EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal
Editor: Gene V Glass
College of Education
Arizona State University

Volume 12 Number 4 **January 28, 2004** **ISSN 1068-2341**

Group and Interaction Effects with "No Child Left Behind": Gender and Reading in a Poor, Appalachian District

Robert Bickel
Marshall University

A. Stan Maynard
Marshall University

Citation: Bickel, R., Maynard, A.S., (2004, January 28). Group and interaction effects with "No Child Left Behind": Gender and reading in a poor, Appalachian district, *Education Policy Analysis Archives, 12*(4). Retrieved [Date] from http://epaa.asu.edu/epaa/v12n4/.

Critics of "No Child Left Behind" judge that it oversimplifies the influence of social context and the place of socially ascribed traits, such as social class, race, and gender, in determining achievement. We hold that this is especially likely to be true with regard to gender-related group effects and gender-implicated interaction effects. We make our concerns concrete in a multilevel, repeated measures analysis of reading achievement in a poor, rural school district located in the southern coalfields of Appalachian West Virginia. Our results suggest that as the percentage of students who are male increases, school mean scores in reading achievement decline for three reasons: individual males do less well than females; the greater the percentage of males, the lower the scores for all students; added to that, the greater the percentage of males, the lower the scores

for males specifically. Given the accountability measures and sanctions proposed by "No Child Left Behind," having a large percentage of males in a school could be disastrous. We conclude that gender effects in reading achievement are complex, easily overlooked, and have no obvious remedy. As such, they lend credence to the view that "No Child Left Behind" oversimplifies the social context of schooling and underestimates the importance of social ascription.

"No Child Left Behind" is the first re-authorization of the Elementary and Secondary Education Act since 1994 (U.S. Department of Education, 2002a). An oft-noted consequence of the revised version of the Act is expansion of the role of the federal government in public education (Seldon, 2001; Rebora, 2002). The controversial nature of the Act is reflected in the Bush Administration's counter assertion that "No Child Left Behind" actually increases flexibility and control at the local level. In this view, what some take to be expansion of federal authority is better construed as redefinition (U.S. Department of Education, 2002b).

The primary purpose of the redefined federal role, as explained by the current Secretary of Education, is to employ federal education funds to close the achievement gap between disadvantaged and minority students and their peers, raising all students to a proficient level (U.S. Department of Education, 2003). Broadly, this is to be accomplished through more rigorous accountability measures, through enabling students to transfer from schools that do not meet prescribed performance levels, and by upgrading required qualifications for teachers and paraprofessional aides (White House, 2003).

Persistent failure to move students toward acceptable performance levels forces schools to invoke a variety of costly correctives. These include providing vouchers to facilitate transfer from poorly performing schools to public alternatives, complemented with supplemental services, including private tutoring (White House, 2001).

Expectations

"No Child Left Behind," is premised on the assumption that effective schools need not be constrained by contextual factors or by students' socially ascribed characteristics. The rationale for this rejection of conventional educational wisdom is often couched in terms of expectations: raise expectations for less-advantaged and minority students, and they will rise to the occasion (White House, 2001; U.S. Department of Education, 2003). Otherwise, students become victims of what the Secretary has termed "the soft bigotry of low expectations" (quoted in Huston, 2003).

School Context and Socially Ascribed Traits

"No Child Left Behind," thus, constitutes an emphatic dismissal of the inevitable intrusiveness of the social context of schooling. Much the same is true of students' socially ascribed traits. If context and social ascription interfere with student achievement, it is because schools are dysfunctional. Otherwise, these extraneous intrusions would be deflected by proper procedures, best practices, and effective school organization (cf. Bush, 2002).

Consistent with this view, the Act mandates that performance measures be disaggregated, reporting separately scores for specified categories of students. Categories include economic disadvantage, ethnicity, gender, English language proficiency, and disability. This permits group comparisons to determine if the achievement gap between members of less-advantaged and socially devalued groups and other students is being closed (White House, 2002).

Whatever its merits, “No Child Left Behind” seems disarmingly straightforward and modern. Scientifically validated methods of accomplishing education, coupled with high expectations for all, enables each student to shake off the constraints of class, race, gender, and other non-meritocratic factors. Deficiencies in curriculum, organization, or personnel that interfere with this process can and must be remedied.

To many professional educators, however, “No Child Left Behind” represents a dangerous oversimplification of the social circumstances of education (Coles, 2001; Bianchini, 2002; Denlinger, 2002; Huston, 2003; Bailey, 2003; Hardy, 2003). In this view, the effects of class, race, gender, and context cannot be explained and remedied with the ease the Act implies.

Group Effects and Interaction Effects

Complicating matters further, group effects and interaction effects are not reducible to readily identifiable individual characteristics or easy-to-see organizational factors (Aiken & West, 1991; Kreft & DeLeeuw, 1998; Raudenbush and Bryk, 2002). In the absence of well-developed theory, such effects are difficult to anticipate and often go undetected (Velicer, 1972; Baron & Kennedy, 1986; Jaccard, Turrisi, and Wan, 1990; Iversen, 1991; Snijders and Bosker, 1999). Nevertheless, group effects and interaction effects which bear on determining measured school performance are ubiquitous and consequential (Heck and Thomas, 2000).

For example, neighborhood effects at the group level suggest that students, in the aggregate, can imbue an entire school with a shared ethos which they jointly import from their out-of-school context (Vartanian and Gleason, 1999; Solon, Page, and Duncan, 2000; Bickel, Smith, and Eagle, 2002: Bickel and Howley, 2003). Depending on neighborhood quality, and net the influence of social class, neighborhood effects may enhance or diminish achievement. Intervening in neighborhoods, however, is beyond the scope of research-based practices and procedures, and raised expectations. As a result, the consequences of such powerful group effects are ignored by “No Child Left Behind.”

As another example, the frequently reported finding that, with class size held constant, the negative association between poverty and achievement is exacerbated as schools get larger represents an interaction effect which has no known remedy, other than to make schools smaller. As such, there is no good reason to believe that the reforms proposed by “No Child Left Behind” will diminish its pernicious consequences (Bickel and Howley, 2000; Bickel, Howley, Glascock, and Williams, 2001).

Research Objectives

In the following we use a small data set collected from all eight elementary schools in an impoverished, rural county in the coalfields of southern West Virginia. Our objective is to focus on one socially ascribed trait, gender, and to assess the plausibility of claims that such extraneous characteristics need not interfere with educational attainment. We do this by examining the group effects of gender and gender-implicated interaction effects in a multilevel, repeated measures analysis.

If we find gender-based group effects or gender-implicated interaction effects which have no available remedy, we will tentatively conclude that “No Child Left Behind” is premised on an unduly simplified view of the social circumstances of education. As a result, efforts to accomplish school reform through focusing on characteristics of individual students and readily manipulable organizational factors will yield, at best, limited success, because group effects and interaction effects will still be at work.

Why Gender?

Economic disadvantage and minority group status are more conspicuous in discussions of “No Child Left Behind” than gender category. Nevertheless, “No Child Left Behind” highlights gender by explicitly permitting use of federal funds for single-sex schools, something that administrators and policy makers had assumed to be inconsistent with Title IX of the Education Amendments of 1972 (Otterbourg, 2001). Proponents of “No Child Left Behind” cite funding for single-sex schools as one means of providing greater flexibility and control at the state and local levels (White House, 2002).

More to the point, while the effects on achievement of economic disadvantage and devalued minority group status are consistent and well known, gender effects are much more difficult to predict and explain (see, for example, Cloer and Dalton, 2001; Lynch, 2002; Phillips, Norris, Osmond, and Maynard, 2002). Sometimes they occur, and sometimes they do not. The same uncertainty applies to their direction, to the advantage of males or females (see, for example, High, 1996). Gender effects seem, therefore, less likely to be detected, especially if they take the form of group effects or interaction effects.

Lack of sensitivity to the importance of group-level effects of gender and gender-implicated interaction effects may lead us to misunderstand the real complexity of the social organization of school achievement. The consequences of gender effects for schools faced with the accountability demands and sanctions promulgated by “No Child Left Behind” may be disguised and damaging.

The County

The poor, rural county which was the source of our data is located in southern West Virginia, bordering on eastern Kentucky. Its population, 26,253, has declined by 24.3 percent since 1980 (U.S. Census Bureau, 2001). The county is 87.7 percent rural, in a state that is 63.9 percent rural; the same figure for the entire U.S. is 24.8 percent. The median family income is $21,347, well below the state median of $29,696 and little more than half the national median of $41,994. Of families with children, 21.4 percent had incomes below the federal poverty level in a state where 17.9 percent of all families with children were below that income level; the same figure for the entire U.S. is 12.4 percent. Among elementary school students in the county, 74.9 percent are eligible for free/reduced cost lunch (U.S. Census Bureau, 2001).

Data

"No Child Left Behind" gives priority to literacy, reflecting the educational priorities of President Bush (International Reading Association, 2003). It posits the existence of research-based, scientifically validated practices and procedures to promote reading achievement, and provides competitive Reading First grants to assist states in implementing reading improvement programs for children in the early elementary grades.

Given the conspicuous role of reading in "No Child Left Behind," it is useful that our multilevel repeated measures analyses are based on successive administrations of the widely used Woodcock-Johnson 22 letter-word identification test and Woodcock-Johnson 23 passage comprehension test as standardized measures of reading achievement (Woodcock and Johnson, 1990). All variables are described in Table 1, and descriptive statistics by gender are reported in Table 2.

Data were originally collected for use in a local, unpublished evaluation of a program designed to provide training for parents and other volunteers to tutor low-achieving students in the lower elementary grades in this poor, rural, Appalachian county. Tutors were paired with students identified by teachers as in danger of being retained because of reading deficiencies.

One hundred-five students from the county's eight elementary schools were referred and tutored. Achievement tests were administered to forty-four first grade students and sixty-one second grade students at the beginning and end of the 1996-97 school year. The number of test takers was constant from the first test administration to the second.

TABLE 1
VARIABLES

W-J 22	Woodcock-Johnson 22: Letter-Word Identification Reading Achievement Test; Internal Consistency Reliability = .92.
W-J 23	Woodcock-Johnson 23: Passage Comprehension Reading Achievement Test; Internal Consistency

	Reliability = .90.
TIME1	Test Administered Twice: Beginning of Grade 1 or 2 and End of Grade 1 or 2 Level 1, Within Subjects; Coded 0 and 1.
GENDER2	Gender Level 2, Between Subjects; Coded 1 (Male) or 0 (Female).
GENDER3	Gender (Aggregated) Level 3, Between Schools.
GRADE2	First or Second Grade, Level 2, Between Subjects.
AGE2	Age in Years Level 2, Between Subjects.
AGE3	Age in Years (Aggregated) Level 3, Between Schools.
SCHLSIZE3	Total School Enrollment Level 3, Between Schools.
CLASSIZE3	Mean Class Size Level 3, Between Schools.
LUNCH3	Percent Eligible for Free/Reduced Cost Lunch, Between Schools.
SPAN3	Grade-Span Configuration, Between Schools.

TABLE 2
DESCRIPTIVE STATISTICS: MALES

	Means	Standard Deviations	Minimum	Maximum
W-J 22	21.76	6.15	8.00	35.00
W-J 23	8.79	4.98	0.00	19.00
TIME1	0.50	0.50	1.00	1.00
GENDER2	1.00	0.00		1.00
GENDER3	0.65	0.17	0.36	0.90
GRADE2	1.60	0.49	1.00	2.00
AGE2	7.48	0.81	6.17	9.00
AGE3	7.52	0.25	7.10	7.98
SCHLSIZE3	296.87	84.09	151.00	381.00
CLASSIZE3	21.54	1.92	18.90	24.50
LUNCH3	74.95	10.98	55.00	95.00
SPAN3	5.19	0.86	5.00	9.00

N = 63

DESCRIPTIVE STATISTICS: FEMALES

	Means	Standard Deviations	Minimum	Maximum
W-J 22	22.45	5.13	14.00	35.00
W-J 23	9.56	4.17	0.00	17.00
TIME1	0.50	0.50	1.00	1.00
GENDER2	0.00	0.00	0.00	0.00
GENDER3	0.60	0.20	1.10	1.67
GRADE2	1.55	0.50	1.00	2.00
AGE2	7.46	0.85	6.17	9.25
AGE3	7.42	0.22	7.10	7.98
SCHLSIZE3	307.24	89.27	151.00	381.00
CLASSIZE3	21.56	1.79	18.90	24.50
LUNCH3	75.19	5.67	55.00	95.00
SPAN3	5.29	1.04	5.00	9.00

N = 42

Data Analysis

Our analysis was done with SPSS 11.0 Mixed Models, using variables measured at three levels: within subjects for repeated measures, between subjects, and between schools (SPSS, 2001). The eight schools in which the one hundred five respondents were located ranged in size from one hundred fifty-one to three hundred eighty-one students. The number of test-takers per school varied from twelve to thirty-eight. This represents approximately twenty percent of the students in first and second grades in each school for 1996-97.

In addition to representing eight schools, the students in our secondary analysis were distributed among an undocumented number of classrooms. Since students were not identified by classroom, this cannot be used as another level in our multilevel analysis.

Reading Achievement Growth as a Linear Process

With only two test administrations, we represent reading achievement growth as a linear process (Raudenbush and Bryk, 2002: 163-169). Moreover, with a small number of observations at the second and third levels, we have sought to be parsimoniously selective in specifying our model (Kreft and De Leeuw, 1998: 58-60). Independent variables are limited to time, to represent movement from the beginning to the end of the school year in our repeated measures analysis; gender at levels two and three, reflecting our interest in reading achievement as a function of gender differences among poor, rural elementary school students; age at levels two and three; grade level at level two; mean classroom size at level three; school size at level three; percent of students eligible for free/reduced cost lunch at level three; and grade-span configuration at level three.

Independent Variables Defined

Time (TIME1) is a first-level, within-subjects measure which corresponds to the two dates of test administration. TIME1 has a random coefficient. This means that the relationship between TIME1 and the repeated measures dependent variable has been permitted to vary from student to student, with the regression coefficient corresponding to TIME1 treated as function of cross-level interactions of TIME1 with second-level and third-level independent variables.

Second-level variables include gender (GENDER2), grade level (GRADE2), and age (AGE2). All second-level variables have fixed coefficients, except GENDER2. The random coefficient corresponding to GENDER2 is permitted to vary from school to school, and is treated as a function of cross-level interactions with third-level variables.

Random coefficients are used with TIME1 and GENDER2 because of the importance of these variables in our analysis: we are working with a growth model, and our primary substantive interest is in the relationship between gender and achievement.

Random coefficients might have been used with other level two independent variables, acknowledging that their regression coefficients may vary from school to school. In addition, use of a random intercept is commonplace, reflecting differences among mean achievement level from school to school. However, use of random coefficients and a random intercept is a case-intensive process, and we are constrained by the small number of students and schools in our secondary analysis. In addition, the primary purpose of second level and third level variables which do not measure gender effects is to serve as controls. We are less concerned with accurately gauging the numerical magnitude and statistical significance of regression coefficients for control variables than for variables gauging gender effects.

Third-level, between-school, variables used in our analysis are gender composition (GENDER3), school size as measured by total enrollment (SCHLSIZE3), mean class size (CLASSIZE3), percent eligible for free or reduced cost lunch (LUNCH3), and grade span configuration (SPAN3). Each of these explanatory factors has a fixed coefficient.

The Absence of Ethnicity

Certainly, ethnicity or race, with their predictably non-meritocratic consequences, could rightly be construed as variables which demand inclusion in any discussion of the relationship between socially ascribed traits and achievement. However, this poor, aging, rural Appalachian county, is 96.4 percent white, and none of the students in our sample was reported to be non-white.

The Absence of Individual Students' Social Class

Information which would enable us to estimate each student's social class or socioeconomic status was not included in the data set used in our secondary analysis. Among our eight elementary schools, however, the percentage of students eligible for free or reduced cost lunch ranges from fifty-five percent to ninety five-percent, with a median of seventy-four percent. This information, in the form of the level three variable LUNCH3, is used as a between-schools explanatory factor.

The Absence of Grade-Level Composition

Our analysis includes a variable which assigns a grade level, first or second, to each student. This is an essential control. However, efforts to aggregate this information to the school level and incorporate it as a level three explanatory factor produced serious multicollinearity problems. When the aggregated grade-level variable is deleted, however, all Variance Inflation Factors and the Condition Index are well within normal limits.

Cross-Level Interactions

Cross-level interaction terms are a staple of multilevel modeling. They are essential in defining the mathematical character of multilevel models (Snijders and Bosker, 1999: 72-83; Angeles and Mroz, 2001), accounting for variability in random regression coefficients (Kreft and De Leeuw, 1998: 72-105), and are of substantive value, as well.

However, as product terms, cross-level interactions proliferate rapidly as the number of independent variables increases. Therefore, cross-level interactions must be selected judiciously (Snijders and Bosker, 1999: 77; Heck and Thomas, 200: 188-89). Because of the substantive importance of gender in our analysis, we have limited our cross-level interaction terms to those which can be created with GENDER2 or GENDER3 and another independent variable.

Use of grand mean and group mean centering helps to avoid intractable multicollinearity problems by rendering multiplicative interaction terms orthogonal to the variables from which they were created. In the present instance, when we use all of the selected independent variables and interaction terms in an ordinary least squares multiple regression equation, collinearity diagnostics yield fourteen variance inflation factors less than 2.00, with the remaining three ranging from 2.16 to 3.00. The value of the condition index is 3.38. All measures are well within acceptable limits (Chatterjee, Hadi, & Price, 2000: 238-241; Kmenta, 1997: 438-439).

Woodcock-Johnson 22 Results: Within Subjects

With the Woodcock-Johnson 22 letter-word identification reading achievement test as our outcome measure, we see in Table 3 that TIME1, the first-level (between-subjects) independent variable with a random coefficient, is statistically significant and positive. Since we are estimating a growth model, this comes as no surprise. Since TIME1 has two levels, coded 0 and 1, the regression coefficient tells us that the passage of time from the first test administration to the second results in an increase in measured math achievement equal, on the average, to 4.08 points. Since the repeated measures dependent variable has a mean of 22.05 and a standard deviation of 5.77 for the entire sample, this is substantial growth, equal to 0.71 standard deviation units in just one school year.

TABLE 3
MAIN EFFECTS: WOODCOCK-JOHNSON 22

LEVEL 1: WITHIN STUDENTS			
PARAMETER	**ESTIMATE**	**t VALUE**	**SIG.**
TIME1	4.08	12.22	.000
LEVEL 2: BETWEEN STUDENTS			
PARAMETER	**ESTIMATE**	**t VALUE**	**SIG.**
GENDER2	-1.27	-2.40	.025
GRADE2	9.12	13.14	.000
AGE2	-1.67	-4.05	.000
LEVEL 3: BETWEEN SCHOOLS			
PARAMETER	**ESTIMATE**	**t VALUE**	**SIG.**
GENDER3	-1.06	-0.67	.509
AGE3	0.69	0.71	.484
SCHLSIZE3	0.02	0.70	.486
CLASSIZE3	0.15	1.03	.313
LUNCH3	-0.24	-9.18	.000
SPAN3	-0.31	-1.40	.172
LEVEL 1 INTERCEPT TERM			
PARAMETER	**ESTIMATE**	**t VALUE**	**SIG.**
INTERCEPT	22.03	110.35	.000

Woodcock-Johnson 22 Results: Between Subjects

Three of the second-level, between-individuals independent variables, have statistically significant regression coefficients: GENDER2, with a random coefficient, and GRADE2 and AGE2, with fixed coefficients. The regression coefficient corresponding to gender tells us that male students, on average, score 1.27 points below female students. This disadvantage for males holds with a reasonable complement of controls in place, including the level two variables GRADE2 and AGE2. As one would expect, our results for GRADE2 tell us that second graders, on average, do better than first graders, with the statistically significant regression coefficient showing a 9.12 test score advantage for students in the higher grade. Furthermore, when controlling for GRADE2 and a variety of less closely related factors, our results show that older students, on average, score 1.67 points per year lower than younger students. This reflects the fact that students' age is positively correlated with retention, and those who are retained tend to do less well on standardized tests than those who do not repeat one or more grades (Thompson & Cunningham, 2000).

Woodcock-Johnson 22 Results: Between Schools

At the third level, between schools, there is one aggregated variable, LUNCH3, with a statistically significant regression coefficient. In this instance, we see that for each one percent increase in our free/reduced cost lunch variable, the Woodcock-Johnson 22 score decreases, on average, by 0.24 points. Since our social class proxy, LUNCH3, can be construed as a school-level measure of the incidence of poverty, this statistically significant and negative relationship is not surprising.

Woodcock-Johnson 22 Results: Cross-Level Interactions

In Table 4, we see that one cross-level interaction term, GENDER2byGENDER3, has a statistically significant regression coefficient. This means that, in addition to the positive main effect relationship due to gender differences at the between-subjects level, it is also the case that males do less well than females as the percentage male in a school increases.

TABLE 4
CROSS-LEVEL INTERACTIONS: WOODCOCK-JOHNSON 22

PARAMETER	ESTIMATE	t VALUE	SIG.
TIME1byGENDER2	-0.62	-0.86	.406
TIME1byGENDER3	-1.62	-0.79	.442
GENDER2byGENDER3	**-8.45**	**-2.16**	**.043**
GENDER2bySCHLSIZE3	-0.05	-0.79	.437
GENDER2byCLASSSIZE3	0.25	1.40	.447
GENDER2byLUNCH3	0.01	0.12	.910
GENDER2bySPAN3	0.15	0.31	.759

Woodcock-Johnson 22 Results: The Influence of Gender in the Complete Model

By way of summarizing our results for the Woodcock-Johnson 22, Table 5 reports values of the -2 log likelihood summary statistic for the empty model and the complete model. With a smaller-is-better summary statistic, when explanatory factors are introduced, the numerical value of the -2 log likelihood measure decreases, and the decrement is statistically significant, meaning an improved model fit (see Snijders & Bosker, 1999: 82-83). For the full model, we also report the R^2_L summary measure. R^2_L is the proportional reduction in the -2 log likelihood statistic due to the independent variables (Menard, 2002: 24), here equal to 14.6 percent.

TABLE 5
Empty Model
Variance Components Error Structure

-2 Log Likelihood	1324.9

Complete Model
Variance Components Error Structure

-2 Log Likelihood	1132.1

R^2_L = 14.6%

Of primary importance with regard to the influence of gender, however, are the results already reported in Tables 3 and 4: gender has a between-individuals main effect and a level-two-by-level-three interaction effect, the product of gender composition at the school level and gender at the individual level. In both instances, with the Woodcock-Johnson 22 letter-word identification test as the outcome measure, gender works to the disadvantage of males.

It is useful to emphasize, moreover, that males' disadvantage is, in part, due to the gender composition of the school they attend. As the percentage of males increases, the male disadvantage is made worse.

Woodcock-Johnson 22 Results: Random Coefficient Parameters

When a simplified analysis is run using TIME1 and GENDER2 with random coefficients as the only independent variables, the variance of the regression coefficient corresponding to GENDER2 is statistically significant. However, in Table 6 we see that when all specified third-level variables and cross-level interactions are included, the variance of the GENDER2 regression coefficient is no longer statistically significant. This means that variability in the random coefficient for GENDER2 has been accounted for by cross-level interaction effects.

TABLE 6
COVARIANCE PARAMETERS: RANDOM EFFECTS

PARAMETER	ESTIMATE	WALD Z	SIG.
TIME1	0.00	0.00	1.000

GENDER2	4.35	1.85	.065

Intraclass Correlation, Levels1&2 = .616   Intraclass Correlation, Levels 2&3 = .162

COVARIANCE PARAMETERS: REPEATED MEASURES

PARAMETER	**ESTIMATE**	**WALD Z**	**SIG.**
BEGIN SCHOOL YEAR	**6.21**	**6.15**	**.000**
END SCHOOL YEAR	**5.46**	**5.26**	**.000**

First-Level Error Covariance Structure

With repeated measures analysis, the Mixed Models procedure for SPSS 11.0 provides a range of choices for the repeated measure error structure, including scaled identity, compound symmetry, first-order autocorrelation, variance components, and unstructured (SPSS, 2001). The variances of the two scores which make up the linear growth measure are substantially different, 6.21 and 5.46, which is consistent with using variance components in modeling our error covariance structure (Schineller, 1997; Bickel and Howley, 2003). Furthermore, running the analysis with the alternatives yields a smaller-is-better -2 log likelihood statistic larger than that obtained with variance components (see Angeles and Mroz, 2001). Table 6 shows us, moreover, that both of the repeated measures covariance parameter estimates are statistically significant.

Woodcock-Johnson 23 Results: Within Subjects

In Table 7 we see that, much as with our Woodcock-Johnson 22 results, TIME1, the first-level (between-subjects) independent variable with a random coefficient, is statistically significant and positive when using the Woodcock-Johnson 23 passage comprehension reading achievement test as the dependent variable. The regression coefficient corresponding to the TIME1 within-subjects variable tells us that, from the first test administration to the second, the test score has increased, on average, by 3.46 points. With a repeated measures dependent variable which has a mean of 9.10 and a standard deviation of 4.68, this is a substantial increase, equal to 0.74 standard deviation units, and comparable to our findings with the Woodcock-Johnson 22.

TABLE 7
MAIN EFFECTS: WOODCOCK-JOHNSON 23

LEVEL 1: WITHIN STUDENTS

PARAMETER	**ESTIMATE**	**t VALUE**	**SIG.**
TIME1	**3.46**	**11.49**	**.000**

LEVEL 2: BETWEEN STUDENTS

PARAMETER	ESTIMATE	t VALUE	SIG.
GENDER2	**-1.15**	**-2.26**	**.043**
GRADE2	**7.50**	**11.65**	**.000**
AGE2	**-1.00**	**-3.17**	**.007**

LEVEL 3: BETWEEN SCHOOLS

PARAMETER	ESTIMATE	t VALUE	SIG.
GENDER3	**-5.35**	**-3.63**	**.002**
AGE3	**3.42**	**3.13**	**.007**
SCHLSIZE3	-0.05	-1.96	.070
CLASSIZE3	0.24	1.79	.095
LUNCH3	**-0.13**	**-5.53**	**.000**
SPAN3	0.11	0.53	.604

LEVEL 1 INTERCEPT TERM

PARAMETER	ESTIMATE	t VALUE	SIG.
INTERCEPT	**9.21**	**49.13**	**.000**

Woodcock-Johnson 23 Results: Between Subjects

Three of the second-level, between-individuals independent variables, have statistically significant regression coefficients. As with the Woodcock-Johnson 22 results, these are GENDER2, with a random coefficient, and GRADE2 and AGE2, with fixed coefficients. The coefficient corresponding to gender tells us that male students, on average, score 1.15 points lower than female students. This disadvantage for males holds with a reasonable complement of controls in place, including the level two variables GRADE2 and AGE2. As before, our results for GRADE2 tell us that second graders, on average, do better than first graders, with the statistically significant regression coefficient showing a 7.50 test score point advantage for students in the higher grade. Furthermore, when controlling for GRADE2 and a variety of less closely related factors, our results show that older students, on average, score 1.00 point per year lower than younger students. Again, age is correlated with retention, with older students more likely to be the retained, and students who are retained tending to do less well on standardized achievement tests (Thompson and Cunningham, 2000).

Woodcock-Johnson 23 Results: Between Schools

At the third level, between schools, Table 7 shows us that GENDER3, AGE3, and LUNCH3 have statistically significant regression coefficients with the Woodcock-Johnson 23 score as the dependent variable. In this instance, we see that for each one percent increase in the percentage of students who are male, the Woodcock-Johnson 23 score decreases, on average, by 5.35 points. Furthermore, for each one year increase in the average age at the school level, average test score increases by 3.42 points. Finally, for each one percent increase in the free/reduced cost lunch variable, the Woodcock-Johnson 23 score decreases, on the average, by 0.13 points.

Woodcock-Johnson 23 Results: Cross-Level Interactions

In Table 8, we see that one cross-level interaction term, GENDER2byGENDER3 has a statistically significant regression coefficient. This means that, in addition to the negative main effect relationships due to gender differences at the between-subjects and between-schools levels, it is also the case that males do less well than females as the percentage male in a school increases.

TABLE 8
CROSS-LEVEL INTERACTIONS: WOODCOCK-JOHNSON 23

PARAMETER	ESTIMATE	t VALUE	SIG.
TIME1byGENDER2	-0.92	-1.41	.195
TIME1byGENDER3	-3.34	-1.81	.108
GENDER2byGENDER3	**-8.07**	**-2.11**	**.040**
GENDER2bySCHLSIZE3	0.02	0.32	.751
GENDER2byCLASSIZE3	-0.21	-0.67	.515
GENDER2byLUNCH3	0.03	0.41	.691
GENDER2bySPAN3	-0.31	-0.69	.498

Woodcock-Johnson 23 Results: The Influence of Gender in the Complete Model

By way of summarizing our results for the Woodcock-Johnson 23, Table 9 reports values of the -2 log likelihood summary statistic for the empty model and the complete model. Again, with the smaller-is-better summary statistic, when explanatory factors are introduced, the numerical value of the -2 log likelihood measure decreases, and the model-to-model decrement is statistically significant.

TABLE 9
Empty Model
Variance Components Error Structure

-2 Log Likelihood	1238.2

Complete Model

Variance Components Error Structure

-2 Log Likelihood	1051.5

$$R^2_L = 15.1\%$$

Since gender effects are our primary concern, however, the findings already reported in Tables 7 and 8 are of special interest: gender has both between-individuals and between-schools main effects, as well as a level-two-by-level-three interaction effect. In all three instances, with the Woodcock-Johnson 23 passage comprehension test as the outcome measure, gender works to the disadvantage of males.

As with the Woodcock-Johnson 22, it is useful to emphasize that gender effects on the Woodcock-Johnson 23 are not limited to the individual level. Instead, as the percentage of students who are male increases, the scores of all students are, on average, diminished, and the scores of male students specifically are diminished still more.

Woodcock-Johnson 23 Results: Random Coefficient Parameters

When the analysis is run using just TIME1 and GENDER2 as independent variables with random coefficients, the variance of neither coefficient is statistically significant. The same is true for results based on the full model, reported in Table 10. This means that the coefficients corresponding to these two explanatory factors do not vary from one higher level unit to another.

TABLE 10
COVARIANCE PARAMETERS: RANDOM EFFECTS

PARAMETER	ESTIMATE	WALD Z	SIG.
TIME1	0.00	0.00	1.000
GENDER2	4.88	1.64	.101

Intraclass Correlation, Levels1&2 = .471   Intraclass Correlation, Levels 2&3 = .311

COVARIANCE PARAMETERS: REPEATED MEASURES

PARAMETER	ESTIMATE	WALD Z	SIG.
BEGIN SCHOOL YEAR	**5.68**	**4.65**	**.000**
END SCHOOL YEAR	**3.77**	**3.04**	**.002**

First-Level Error Covariance Structure

As with the Woodcock-Johnson 22, when using repeated measures analysis with Woodcock-Johnson 23, the variances of the two scores which make up the linear growth measure differ substantially, having values of 5.74 and 3.68. Again, this is consistent with using variance components error structure. As before, variance components error structure yielded the smallest value for the smaller-is-better -2 log likelihood summary statistic, and Table 11 shows us that both repeated measures covariance parameter estimates are statistically significant.

Discussion

With unusual consistency across two widely used measures of reading achievement, we have found that first and second grade males in a poor, rural, Appalachian school district do less well than females. For the both Woodcock-Johnson 22 and 23, individual male students, on average, do less well than female students. In addition, for the Woodcock-Johnson 23, as the percentage of students in a school who are male increases, the scores of all students tend to decline. Furthermore, for the Woodcock-Johnson 22 and 23, as the percentage of a school's students who are male increases, the scores of male students specifically are further diminished.

Of special importance for our research objectives are the group effect of gender with the Woodcock-Johnson 23, and the interaction effects involving gender with both the Woodcock-Johnson 22 and the Woodcock-Johnson 23. Both sets of effects make clear that the role of the socially ascribed trait gender in determining reading achievement is not limited to the individual level. As such, gender effects take forms that may be difficult to anticipate. How one remedies group-level gender group effects and gender-implicated interaction effects, moreover, is not clear. It does seem clear, however, that "No Child Left Behind" presumes a social world wherein schooling is less complex, and easier to understand and reform, than is actually the case.

Imagine, for example, a distribution of schools which vary with regard to gender composition. Our results suggest that as the percentage of students who are male increases, school mean scores in reading achievement may decline for three reasons: individual males do less well than females; the greater the percentage of males, the lower the scores for all students; and the greater the percentage of males, the lower the scores for males specifically. Given the accountability measures and sanctions proposed by "No Child Left Behind," having a large percentage of males in a school could be disastrous.

Conclusion

At the outset, we noted that the disarmingly straightforward and science-focused character of "No Child Left Behind" is judged by many professional educators to be misleading. In their view, the effects of class, race, gender, and context cannot be explained and remedied with the ease the Act implies. We added that the involvement of social ascription in group effects and interaction effects could further complicate matters with regard to both substance and method. We have now demonstrated that gender effects for elementary reading can be complex, indeed, taking the form of individual effects, group effects, and interaction effects. This makes it likely that the socially ascribed trait gender will intrude in unanticipated and undetected ways in determining the achievement objectives and accountability measures mandated by “No Child Left Behind.” Our findings lend credence to the view that “No Child Left Behind” oversimplifies the social context of schooling and underestimates the importance of socially ascribed traits.

References

Aiken, L. and West, G. (1991) Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: Sage.

Angeles, A. and Mroz, T. (2001) A Guide to Using Multilevel Models for the Evaluation of Program Impact. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Baron, R. and Kenny, D. (1986) The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. Journal of Personality and Social Psychology. 51: 1173-1182.

Bianchini, L (2002) NCTE Resolution on the Reading First Initiative. News from NCTE.org/news/2002/resolution.shtml.

Bickel, R. and Howley, C. (2000) The Influence of Scale on School Performance: A Multilevel Extension of the Matthew Principle. Education Policy Analysis Archives. 8. epaa.asu.edu/epaa/v8n22.html

Bickel, R. and Howley, C. (2003) Elementary Math Achievement and Rural Development: Effects of Contextual Factors Intrinsic to the Modern World. Athens, OH: ACCLAIM Working Paper No. 15, Appalachian Collaborative for Learning, Assessment, and Instruction in Mathematics.

Bickel, H., Howley, C., Glascock, C., and Williams, T. (2001) High School Size, Achievement Equity, and Cost: Robust Interaction Effects and Tentative Results. Education Policy Analysis Archives. 9. http://epaa.asu.edu/epaa/v9n40.html

Bickel, R., Smith, C., and Eagle, T. (2002) Poor, Rural Neighborhoods and Early School Achievement. Journal of Poverty. 6: 89-108.

Bracey, G. (2003) NCLB – A Plan for the Destruction of Public Education. NoChildLeft.Com. Februray. http://NoChildLeft.com/2003/feb03no.html

Bush, G. (2002) President Launches Quality Teacher Initiative. Washington, D.C.: Executive Office of the President.

Chatterjee, S., Hadi, A., and Price, B. (2000) Regression Analysis by Example. New York: Wiley.

Cloer, T. and Dalton, S. (2001) Gender and Grade Differences in Reading Achievement and in Self-Concept as Readers. Journal of Reading Achievement. 26: 31-36.

Coles, G. (2001) Learning to Read “Scientifically.” Rethinking Schools Online. Summer. http://www.rethinkingschools.org/archives/15_04/read154.htm

Denlinger, S. (2002) Teaching as a Profession: A Look at the Problem of Teacher Deficits. Clearing House. 75: 116-117.

Hardy, L. (2003) Overburdened and Overwhelmed. ASBJ.Com. April. http://www.sdbj.com/current/coverstory.html

Heck, R. and Thomas, L. (2000) An Introduction to Multilevel Modeling Techniques. Mahwah, NJ: Lawrence Earlbaum.

High, C. (1996) The Texas Study: A Regression Analysis of Selected Factors that Influence the Scores of Students on the TASP Test. Houston: Texas Association of College Testing

Personnel.

Huston, P. (2003) The Bigotry of Expectations. School Administrator. January. ht tp://www.aasa.org.publications/sa/2003 01/execeper.htm

International Reading Association (2003) IRA Survey Examines Process for Reading First Applications. Reading Today. February/March. http://www/reading.org/publications/rty/archives/o3feb_survey.html

Iversen, G. (1991) Contextual Analysis. Newbury Park, CA: Sage.

Jaccard, J., Turrisi, T., and Choi, K. (1990) Interaction Effects in Multiple Regression. Newbury Park, CA: Sage.

Kmenta, J. (1997) Elements of Econometrics. Ann Arbor, MI: University of Michigan.

Kreft, I. and De Leeuw (1998) Introducing Multilevel Modeling. Thousand Oaks, CA: Sage.

Lynch, J. (2002) Parents' Self-Efficacy Beliefs, Parents' Gender, Children's Reader Self-Perceptions, Reading Achievement, and Gender. Journal of Research in Reading. 25: 54-67.

Menard, S. (2002) Applied Logistic Regression Analysis. Newbury Park, CA: Sage.

Otterbourg, S. (2001) The Partnership for Family Involvement in Education: Who We and What We Do. Jessup, MD: The Partnership for Family Involvement in Education.

Phillips, L., Norris, S., Osmond, W. & Maynard, A. (2002) Relative Reading Achievement of 187 Children from First through Sixth Grades. Journal of Educational Psychology. 94: 3-13.

Raudenbush, S. and Bryk, A. (2002) Hierarchical Linear Models. Thousand Oaks, CA: Sage.

Rebora, A. (2002) No Child Left Behind. Education Week on the Web, April 2. http://www.edweek.org/context/topics/issuespage.cfm?id=59

Schineller, L. (1997) An Econometric Model of Capital Flight from Developing >Countries. Washington, D.C.: International Finance Discussion Paper, Number 579, Board of Governors of the Federal Reserve System.

Seldon, R (2001) Parent Power: Why National Standards Won't Improve Education. Washington, D.C.: The Cato Institute.

Snijders, T. and Bosker, R. (1999) Multilevel Analysis. Thousand Oaks, CA: Sage.

Solon, G., Page, M., and Duncan, G. (2000) Correlations Between Neighboring Children in Their Subsequent Educational Attainment. Review of Economics and Statistics. 82: 383-393.

SPSS (2001) SPSS Advanced Models 11.0. Chicago, IL: SPSS.

Thompson, C. and Cunningham, E. (2000) Retention and Social Promotion: Implications for Policy. New York: Teachers College, Columbia University, ERIC Clearinghouse on Urban Education.

U.S. Census Bureau (2001) QuickFacts. Washington, D.C.: U.S. Census Bureau, Department of Health and Human Services.

U.S. Department of Education (2002a) The "No Child Left Behind Act of 2001," Executive Summary (Updated). Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Education (2002b) The "No Child Left Behind" Act: Reauthorization of the Elementary and Secondary Act Legislation and Policies Website. July 11. <www.ed.gov/offices/OES E/esea/>

U.S. Department of Education (2003) Comments by Secretary Paige to the Commonwealth Club of California. March 12. www.ed.gov/02-2003 /03122003a.html

Velicer, W. (1972) The Moderator Variable Viewed as Heterogeneous Regression. Journal of Applied Psychology. 56: 266-269.

Vartanian, T., and Gleason, P. (1999) Do Neighborhood Conditions Affect High School Dropout and College Graduation Rates. Journal of Socioeconomics. 28: 21-42.

White House (2002) Fact Sheet: No Child Left Behind. January 8. www.whitehouse/gov/news/releases/2002/01/20020108.html

White House (2001) Transforming the Federal Role in Education so that No Child No Child is Left Behind. December 12. www.whitehouse.gov/news/reports/no-child-left-behind.html

Woodcock, R. and Johnson, M. (1990) Woodcock-Johnson Tests of Achievement. Allen, TX: DLM Teaching Services.

About the Authors

Robert Bickel is Professor of Advanced Educational Studies at Marshall University. His recent research is concerned with correlates of crime on school property, the limits of educational reform in promoting rural development, and adverse consequences of schools' efforts to meet the requirements of "No Child Left Behind." He has recently completed a monograph on multi-level analysis for education policy analysts.

Stan Maynard is Professor of Secondary Education at Marshall University. He is also Executive Director of the June Harless Center for Rural Educational Research and Development. His primary interest has long been development of innovative, cost-effective ways to assure high-quality public education for less-advantaged students living in isolated rural areas

---

The World Wide Web address for the *Education Policy Analysis Archives* is **epaa.asu.edu**

Editor: Gene V Glass, Arizona State University

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

EPAA is published by the Education Policy Studies Laboratory, Arizona State University