Oklahoma School Grades: Hiding “Poor” Achievement

Technical Addendum

The Oklahoma Center for Education Policy (University of Oklahoma) and
The Center for Educational Research and Evaluation (Oklahoma State University)

October 2013

© OCEP and CERE

The contributing authors are responsible for the accuracy of the information and claims made herein; the opinions expressed are theirs and they do not necessarily represent the views of The University of Oklahoma or Oklahoma State University.
Contributing authors:

The Oklahoma Center for Education Policy (University of Oklahoma)
- Curt M. Adams, Senior Research Scientist
- Ellen A. Dollarhide, Research Associate
- Patrick B. Forsyth, Senior Research Scientist
- Ryan C. Miskell, Research Associate
- Jordan K. Ware, Research Associate

The Center for Educational Research and Evaluation (Oklahoma State University)
- Laura L.B. Barnes, Senior Research Scientist
- Jam Khojasteh, Senior Research Scientist
- Mwarumba Mwavita, Senior Research Scientist

Corresponding authors:

Curt M. Adams, Curt.Adams-1@ou.edu
Patrick B. Forsyth, patrick.forsyth@ou.edu
TECHNICAL ADDENDUM SUMMARY

This technical addendum presents findings on our analysis of A-F grades using scale scores from the Oklahoma curricular exams. We wanted to include these findings for two reasons. First, some have suggested that the findings presented in Oklahoma School Grades: Hiding “Poor” Achievement are flawed because raw scores do not yield the same performance interpretations as scale scores. Second, there is a national interest in knowing if the spotlight on subgroup performance has diminished with accountability systems approved under state waivers to No Child Left Behind.

An accountability system weighted to measure equitable achievement distributions would see performance indicators decline as within-school achievement gaps increased. As seen in the figures below, achievement gaps in reading and math moved in a direction opposite from what one would expect. Achievement gaps in reading and math increased as school GPA and grades improved. The problem is not attributed to our sample of schools; but rather, the inability of the A-F components to measure achievement variance within schools.

Figure 1. Moderating effect of school A-F GPA on the reading achievement of FRL and non FRL students

![Figure 1](image1.png)

Figure 2. Moderating effect of school A-F GPA on the math achievement of FRL and non FRL students

![Figure 2](image2.png)
Oklahoma School Grades: Hiding “Poor” Achievement

TECHNICAL ADDENDUM

We decided to add this technical addendum to the report Oklahoma School Grades: Hiding “Poor” Achievement for two reasons. First, some have suggested that the findings are flawed because raw scores do not yield the same performance interpretation as scale scores. Second, national advocacy groups have raised concerns that efforts to fix problems of No Child Left Behind through flexibility waivers have unintentionally shifted attention away from achievement gaps and the performance of poor, minority students.¹ Thus, we wanted to include our findings on within-school achievement gaps using scale scores from Oklahoma curricular exams.

Method

As described in Oklahoma School Grades: Hiding “Poor” Achievement, we examined test scores of over 15,000 students in 63 urban schools² by state-assigned letter grade. We used scale scores in this analysis. Scale scores ranged from 440-990. The average math scale score for the sample was 701 with a standard deviation of 97, reading was 699 with a standard deviation of 96, and science was 744 with a standard deviation of 79. The school sample shows that the average FRL rate was 86%; 5% of the schools earned school grades of “A”; 13% earned


² The charge that the sample is not representative of schools in the state is true, but irrelevant to the fact that for the more than 15,000 students and 63 schools in our sample the state assigned letter grade obscured actual achievement differences among schools, failed to measure achievement equity within schools, and obscured inconsistent achievement across subjects. The problem is a measurement one.
grades of “B”; 21 % earned grades of “C”; 54 % earned grades of “D”; 8 % grades of “F”; 49 schools were elementary; and 14 were middle schools.

Analysis

We tested a multi-level, random coefficient regression model (intercepts and slopes as outcomes). Student predictors were group-mean centered because our primary interest was in testing the efficacy of letter grades in measuring achievement equity within schools. Group-mean centering provides a more accurate estimate of variation in achievement gaps across schools. To further increase the reliability of the achievement gap estimation, we used the State calculated school GPA as a single predictor variable. GPA is a continuous variable that is used to determine the categorical letter grade. The slopes and intercepts as outcomes models had less error and best fit with the data. In other words, these models provided an unbiased assessment of mean differences between “A”, “B”, “C”, “D”, and “F” schools, as well as the moderating effect of school GPA on the achievement of FRL and minority students. Estimates represent the actual difference in scale scores after controlling for factors not related to teaching practices and school performance.

\[
\text{Level 1: } \quad A_{ij} = \beta_0j + \beta_1j \text{(Minority Status}_{ij}) + \beta_2j \text{(Gender}_{ij}) + \\
\beta_3j \text{(FRL Status}_{ij}) + r_{ij}
\]

\[
\text{Level 2: } \quad \beta_0j = \gamma_{00} + \gamma_{01} (B) + \gamma_{02} (C) + \gamma_{03} (D) + \gamma_{04} (F) + \\
\gamma_{05} (\% \text{Minority}) + \gamma_{06} (\text{Prior Achievement}) + u_{0j}
\]

\[
\beta_1j = \gamma_{00} + \gamma_{11} (\text{GPA}) + u_{0j}
\]

\[
\beta_2j = \gamma_{00} + u_{0j}
\]

\[
\beta_3j = \gamma_{00} + \gamma_{31} (\text{GPA}) + u_{0j}
\]

\( \beta_{0j} = \) is the school achievement mean for math achievement

---

Technical Addendum 6

\[ \beta_{1j} = \text{Minority achievement gap} \]
\[ \beta_{2j} = \text{distributive effects of grade} \]
\[ \beta_{3j} = \text{FRL achievement gap} \]
\[ \gamma_{00} = \text{grand mean for achievement} \]
\[ \gamma_{01} = \text{is the difference in average achievement between A schools and B schools} \]
\[ \gamma_{02} = \text{is the difference in average achievement between A schools and C schools} \]
\[ \gamma_{03} = \text{is the difference in average achievement between A schools and D schools} \]
\[ \gamma_{04} = \text{is the difference in average achievement between A schools and F schools} \]
\[ \gamma_{05} = \text{is the effect of school \% Minority on achievement} \]
\[ \gamma_{06} = \text{is the effect of prior school achievement on student achievement} \]
\[ \gamma_{11} = \text{cross-level interaction of minority achievement and school GPA} \]
\[ \gamma_{31} = \text{cross-level interaction of FRL achievement and school GPA} \]

Results

Consistent with findings using raw test scores, achievement differences in reading, math, and science were small and for several of the letter grades not statistically different from zero (Table 1). For reading, the small average differences between students in “A”, “B”, and “C” schools were more likely the result of chance and sampling error than systematic differences in school performance. These small differences also fall within the average measurement error of the 3rd – 8th grade State reading assessment (average SEM = 33),\(^4\) suggesting that there is no difference in average reading achievement between students in “A”, “B”, and “C” schools.

Average reading differences between “A” and “D” (\(\gamma_{03} = -51, p<.01\)) and “A” and “F” schools (\(\gamma_{04} = -40, p < .01\)) were statistically significant with small effect sizes (Cohen’s \(d = .4\)).\(^5\) The average student in a “D” school scored about a ½ standard deviation lower than the average student in an “A” school. Average achievement was actually higher in “F” schools than “D” schools with “F” students scoring less than a ½ standard deviation lower than “A” schools. The small effect sizes translate to approximately a 73% overlap in the achievement distributions of “A” schools and “D/F” schools.

---


Mean differences between letter grades in math were larger than reading. The difference between “A” and “B” schools (γ01 = -32) was not significantly different from zero and fell around the average measurement error of 3rd – 8th grade State math assessments (average SEM = 31). Mean differences between “A” and “C” (γ02 = -53, p< .01), “A” and “D” (γ03 = -59, p <.01), and “A” and “F” (γ04 = -56, p < .01) were statistically significant with medium effect sizes (d for “A/C” = .5; d for “A/D” = .6; d for “A/F” = .57). These effect sizes translate to a 67% overlap in the achievement distributions of “A” and “C” schools and a 62% overlap in the distributions of “A” and “D/F” respectively. Findings also indicate small mean differences between “C” and “D” (6 scale points) and “C” and “F” (3 scale points) schools. Additionally, similar to reading achievement, “F” schools had higher average math achievement than “D” schools.

School grades were least effective at predicting differences in science achievement. Average differences among “A”, “B”, “C”, and “D” schools were not significantly different from zero and well within the average measurement error of the 5th and 8th grade State science exams (average SEM = 29). The only statistically significant difference was between “A” and “F” (γ04 = -42, p < .01) with a small effect size (d = .4). This difference translates to a 73% overlap in the achievement distributions of “A” and “F” schools.
Table 1. 
Results of the intercepts and slopes-as-outcomes models for reading, math, and science achievement

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Reading</th>
<th>Math</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>697.49 (1.6)**</td>
<td>699.31 (2.1)**</td>
<td>742.63 (2.02)**</td>
</tr>
<tr>
<td>“B”</td>
<td>-8.36 (10.7)</td>
<td>-32.13 (24.4)</td>
<td>-18.60 (10.6)</td>
</tr>
<tr>
<td>“C”</td>
<td>-23.52 (12.7)</td>
<td>-53.45 (15.1)**</td>
<td>-19.15 (11.7)</td>
</tr>
<tr>
<td>“D”</td>
<td>-51.74 (14.6)**</td>
<td>-59.11 (20.1)**</td>
<td>-30.80 (16.5)</td>
</tr>
<tr>
<td>“F”</td>
<td>-40.60 (14.9)**</td>
<td>-56.51 (20.7)**</td>
<td>-42.61 (16.2)**</td>
</tr>
<tr>
<td>%Minority</td>
<td>-17.26 (3.0)**</td>
<td>-16.13 (3.1)**</td>
<td>-15.83 (2.67)**</td>
</tr>
<tr>
<td>Prior Achievement</td>
<td>9.64 (2.7)**</td>
<td>20.8 (9.4)*</td>
<td>10.91 (2.9)**</td>
</tr>
<tr>
<td>Minority Slope</td>
<td>-16.2 (2.0)**</td>
<td>-23.29 (1.8)**</td>
<td>-23.86 (2.2)**</td>
</tr>
<tr>
<td>GPA</td>
<td>-6.6 (1.5)*</td>
<td>-5.56 (1.6)*</td>
<td>-2.19 (2.2)</td>
</tr>
<tr>
<td>FRL Slope</td>
<td>-11.7 (2.3)**</td>
<td>-12.36 (2.4)**</td>
<td>-13.32 (3.1)**</td>
</tr>
<tr>
<td>GPA</td>
<td>-12.8 (2.1)**</td>
<td>-11.68 (1.9)**</td>
<td>-7.05 (2.7)**</td>
</tr>
<tr>
<td>Male Slope</td>
<td>-13.35 (1.3)**</td>
<td>3.46 (1.23)**</td>
<td>6.11 (1.9)**</td>
</tr>
</tbody>
</table>

Deviance (-2 Log likelihood) | 178078 | 181590 | 55583
Δ Deviance | 638** | 586** | 343**
Explained Between School Variance | 92% | 89% | 91%

Note. * p<.05, **p<.01. We had valid reading data for 15,380 students, valid math data for 15,315 students, and valid science data for 4,935 students. Estimates come from random intercept and slopes as outcomes models. Standard errors are reported in parentheses. Student controls include FRL status, minority status, and male. Contextual controls include prior achievement and percent minority. Student variables were group-mean centered and full maximum likelihood estimation was used. Dummy coding was used for school letter grade. “A” schools were set as the referent group; thus, regression estimates represent the mean difference with students in “A” schools.

Achievement of Poor, Minority Students

Similar to findings using raw scores, within-school achievement gaps moved in a direction opposite from what would be expected if letter grades were weighted to account for achievement equity. Schools with higher grade point averages (GPA) had significantly larger
within-school FRL\(^6\) achievement gaps in reading \((\gamma 31 = -13, \ p < .01)\), math \((\gamma 31 = -12, \ p < .01)\), and science \((\gamma 31 = -7, \ p < .01)\) (Table 1). For reading, the average FRL gap within schools was approximately 12 points and increased to approximately 25 points in schools with the highest GPA (“A” and “B” schools). For math, the average FRL gap within schools was approximately 12 points and increased to 24 points in schools with the highest GPA. For science, the average FRL gap within schools was 13 scale points and increased to approximately 21 points. Minority achievement gaps within schools followed a similar pattern as FRL, but the achievement gap increase was not as large. As GPA increased so did the minority gap in reading \((\gamma 11 = -6.6, \ p < .05)\) and math \((\gamma 11 = -5.6, \ p < 05)\). In science, the within-school achievement gap did not increase by a statistically significant margin (Table 1).

Figures 1-4 show both the decline in average achievement of poor, minority students and the increase in achievement gaps as school grades increase. Average achievement of FRL students depreciated significantly as school GPA increased (Figures 1 and 2). FRL students actually had the lowest average reading and math achievement in schools with the highest State calculated GPA (“A” and “B” schools). Moreover, the FRL achievement gap widened considerably as school GPA increased. Schools with the highest GPA had inequitable achievement distributions, yet these schools received either an “A” or “B” even though their FRL students had lower average achievement than comparable peers in “D” and “F” schools.

\(^6\) FRL stands for students who qualify for the federal Free or Reduced Lunch Program. It is commonly used as a proxy for poverty.
Figure 1. Moderating effect of school A-F GPA on the reading achievement of FRL and non FRL students

Figure 2. Moderating effect of school A-F GPA on the math achievement of FRL and non FRL students

The minority achievement gap followed a similar pattern as the FRL gap (Figures 3 and 4). Minority students in schools with a lower GPA ("D" and "F" schools) had higher average math and reading achievement than minority students in “A” and “B” schools. Additionally, the minority achievement gap increased as school GPA increased. Again, schools with the most
inequitable achievement distributions received an “A” or “B” even though their minority students had lower average achievement than comparable peers in “D” and “F” schools. Overall, figures 1-4 show an accountability system that hides FRL and minority achievement gaps and obscures the performance of poor, minority students.

**Figure 3.** Moderating effect of school A-F GPA on the reading achievement of minority and non-minority students

![Figure 3](image)

**Figure 4.** Moderating effect of school A-F GPA on the math achievement of minority and non-minority students

![Figure 4](image)
In summary, as with raw test scores, actual achievement differences between letter grades were small and in most cases not statistically significant when student and school characteristics were held constant. School grades did moderate achievement gaps, but gaps moved in a direction opposite from what would be desired of an accountability system that measured achievement equity. To account for achievement equity, the FRL and minority gaps need to depreciate as GPA and letter grades improve. This, however, was not the effect of letter grades. Achievement gaps were larger in schools with higher grades. Although our evidence is limited to urban schools in Oklahoma, the system is likely to behave the same way in a different context with different schools. It is not variability in schools that presents a problem, but rather weaknesses of the components in measuring achievement variance within schools.