the current study is unique in that it models both the within classroom and between classroom variance in report card grade outcomes, and includes classroom-level covariates.

Even after controlling for measured academic achievement *in the outcome year*, the influence of problem behavior on grades persisted. Several supplemental analyses were conducted to examine the influence of problem behavior on standardized test scores, controlling for student grades. The association between behavior and test scores appears smaller than the relation between behavior and grades, and this relation became nonsignificant once student grades were added as a covariate predicting test scores.

**Does teacher tolerance of problem behavior predict student grades?**

Results for the influence of teacher tolerance on student grades were opposite of the direction hypothesized. Fifth grade students receive higher math and reading grades in classrooms with less tolerant teachers, and fourth grade students receive higher math grades with less tolerant teachers. Although I hypothesized that students would receive lower grades if their teachers were less tolerant, in hindsight, the results obtained here seem consistent with prior research on teacher tolerance. Most of the prior research focused on teacher tolerance of *extreme* problem behaviors which students previously excluded from the general education environment might be expected to demonstrate (e.g. Algozinne & Curran, 1979). The population of students involved in the current study mostly included well-behaved students, and only included a few students rated on the extreme end of the problem behavior scale. Findings from previous studies, therefore, might not apply to the sample in the current study.

Another possibility is that my teacher tolerance variable does not actually measure teacher tolerance, and measures some other construct such as teacher strictness or classroom management skills.

**Does teacher tolerance of problem behavior moderate the relation between student behavior ratings and student grades?**

I hypothesized that the relation between problem behavior and grades would be weaker in classrooms with tolerant teachers, and stronger in classrooms with less tolerant teachers. What I found, however, was the opposite. In classrooms with less tolerant teachers, the slope of the regression predicting grades from problem behavior is *weaker*. This means that students with extreme behavior problems get *higher* grades in classrooms with *less tolerant* teachers, and *lower* grades in classrooms with *more tolerant* teachers. This finding seems inconsistent with prior research on teacher tolerance, which found that teachers who were less tolerant of "disturbing" behavior gave students *lower* ratings (Algozzine & Curran, 1979). As discussed above, the results in the current study could differ from prior research on teacher tolerance because my sample included mostly well-behaved students, so effects for those with extreme behavior problems may not be well estimated.

As discussed above, it is also possible that the teacher tolerance variable measures something other than what I intended. I attempted to measure teacher tolerance creatively, using the data available from prior years problem behavior ratings. Perhaps seemingly "tolerant" teachers are better at classroom management than other teachers and so actually get less problem behavior from their students. Another possibility is that my teacher tolerance variable actually measures teacher

strictness. Seemingly "intolerant" teachers might have stricter behavioral standards and expectations. My ability to measure the teacher tolerance construct was limited because I did not have any attitudinal or observational measures of the teacher's tolerance. These additional measures could help me to validate my teacher tolerance measure in the future. Limited construct validity of the teacher tolerance variable, as currently measured, could account for my unexpected findings.

**Potential Limitations**

Ambiguous temporal precedence of the measured variables (grades, behavior ratings, achievement scores and teacher tolerance) is an internal-validity threat to causal inference. Here I propose that behavior problems cause students to receive lower grades. However, it is possible that a negative relation between problem behavior and grades is caused by other factors not measured in this study or that low teacher grades cause problem behavior.

As with most social science research, it is possible that the variables I use in this study do not measure the constructs that I intend to measure. Student problem behavior measures are teacher reports. These reports represent teacher perceptions of student behavior, and do not necessarily capture the actual incidence of problem behavior. As discussed earlier, construct validity of the teacher tolerance variable is limited, and this variable should be interpreted with caution.

The results are restricted to third through fifth grade students and teachers in a suburban public school system in the mid-Atlantic region and to the demographic characteristics unique to this school system, such as SES percentages, ethnicity classifications, or teacher education. Results may not generalize to students younger

or older than those in this study. In addition, results may not generalize to teachers of other grade levels. This study only includes students from grades 3-5 because the standardized achievement tests are only give to Prince William County elementary students in third, fourth, and fifth grade. Future studies should examine K-2 and 6-12 grade students and teachers.

Finally, skewed distributions for some variables (i.e. problem behavior, SOL, and grades) violate the assumption of normality for regression, which threatens statistical conclusion validity. Results should be interpreted with caution because the skewed distributions could inflate significance values.

**Implications and Future Directions**

Results of the current study have implications for understanding how teachers assign grades, and how well student grades reflect the categories outlined on grading rubrics. The grading rubric for Prince William County (see Appendices A & B) does not instruct teachers to incorporate their assessments of student problem behavior into student report card grades, yet the current study shows that problem behavior ratings do account for a significant portion of variance in grades. There are several explanations for this finding. First, it is possible that teachers include problem behavior when they assign report card grades (either consciously or unconsciously), even though they are not instructed to do so. Second, it is possible that student problem behavior is related to student effort, neatness, ability to work independently, and class participation (behaviors which county teachers *are* instructed to consider), and that these other unmeasured variables actually account for the problem behavior-grades relation. Future research should examine whether problem behavior

influences grades above and beyond student standardized achievement scores *and* student effort/concentration. In addition, future research might ask Prince William County teachers to describe their grading practices, or to respond to a self-report questionnaire, similar to McMillan and colleagues (2002), to indicate the factors they consider when grading.

Grades are often overlooked as achievement measures because they are "contaminated" with additional information, such as behavior, effort, or biases. Results from this study support this idea that grades include factors other than academic achievement; the problem behavior factor seems to make a robust contribution to grades across subject areas and grade levels. These results have implications for educational researchers searching for valid measures of academic achievement. It seems prudent to investigate the factors that contribute to grades any time that grades are used in educational research. The current study provides a clearer picture of the factors that contribute to grades for 3$^{rd}$ to 5$^{th}$ graders in Prince William County, VA, but might differ in school systems employing different grading rubrics.

Future investigations of the influence of problem behavior on grades might include additional student behavior measures, including effort, neatness, or participation, to determine if problem behavior remains a significant predictor of grades, controlling for other behavioral measures. In addition, further research could replicate the current study with data from a school system that uses a different type of rubric for assigning student grades. For example, would results differ in school systems that tell teachers to base grades solely on academic achievement factors,

excluding behavior altogether? Or do teachers grade similarly regardless of rubrics? The literature indicates that grading practices are inconsistent across systems, and it would be interesting to test my models in other school systems.

Although there are limitations to my teacher tolerance measure, the current study contributes to the literature on teacher tolerance because it attempts to study this construct in a new way, and can inform future studies of teacher tolerance. The practical significance of teacher tolerance, however, remains unclear. Results suggest that students may receive higher grades in classrooms with less tolerant teachers, and that there may be a differential effect of tolerance on students depending on their degree of behavior problems. Further research should investigate the beneficial and detrimental effects of teacher tolerance. It may be that teachers who lack tolerance are stricter, manage their classrooms more effectively, and that students experience more engaged time and perform better. Teachers who lack tolerance may have greater skills or motivation to manage extreme student problem behaviors, which could explain why 5$^{th}$ grade students with extreme behaviors receive slightly higher math and reading grades in classrooms with less tolerant teachers.

Additional research is necessary before the influence of teacher tolerance on grades can be clearly understood, as questions about the meaning and significance of this construct persist. A future study could combine the teacher self-report tolerance measures used in previous studies (Safran & Safran, 1984; Algozzine et al., 1982; Algozzine, 1979 ) with the aggregated residuals method employed here. In addition, future studies could include additional measures of student behavior in the calculation of teacher tolerance. For example, a future study might examine the discrepancy

between teacher ratings of student problem behavior, student self-reported problem behavior, and direct observations of student behavior. Additional behavior measures could provide a method for checking the construct validity of the teacher tolerance measure developed for the current study.

## Footnotes

[1] Teacher tolerance was also calculated from residuals from a second regression in which prior behavior ratings from 2005-06 and 2006-07 were entered as separate predictors. These residuals were aggregated to the classroom level and included in analyses. Since results were highly similar for both teacher tolerance variables, results are only reported for the first one.

[2] Initial plans for the current study involved using both grand-mean and group-mean centering for the problem behavior predictor. I had hypothesized that the relation between problem behavior and grades might differ according to centering method. However, preliminary analyses showed that the problem behavior coefficients were nearly identical when classroom and teacher variability were retained (with grand-mean centering) as when classroom variability was taken out (with group-mean centering). Since results were so similar, I decided to only report and interpret results for the group-mean centered problem behavior predictor.

Table 1

*Participant Characteristics*

| Teachers ($N = 562$) | Percentage | Students ($N = 12{,}993$) | Percentage |
|---|---|---|---|
| Gender | | Gender | |
| Female | 88 | Female | 50 |
| Male | 12 | Male | 50 |
| Ethnicity | | Ethnicity | |
| Caucasian | 85 | Caucasian | 43 |
| African-American | 9 | African-American | 21 |
| Hispanic | 2 | Hispanic | 24 |
| Asian | 1 | Asian | 07 |
| Other | 3 | Other | 03 |
| Grade Level | | Grade Level | |
| 3rd grade | 36 | 3rd grade | 34 |
| 4th grade | 33 | 4th grade | 33 |
| 5th grade | 31 | 5th grade | 33 |
| Class Size | | | |
| 12-16 students | 2 | FARM | 31 |
| 17-20 students | 19 | ESOL | 22 |
| 21-24 students | 43 | Special Education | 10 |
| 25-28 students | 34 | | |
| 29-33 students | 2 | | |

*Note*. FARM is Free and/or reduced meals. ESOL is English speakers of other languages.

Table 2

*Descriptive Statistics for Student- and Classroom-Level Variables*

| Variables | Mean/ proportion | *SD* | Min | Max | Reliability |
|---|---|---|---|---|---|
| Student-level variables | | | | | |
| (outcome) GPA | 0.00 | 1.00 | -4.32 | 1.29 | 0.91[a] |
| (outcome) Reading grade | 0.00 | 1.00 | -4.14 | 1.19 | - |
| (outcome) Math grade | 0.00 | 1.00 | -4.05 | 1.15 | - |
| Problem behavior | 0.00 | 1.00 | -0.65 | 5.39 | 0.90[a] |
| FARM | 0.00 | 1.00 | -0.67 | 1.48 | - |
| SOL composite | 0.00 | 1.00 | -4.60 | 1.83 | 0.87[a] |
| SOL reading | 0.00 | 1.00 | -3.72 | 1.65 | 0.88[b] |
| SOL math | 0.00 | 1.00 | -3.83 | 1.36 | 0.89[b] |
| Classroom-level variables | | | | | |
| Prior problem behavior | 0.00 | 1.00 | -1.79 | 5.71 | 0.43[c] |
| SOL & FARM composite | 0.00 | 1.00 | -3.12 | 3.33 | 0.89[c] |
| Teacher tolerance | 0.00 | 1.00 | -3.06 | 4.18 | 0.83[c] |

*Note*. [a] $\alpha$ (alpha) measuring internal consistency, [b] Kuder-Richardson Formula 20 (KR20), measuring internal consistency, [c] $\hat{\lambda}$ lamda hat, measuring average reliability for level-2 units. Descriptive statisctics include grades 3-5. Reliablities for SOL reading and math are estimates based on information provided in the 2004-05 SOL technical report (Virgina Standards of Learning, 2005). GPA is the grade point average of final math, reading, social studies, science, and writing grades. Reading grade is the final report card grade received in reading. Math grade is the final report card grade received in math. Problem behavior is the score a student received for the externalizing behavior scale on the TRSB survey. FARM is receiving free and/or reduced meals. SOL composite is the mean of scores received on all SOL subject tests. SOL reading is the standard score received on the SOL reading test. SOL math is the standard score received on the SOL math test.

Prior problem behavior is the classroom aggregate of student prior problem behavior. SOL & FARM composite is the classroom-level composite of mean classroom SOL achievement and proportion FARM. Teacher tolerance is the aggregate of student-level residuals from the regression predicting outcome year problem behavior ratings from the mean of prior years' ratings.

Table 3

*Correlations Between Student-Level and Classroom-Level Variables for Grades 3 through 5*

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| | Student-level variables (*N*=11,159) | | | | | | | |
| 1 FARM | - | | | | | | | |
| 2 Problem behavior | 0.08** | | | | | | | |
| 3 Reading SOL | -0.29** | -0.20** | | | | | | |
| 4 Math SOL | -0.28** | -0.22** | 0.60** | | | | | |
| 5 SOL composite | -0.35** | -0.23** | 0.83** | 0.83** | | | | |
| 6 (outcome) Math grades | -0.27** | -0.23** | 0.52** | 0.65** | 0.66** | | | |
| 7 (outcome) Reading grades | -0.30** | -0.26** | 0.56** | 0.55** | 0.63** | 0.71** | | |
| 8 (outcome) GPA | -0.30** | -0.27** | 0.60** | 0.63** | 0.71** | 0.82** | 0.84** | - |

| Variables | 1 | 2 | 3 |
|---|---|---|---|
| | Classroom-level variables (*N*=532) | | |
| 1 SOL & FARM composite | - | | |
| 2 Teacher Tolerance | -0.09* | | |
| 3 Prior Problem Behavior | 0.04 | -.09* | - |

*Note*. FARM is receiving free and/or reduced meals. Reading SOL is the standard score received on the SOL reading test. Math SOL is the standard score received on the SOL math test. SOL composite is the mean of scores received on all SOL subject tests. GPA is grade point average across reading, math, social studies, science, and writing. SOL & FARM composite is the classroom-level composite of mean classroom SOL achievement and proportion FARM.

* = p< .05 ** = *p* <.01

Table 4

*Variance Components and Proportion of Variance Explained for Student GPA, Reading, and Math Grades for Multilevel Models Testing Questions 1, 2, & 3*

| | 3rd Grade | | | 4th Grade | | | 5th Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| | GPA | Reading | Math | GPA | Reading | Math | GPA | Reading | Math |
| Unconditional Model | | | | | | | | | |
| Sigma squared ($\sigma^2$) | 0.88 | 0.88 | 0.83 | 0.77 | 0.78 | 0.83 | 0.83 | 0.84 | 0.87 |
| Tau ($\tau$) | 0.18 | 0.17 | 0.15 | 0.17 | 0.17 | 0.16 | 0.20 | 0.18 | 0.16 |
| Intraclass Correlation (ICC) | 0.17 | 0.16 | 0.15 | 0.18 | 0.18 | 0.16 | 0.19 | 0.18 | 0.16 |
| Models Testing Question 1 | | | | | | | | | |
| Final Model: | | | | | | | | | |
| Sigma squared ($\sigma^2$) | 0.29 | 0.44 | 0.38 | 0.25 | 0.43 | 0.36 | 0.27 | 0.49 | 0.37 |
| Tau ($\tau$) | 0.15 | 0.11 | 0.13 | 0.13 | 0.12 | 0.15 | 0.18 | 0.15 | 0.16 |
| Variance Explained: | | | | | | | | | |
| Proportion of $\sigma^2$ explained | 0.67 | 0.50 | 0.54 | 0.67 | 0.45 | 0.57 | 0.68 | 0.41 | 0.57 |
| Proportion of $\tau$ explained | 0.15 | 0.33 | 0.15 | 0.20 | 0.33 | 0.09 | 0.08 | 0.18 | 0.03 |
| Models Testing Question 2 | | | | | | | | | |
| Final Model: | | | | | | | | | |
| Sigma squared ($\sigma^2$) | 0.29 | 0.44 | 0.38 | 0.25 | 0.43 | 0.36 | 0.27 | 0.49 | 0.37 |
| Tau ($\tau$) | 0.15 | 0.11 | 0.13 | 0.14 | 0.12 | 0.15 | 0.19 | 0.15 | 0.16 |
| Variance Explained: | | | | | | | | | |
| Proportion of $\sigma^2$ explained | 0.67 | 0.50 | 0.54 | 0.67 | 0.45 | 0.57 | 0.68 | 0.41 | 0.57 |
| Proportion of $\tau$ explained | 0.15 | 0.34 | 0.16 | 0.19 | 0.33 | 0.09 | 0.06 | 0.16 | 0.03 |
| Models Testing Question 3 | | | | | | | | | |
| Final Model: | | | | | | | | | |
| Sigma squared ($\sigma^2$) | 0.29 | 0.44 | 0.38 | 0.25 | 0.43 | 0.36 | 0.27 | 0.49 | 0.37 |
| Tau ($\tau$) | 0.15 | 0.11 | 0.13 | 0.14 | 0.11 | 0.15 | 0.18 | 0.15 | 0.16 |
| Variance Explained: | | | | | | | | | |
| Proportion of $\sigma^2$ explained | 0.67 | 0.50 | 0.54 | 0.67 | 0.45 | 0.57 | 0.68 | 0.41 | 0.57 |
| Proportion of $\tau$ explained | 0.15 | 0.33 | 0.15 | 0.20 | 0.34 | 0.08 | 0.08 | 0.17 | 0.04 |

*Note.* $\sigma^2$ = within-classroom variance, $\tau$ = between-classroom variance. Proportion of $\sigma^2$ explained is computed using the following formula: ($\sigma^2$ unconditional - $\sigma^2$ final) / $\sigma^2$ unconditional. Proportion of $\tau$ explained is computed using the following formula: ($\tau$ unconditional - $\tau$ final) / $\tau$ unconditional. Intraclass correlation (ICC) is the proportion of total variance in the outcome that lies between classrooms. ICC = $\tau$ unconditional / ( $\tau$ unconditional + $\sigma^2$ unconditional).

Table 5

*Summary of Analyses Predicting 3rd Grade GPA, Reading, and Math Grades from Student Problem Behavior and Teacher Tolerance*

| | Model 1[a] | | | Model 2[a] | | | Model 3[b] | | | Model 4[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Testing Question 1 | | | Testing Question 2 | | | Testing Question 3 | | |
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| GPA (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.13** | **0.01** | *** | **-0.14** | **0.01** | *** | -0.14 | 0.01 | *** | -0.15 | 0.01 | *** |
| FARM status | -0.06 | 0.01 | *** | -0.06 | 0.01 | *** | -0.06 | 0.01 | *** | -0.06 | 0.01 | *** |
| SOL Composite | 0.77 | 0.02 | *** | 0.77 | 0.02 | *** | 0.77 | 0.02 | *** | 0.77 | 0.02 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | 0.02 | 0.02 | | 0.04 | 0.02 | | 0.02 | 0.02 | |
| SOL & FARM Composite | | | | -0.10 | 0.03 | ** | -0.10 | 0.03 | ** | -0.10 | 0.03 | ** |
| Teacher Tolerance | | | | | | | **0.05** | **0.03** | | -0.01 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.05** | **0.01** | *** |
| Reading Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.15** | **0.01** | *** | **-0.15** | **0.01** | *** | -0.15 | 0.02 | *** | -0.17 | 0.02 | *** |
| FARM status | -0.12 | 0.02 | *** | -0.11 | 0.02 | *** | -0.11 | 0.02 | *** | -0.11 | 0.02 | *** |
| SOL Reading | 0.57 | 0.01 | *** | 0.57 | 0.01 | *** | 0.57 | 0.01 | *** | 0.57 | 0.01 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.01 | 0.02 | | 0.00 | 0.02 | | -0.02 | 0.02 | |
| SOL & FARM Composite | | | | -0.12 | 0.03 | *** | -0.13 | 0.03 | *** | -0.13 | 0.03 | *** |
| Teacher Tolerance | | | | | | | **0.02** | **0.03** | | -0.04 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.03** | **0.01** | |
| Math Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.11** | **0.01** | *** | **-0.11** | **0.01** | *** | -0.11 | 0.01 | *** | -0.12 | 0.02 | *** |
| FARM status | -0.08 | 0.01 | *** | -0.08 | 0.01 | *** | -0.08 | 0.01 | *** | -0.07 | 0.01 | *** |
| SOL Math | 0.68 | 0.02 | *** | 0.68 | 0.02 | *** | 0.68 | 0.02 | *** | 0.68 | 0.02 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.01 | 0.02 | | 0.00 | 0.02 | | -0.01 | 0.02 | |
| SOL & FARM Composite | | | | -0.13 | 0.03 | *** | -0.13 | 0.03 | *** | -0.13 | 0.03 | *** |
| Teacher Tolerance | | | | | | | **0.04** | **0.03** | | 0.01 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.04** | **0.01** | ** |

*Note.* [a] Problem Behavior predictor is group-mean-centered, all other student and classroom-level predictors are grand-mean-centered. [b] All student and classroom-level predictors are grand-mean centered. Boldface entries indicate the coefficients providing tests of the research question examined in each model.

* < .05. ** < .01. *** < .001.

Table 6

*Summary of Analyses Predicting 4th Grade GPA, Reading, and Math Grades from Student Problem Behavior and Teacher Tolerance*

| | Model 1[a] | | | Model 2[a] Testing Question 1 | | | Model 3[b] Testing Question 2 | | | Model 4[a] Testing Question 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| GPA (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.18** | **0.01** | *** | **-0.18** | **0.01** | *** | -0.18 | 0.01 | *** | -0.19 | 0.01 | *** |
| FARM status | -0.05 | 0.01 | *** | -0.05 | 0.01 | *** | -0.05 | 0.01 | *** | -0.05 | 0.01 | *** |
| SOL Composite | 0.65 | 0.01 | *** | 0.65 | 0.01 | *** | 0.65 | 0.01 | *** | 0.65 | 0.01 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.01 | 0.02 | | 0.02 | 0.02 | | -0.01 | 0.02 | |
| SOL & FARM Composite | | | | -0.08 | 0.02 | ** | -0.08 | 0.02 | ** | -0.08 | 0.02 | ** |
| Teacher Tolerance | | | | | | | **0.06** | **0.03** | **0.06** | 0.00 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.02** | **0.01** | |
| Reading Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.18** | **0.02** | *** | **-0.19** | **0.02** | *** | -0.19 | 0.02 | *** | -0.20 | 0.02 | *** |
| FARM status | -0.09 | 0.01 | *** | -0.08 | 0.01 | *** | -0.08 | 0.01 | *** | -0.08 | 0.01 | *** |
| SOL Reading | 0.48 | 0.01 | *** | 0.49 | 0.01 | *** | 0.48 | 0.01 | *** | 0.48 | 0.01 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.03 | 0.03 | | -0.01 | 0.02 | | -0.03 | 0.02 | |
| SOL & FARM Composite | | | | -0.05 | 0.02 | * | -0.06 | 0.02 | * | -0.05 | 0.02 | * |
| Teacher Tolerance | | | | | | | **0.01** | **0.04** | | -0.06 | 0.04 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.02** | **0.02** | |
| Math Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.15** | **0.01** | *** | **-0.15** | **0.01** | *** | -0.15 | 0.01 | | -0.15 | 0.01 | |
| FARM status | -0.07 | 0.01 | *** | -0.07 | 0.01 | *** | -0.07 | 0.01 | | -0.07 | 0.01 | |
| SOL Math | 0.67 | 0.02 | *** | 0.68 | 0.02 | *** | 0.68 | 0.02 | | 0.68 | 0.02 | |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | 0.02 | 0.03 | | 0.04 | 0.03 | | 0.02 | 0.03 | |
| SOL & FARM Composite | | | | -0.06 | 0.03 | * | -0.06 | 0.03 | * | -0.06 | 0.03 | * |
| Teacher Tolerance | | | | | | | **0.08** | **0.03** | * | 0.02 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.01** | **0.01** | |

*Note.* [a] Problem Behavior predictor is group-mean-centered, all other student and classroom-level predictors are grand-mean-centered. [b] All student and classroom-level predictors are grand-mean centered. Boldface entries indicate the coefficients providing tests of the research question examined in each model.

* < .05. ** <.01. *** <.001.

Table 7

*Summary of Analyses Predicting 5th Grade GPA, Reading, and Math Grades from Student Problem Behavior and Teacher Tolerance*

| | Model 1[a] | | | Model 2[a] | | | Model 3[b] | | | Model 4[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Testing Question 1 | | | Testing Question 2 | | | Testing Question 3 | | |
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| GPA (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.16** | **0.01** | *** | **-0.16** | **0.01** | *** | -0.16 | 0.01 | *** | -0.18 | 0.01 | *** |
| FARM status | -0.04 | 0.01 | ** | -0.04 | 0.01 | ** | -0.04 | 0.01 | ** | -0.04 | 0.01 | ** |
| SOL Composite | 0.80 | 0.02 | *** | 0.80 | 0.02 | *** | 0.80 | 0.02 | *** | 0.80 | 0.02 | *** |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.03 | 0.02 | | -0.01 | 0.03 | | -0.04 | 0.02 | |
| SOL & FARM Composite | | | | -0.03 | 0.03 | | -0.04 | 0.03 | ** | -0.03 | 0.03 | |
| Teacher Tolerance | | | | | | | **0.06** | **0.03** | **0.07** | 0.01 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.05** | **0.01** | *** |
| Reading Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.25** | **0.02** | *** | **-0.25** | **0.02** | *** | -0.25 | 0.02 | *** | -0.27 | 0.02 | |
| FARM status | -0.09 | 0.02 | *** | -0.08 | 0.02 | *** | -0.08 | 0.02 | *** | -0.09 | 0.02 | |
| SOL Reading | 0.49 | 0.02 | *** | 0.49 | 0.02 | *** | 0.49 | 0.02 | *** | 0.49 | 0.02 | |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | -0.06 | 0.03 | ** | -0.03 | 0.03 | | -0.06 | 0.03 | * |
| SOL & FARM Composite | | | | -0.02 | 0.03 | | -0.03 | 0.03 | | -0.02 | 0.03 | |
| Teacher Tolerance | | | | | | | **0.07** | **0.03** | ** | -0.02 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.05** | **0.02** | ** |
| Math Grades (outcome) | | | | | | | | | | | | |
| Student-level | | | | | | | | | | | | |
| Problem Behavior | **-0.13** | **0.01** | *** | **-0.13** | **0.01** | *** | -0.13 | 0.01 | | -0.14 | 0.02 | |
| FARM status | -0.07 | 0.01 | *** | -0.07 | 0.01 | *** | -0.07 | 0.01 | | -0.07 | 0.01 | |
| SOL Math | 0.65 | 0.02 | *** | 0.65 | 0.02 | *** | 0.65 | 0.02 | | 0.65 | 0.02 | |
| Classroom-level | | | | | | | | | | | | |
| Prior Problem Behavior | | | | 0.01 | 0.03 | | 0.03 | 0.03 | | 0.01 | 0.03 | |
| SOL & FARM Composite | | | | -0.07 | 0.03 | ** | -0.08 | 0.03 | ** | -0.07 | 0.03 | * |
| Teacher Tolerance | | | | | | | **0.06** | **0.03** | * | 0.02 | 0.03 | |
| Cross-Level Interaction | | | | | | | | | | | | |
| Teacher Tolerance*Problem Behavior | | | | | | | | | | **0.04** | **0.01** | ** |

*Note.* [a] Problem Behavior predictor is group-mean-centered, all other student and classroom-level predictors are grand-mean-centered. [b] All student and classroom-level predictors are grand-mean centered. Boldface entries indicate the coefficients providing tests of the research question examined in each model.

* < .05. ** < .01. *** < .001.

Table 8

*Across-Grade Comparisons of Coefficients Predicting Grades Outcomes from Student Problem Behavior and Teacher Tolerance*

| | 3rd grade | | 4th grade | | 5th grade | |
|---|---|---|---|---|---|---|
| Variable | coef. | *SE* | coef. | *SE* | coef. | *SE* |
| Student problem behavior | | | | | | |
| GPA | -0.20 | 0.02 | -0.26 | 0.02 | -0.23 | 0.02 |
| Reading grades | -0.23 | 0.02 | -0.28 | 0.03 | -0.38 | 0.03 |
| Math Grades | -0.17 | 0.02 | -0.23 | 0.02 | -0.20 | 0.02 |
| Teacher Tolerance | | | | | | |
| GPA | 0.14 | 0.09 | 0.19 | 0.10 | 0.17 | 0.09 |
| Reading grades | 0.08 | 0.09 | 0.03 | 0.12 | 0.22 | 0.09 |
| Math Grades | 0.13 | 0.11 | 0.25 | 0.11 | 0.21 | 0.10 |

*Note.* coef. = unstandardized coefficient calculated using raw scores for outcomes (grades, 0 through 4) and predictors (problem behavior, 0 through 3; teacher tolerance, -.63 through .86). Unstandardized coefficients provide a method for comparing effects across grade levels (Richards, 1982). Student problem behavior is a student-level variable. Teacher tolerance is a classroom-level variable.

*Figure 1.* Relation between problem behavior ratings and reading grades as a function of teacher tolerance for fifth grade. High Tolerance = 15th percentile and Low Tolerance = 85th percentile. Remember that the tolerance variable is coded such that teachers with low scores are more tolerant, while those with high scores are less tolerant. Reading grades and problem behavior are shown in standardized score form (*M*=0, *SD*=1). The range for problem behavior goes from the 10th percentile to the 90th percentile.

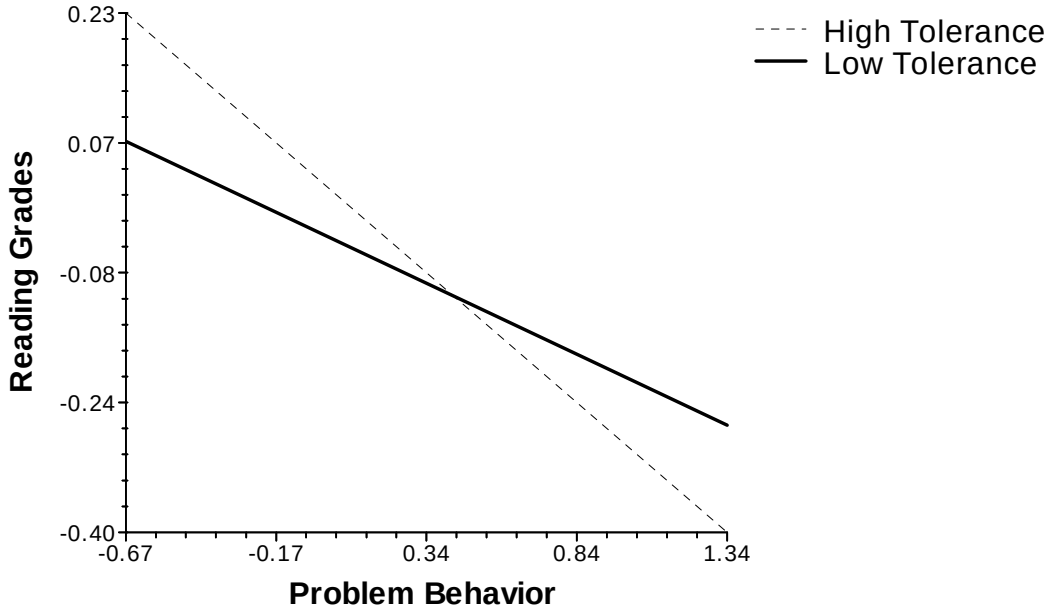
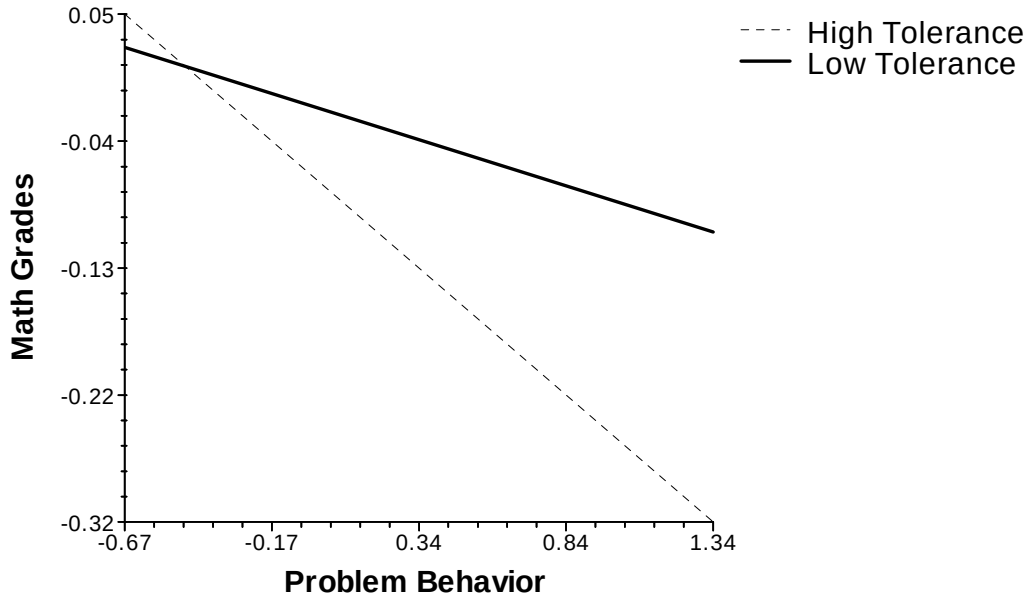


*Figure 2.* Relation between problem behavior ratings and math grades as a function of teacher tolerance, for fifth grade. High Tolerance = 15th percentile and Low Tolerance = 85th percentile. Remember that the tolerance variable is coded such that teachers with low scores are more tolerant, while those with high scores are less tolerant. Math grades and problem behavior are shown in standardized score form (*M*=0, *SD*=1). The range for problem behavior goes from the 10th percentile to the 90th percentile.



# Appendix A

## Prince William County Grading Rubric for Grades 3-5 Core Subjects: A Through C+

Regulation 661-1, INSTRUCTION, June 23, 2004, Pages 3-5

| GRADE | RANGE | ACHIEVEMENT IN SUBJECT | CLASS PERFORMANCE | INDEPENDENCE IN WORK |
|---|---|---|---|---|
| A Excellent | 93 - 100 % | - demonstrates outstanding achievement and mastery of the subject area<br>- evidences understanding and appreciation of the fundamental concepts of the subject area<br>- exercises superior ability in problem solving and in arriving at logical conclusions<br>- expresses ideas clearly both orally and in writing | - fully participates and demonstrates effort in all class activities<br>- exhibits originality in thinking, expression, and work products<br>- submits all work on or before due date<br>- displays neatness, legibility, and accuracy in work | - is self-directed<br>- shows originality in preparation of assignments<br>- consistently contributes independent work in addition to required assignments<br>- submits all work on or before due date |
| B+ Very Good | 90 – 92 % | - demonstrates very good achievement and mastery of the subject area<br>- evidences understanding and appreciation of the fundamental concepts of the subject area<br>- expresses ideas clearly both orally and in writing | - usually participates and demonstrates effort in class activities<br>- exhibits originality in thinking, expression, and work products<br>- submits all work on or before due date<br>- displays neatness, legibility, and accuracy in work | - completes assignments on time, thoroughly and accurately<br>- is self-directed<br>- sometimes contributes independent work in addition to required assignments |
| B Good | 84 - 89 % | - demonstrates above average achievement and mastery<br>- usually evidences understanding and appreciation of the fundamental concepts of the subject area | - usually participates and demonstrates effort in class activities<br>- usually submits work on or before due date<br>- displays neatness, legibility, and accuracy in work | - usually completes assignments on time, thoroughly and accurately<br>- is self-directed<br>- sometimes contributes independent work in addition to required assignments |
| C+ High Average | 81 – 83 % | - achieves sufficient subject mastery to proceed to the next level<br>- objectives are usually mastered, but not always | - sometimes participates and demonstrates effort in class activities<br>- inconsistently submits work on due date<br>- does not always display neatness, legibility, and accuracy in work | - usually completes assignments on time<br>- is sometimes self-directed, but sometimes needs encouragement to complete tasks |

# Appendix B

## Prince William County Grading Rubric for Grades 3-5 Core Subjects: C through F

Regulation 661-1, INSTRUCTION, June 23, 2004, Pages 3-5

| GRADE | RANGE | ACHIEVEMENT IN SUBJECT | CLASS PERFORMANCE | INDEPENDENCE IN WORK |
|---|---|---|---|---|
| C Average | 74 – 80 % | - achieves sufficient subject mastery to proceed to the next level<br>- objectives are sometimes mastered, but not always | - sometimes participates and demonstrates effort in class activities<br>- inconsistently submits work on due date<br>- does not always display neatness, legibility, and accuracy in work | - sometimes completes assignments on time<br>- is sometimes self-directed, but sometimes needs encouragement to complete tasks |
| D+ Below Average | 71 – 73 % | - frequently falls below the average level of achievement<br>- lacks sufficient subject mastery to proceed to the next level | - often does not participate and demonstrate effort in class activities<br>- submits poor work, but effort is in evidence | - frequently requires individual direction<br>- often does not complete assignments on time, or at all |
| D Poor | 65 – 70 % | - demonstrates limited achievement of grade level objectives<br>- consistently falls below grade level requirements | - may be irregular in attendance and generally fails to make up missed work<br>- shows little interest in class and rarely contributes | - seldom completes an undertaking without teacher direction and encouragement |
| F Failure | 64 % and below | - fails to meet minimum requirements | - frequently fails to complete assignments<br>- demonstrates little or no effort<br>- may have excessive unexcused absences<br>- fails to complete 65% of the assigned, evaluated work | - seldom completes an undertaking without teacher direction and encouragement |

## References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behavior: Its short- and long-term consequences for school performance. *Child Development, 64*, 801-814.

Algozzine, B. (2003a). *Technical Manual for the Disturbing Behavior Checklists*. Charlotte, NC: University of North Carolina at Charlotte, Department of Educational Leadership: Author.

Algozzine, B. (2003b). *Disturbing Behavior Checklist I*. Charlotte, NC: University of North Carolina at Charlotte, Department of Educational Leadership: Author.

Algozzine, B., & Curran, T. J. (1979). Teachers' predictions of children's school success as a function of their behavioral tolerance. *Journal of Educational Research*, *72*, 344-347.

Algozzine, B., Mercer, C., & Countermine, T. (1977). The effects of labels and behavior on teacher expectations. *Exceptional children*, *44*, 131-132.

Algozzine, B., Ysseldyke, J., & Christenson, S. (1982). The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student. *Alberta Journal of Educational Research, 29*, 89-97.

Beswick, J. F., Willms, J. D. & Sloat, E. A. (2005). A comparative study of teacher ratings of emergent literacy skills and student performance on a standardized measure. *Education, 126*, 116-137.

Brookhart, S. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30, 123-142.

Brookhart, S. (1994). Teacher's grading: Practice and theory. *Applied Measurement in Education, 7,* 279-301.

Bruckman, K., Vu., P., Vaganek, M., Rosenfield, S., Gottfredson, G., & Berger, J. (2009). The effects of Instructional Consultation Teams on student achievement and teacher ratings after two years of implementation. Unpublished manuscript, University of Maryland, College Park, Maryland.

Bub, K. L., McCartney, K., & Willet, J. B. (2007). Problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology, 99,* 653-670.

Bursuck, W., Polloway, E., Plante, L., Epstein, M., Jayanthi, M., & McConeghy, J. (1996). Report card grading and adaptations: A national survey of classroom practices. *Exceptional Children, 62,* 1-10.

Conners, C. K. (2001). *Conners' rating scales - revised: Technical manual.* Toronto: Multi-Health Systems, Inc.

Crosby, E. G. & French, J. L. (2002). Psychometric data for teacher judgments regarding the learning behaviors of primary grade children. *Psychology in the Schools,* 39, 235-244.

Cross, L. H., Frary, R. B. (1996, April). Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Education for the Handicapped Act of 1975. P.L. 94-142 (1975).

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.

Flynn, J. (1997). *Teacher Rating Scale.* LaCrosse, WI: LaCrosse Area Dyslexia Research Institute, Inc.

Friedman, S. J. & Frisbie, D. A. (1995) The influence of report cards on the validity of grades reported to parents. *Educational & Psychological Measurement*, 55, 5-26.

Gottfredson, G. D. (1981). Schooling and delinquency. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), *New directions in the rehabilitation of criminal offenders* (pp. 424-469). Washington, DC: National Academy Press.

Gottfredson, G. D. & Gottfredson, D. C. (1999). *Development and Applications of Theoretical Measures for Evaluatign Drug and Delinquency Prevention Programs: Technical Manual for Research Editions of What About You? (WAY).* Marriottsville, MD: Gottfredson Associates: Authors.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127-155.

Johnson, H. L. & Fullwood, H. L. (2006). Disturbing behaviors in the secondary classroom: How do general educators perceive problem behaviors? *Journal of Instructional Psychology, 33*, 20-39.

Johnson, W., McGue, M. & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology, 42*, 514-532.

Liljequist, L. & Renk, K. (2007). The relationship among teachers' perceptions of student behaviour, teachers' characteristics, and ratings of students' emotional and behavioral problems. *Educational Psychology, 27*, 557-571.

McMillan, J. H., Myran, S. & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95,* 203-213.

Psychological Corporation. (2002). *Wechsler individual achievement test: Second edition.* San Antonio, TX: Psychological Corporation.

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

Richards, J. (1982). Standardized versus unstandardized regression weights. *Applied Psychological Measurement, 6,* 201-212.

Safran, S. P. & Safran, J. S. (1984). Elementary teachers' tolerance of problem behaviors. *Elementary School Journal, 85,* 236-243.

Stone, B. J. (1994). Group ability test versus teachers ratings for predicting achievement. *Psychological Reports*, 75, 1487-1490.

Strein, W., & Meshbesher, N. (2006). Grades and grading. In G. Bear and K. Minke (Eds.). *Children's Needs: Psychological Perspectives III*. Bethesda, MD: National Association of School Psychologists.

Virginia Standards of Learning Assessments. (2005). Technical Report: 2003-2004 Administration. Retrieved from http://www.doe.virginia.gov/VDOE/Assessment/VA0304TechnicalReport.pdf

Werthamer-Larsson, L., Kellam, S.G., & Wheeler, L. (1991). Effect of first-grade classroom environment on child shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19*, 585-602.