

How Good is Your School?

Steve Fleming, National Center for Educational Achievement, Austin, TX

ABSTRACT

With the widespread standardized testing put in place by the No Child Left Behind Act (NCLB), there is more performance data available on schools than ever before. Can this data be used to judge school effectiveness? For example, are performance results attributable to the school or to the community from which the school draws? These questions will be explored in the context of four accountability models: status, improvement, growth, and value-added. SAS code for implementing the four models and reporting the results will be demonstrated. For the data in this study, the first three models yielded similar results while the value-added model was less related to the other three and to school demographics. A familiarity with regression analysis will be helpful to get the most from this presentation.

INTRODUCTION

The No Child Left Behind Act of 2001 specified that students be tested annually in reading and mathematics in Grades 3 through 8. This has led to an explosion in the amount of academic data available on schools. Previous work showing how SAS can be used for school accountability has focused on data management, aggregation, and reporting (Mulvenon, et. al., 2000). In this paper, the question will be explored from the standpoint of determining the effectiveness of middle schools (Grades 6 through 8).

There are four general models used in education research to judge school effectiveness: status, improvement, growth, and value-added (Table 1; Goldschmidt, et. al., 2005). Previous work has shown the results from status, improvement, and growth models to be more similar to each other than to value-added models (Yu, et. al., 2007). However, as more years of assessment data are considered, status and growth measures diverge (Goldschmidt, et. al., 2005). Each of these models will be described in turn along with SAS code to implement them. A discussion of the advantages and disadvantages of each will also be given. To begin, however, a description of the data used to demonstrate the models is provided.

Table 1: Four general models used in education research

Model	Description	Question (Goldschmidt, et. al., 2005)	
Status	Uses a snapshot of assessment data from a single year as an indicator of school effectiveness	On average, how are students performing this year?	Simple
Improvement	Uses assessment data from multiple years at the same grade-level to project a school's status in the future	On average, are student doing better this year as compared to students in the same grade last year?	↓ Complex
Growth	Uses individual student assessment data to project their status in the future. A school's effectiveness is then aggregated from predicted student achievement	How much, on average, did individual students' performance change?	
Value-added	Uses individual student assessment data to estimate how much value a school has added to a student's learning. School effectiveness under this model is judged by how different a school's results are from typical (Allen et. al., 2009)	By how much did the average change in student performance miss or exceed the growth expectation?	

DATA

Data on student test performance in mathematics was analyzed for 39 schools that had a grade span of 6 to 8 for each school year from 2007-08 to 2009-10. A range of identification and demographic variables were available for each school (Table 2). Of the 39 schools, 4 were magnet schools. The 39 middle schools averaged about 530 students but ranged from 155 to 1,030 (Table 3). For the average school, 60% of the students qualified for free and reduced lunch although at least one school had as low as 20% and at least one as high as 100%. The ethnic distribution of the schools varied greatly although around 90% of students were either African American or White.

Table 2: School variables

Variable	Type	Length	Label
Pschl_AfrAm	Num	8	% African American
Pschl_Asian	Num	8	% Asian
Pschl_Hisp	Num	8	% Hispanic
Pschl_NatAm	Num	8	% Native American
Pschl_White	Num	8	% White
Pschlep	Num	8	% having limited English proficiency
Pschspec	Num	8	% identified as special education
campus_id	Char	7	Campus ID
campus_name	Char	30	Campus Name
high_grade	Num	8	High grade in school
low_grade	Num	8	Low grade in school
magflag	Num	8	Magnet school flag; 0=no, 1=yes
nschool	Num	8	Number in school
pschlow	Num	8	% receiving free and reduced lunch

Table 3: School statistics

Label	N	Mean	Std Dev	Minimum	Maximum
Number in school	39	532.7	228.3	155.0	1030.0
% receiving free and reduced lunch	39	60.5	19.2	20.4	100.0
% having limited English proficiency	39	3.3	4.5	0.0	16.6
% identified as special education	39	10.2	3.4	0.6	16.9
% African American	39	34.3	31.7	0.0	96.1
% Hispanic	39	5.7	5.6	0.5	21.7
% Asian	39	2.1	5.4	0.0	33.8
% Native American	39	0.9	2.1	0.0	10.7
% White	39	56.5	31.2	3.0	99.0

To demonstrate the results of each model, a smaller set of five schools was selected by a systematic random sample controlling for the percentage of students in each school that qualified for the free and reduced lunch program (Code Sample 1). This helps ensure the sample schools have a range of demographic composition.

Code Sample 1

```
proc surveysselect data=scsug.middle_schools
    out=middle method=sys /* Systematic Random Sample */
    sampsize=5 seed=641344710; /* Specifying the seed allows results to be
                                reproducible. Value created using the RAND
                                function of a TI-30X IIS calculator */
    control pschlow; /* The data is sorted by the % of students in the school receiving free
                    and reduced lunch before taking the sample. This ensures that the
                    sample of schools has a spread of values on this variable */
run;
```

To preserve school anonymity the five sample schools were renamed for the first five presidents of the United States. They range in size from just over 300 students to over 1,000 students (Table 4). The percentage of students receiving free and reduced lunch varies from 39% to 97%. The ethnic distribution in the schools also varies greatly with two having more the 80% White students (Washington and Madison) and one having 80% African American students (Monroe). Adams Middle School is identified as a magnet school.

Table 4: Sample school demographics

School	Number in school	% receiving free and reduced lunch	% having limited English proficiency	% identified as special education	% African American	% Hispanic	% Asian	% Native American	% White	Magnet School?
Washington	1,030	39	2	12	11	5	2	0	82	No
Adams	867	53	4	5	51	6	3	0	40	Yes
Jefferson	376	58	1	6	22	6	0	0	72	No
Madison	328	70	8	11	0	7	5	3	84	No
Monroe	623	97	16	14	80	16	1	0	3	No

The student mathematics test data contains fields indicating the student ID, the test score and other attributes (Table 5). Of the students in the middle schools in 2009-10, 97% were tested in mathematics, 97% were enrolled at their school for the full academic year, 33% met the College and Career Readiness Target in mathematics, and about a third were in each grade level. The average student score was 25 points higher in the current year than the prior year (Table 6).

Table 5: Student variables

Variable	Type	Length	Label
FAY	Num	8	Student was enrolled in the tested school for the full academic year; 0=no, 1=yes
campus_id	Char	7	Campus ID of the school where the student was tested
campus_name	Char	30	Campus name of the school where the student was tested
ccr	Num	8	Student met the CCR Target on the test; 0=no, 1=yes
grade	Num	8	Grade level in which the student was tested; 6, 7, or 8
pssc_mt	Num	8	The student's scale score in mathematics in the prior year and grade
sid	Char	10	Student Identifier
ssc	Num	8	The student's scale score in mathematics in the current year
subject	Char	11	Subject tested
tested	Num	8	Student was tested; 0=no, 1=yes
year	Num	8	Year tested, 2010 means the 2009-10 school year

Table 6: Student scale score statistics 2009-10

Label	N	Mean	Std Dev	Minimum	Maximum
The student's scale score in mathematics in the current year	20,187	714	99	115	993
The student's scale score in mathematics in the prior year and grade	18,679	689	107	58	999

COLLEGE AND CAREER READINESS

Most state accountability systems are based around getting students to a state-determined level of proficiency on standardized tests. Research by the National Center for Educational Achievement (NCEA), a department of ACT, Inc. has shown that students who just achieve the proficiency standard on a state's Grade 11 mathematics exam have a less than 10% chance of reaching the College Readiness Benchmark (CRB) on the ACT mathematics test by Grade 12 (Dougherty, 2008b). The CRBs on the ACT exams are in turn based on student success in college as determined by course grades (Allen and Sconing, 2005). NCEA defines College and Career Readiness (CCR) as when a student has reached an academic achievement level that indicates they are likely to have success in postsecondary learning or training that leads to skilled careers. We call the lowest score that achieves this academic achievement level the CCR Target which was used as the standard for student performance in this study.

STATUS MODELS

Status models refer to a snapshot performance statistic from a particular moment in time which may be compared to a goal. A well-known status model goal is set in the NCLB legislation of having every student meet the proficiency standard by 2014.

These measures have the advantage of being easy to calculate and explain. However, status models do not account for incoming student ability. For example a student who fell from an advanced achievement level to just above the CCR Target is given credit while a student who advanced from well below the CCR Target to just below it is not. In this regard, many educators view status measures as unfair and a reflection of student background rather than school effectiveness (Allen, et. al., 2009). Status models, however, are the primary means by which school performance is judged under NCLB. For this study, the percentage of students meeting the CCR Target on the mathematics test in 2010 will be used as the status measure (Table 7).

Table 7: Cohort grade levels contributing to the status measure

Cohort	2007	2008	2009	2010
1	7	8		
2	6	7	8	
3		6	7	8
4			6	7
5				6

% CCR 2010

Contributes to the status measure

Status measures can be calculated in many ways in SAS. One version is shown below using PROC SUMMARY and PROC REPORT (Code Sample 2). On the status measure, Jefferson and Washington Middle Schools appear to have distinguished themselves (Table 8). Monroe Middle school lags behind with only 9% of students reaching the CCR Target.

Code Sample 2

```

/* proportion of students meeting
the CCR Target */
proc summary data=arm_10 nway ;
  class campus_id;
  id campus_name;
  var ccr;
  output out=scsug.status mean=pCCR ;
run;

proc sort data=scsug.status;
  by descending pCCR ;
run;

proc report data=scsug.status nowd; /* Table 8 */
  column campus_name pCCR ;
  define campus_name / display "School";
  define pCCR / display "% CCR 2010"
    format=percent8.;
run;

```

Table 8: The percentage of students meeting the CCR Target in mathematics in 2010 varies by 40%

School	% CCR 2010
Jefferson	49%
Washington	48%
Adams	33%
Madison	31%
Monroe	9%

IMPROVEMENT MODELS

Improvement measures project a school's performance in the future based on a performance trend over time. For example, if a school had 50% of its Grade 8 students meeting the CCR Target in mathematics in 2009 and 60% of its Grade 8 students meeting the target in 2010, a simplistic improvement measure would project 70% of students will reach the CCR Target in 2011. Improvement measures are popular under NCLB because they can project future achievement and compare it to the goal of 100% proficiency in 2014 (Allen, et. al., 2009). However, improvement measures suffer from some deficiencies. Most importantly, an extrapolation into the future is not necessarily wise using a regression model because the assumptions of the model may not hold into the future (Stevens, 2007). For example, an assumption is made that the academic preparation of students has held steady over the years of input data and will hold steady in the future. For school's experiencing rapid changes in student academic preparation, improvement measures can be wildly off target. In addition, the independence of results assumption is violated because about two-thirds of the students from one year will also contribute to the test results for the following year. For our data, an improvement measure will be based on a projection of the percentage of students meeting the CCR Target in 2011 based on status model of data from 2008-2010 (Table 9).

Table 9: Cohort grade levels contributing to the improvement measure

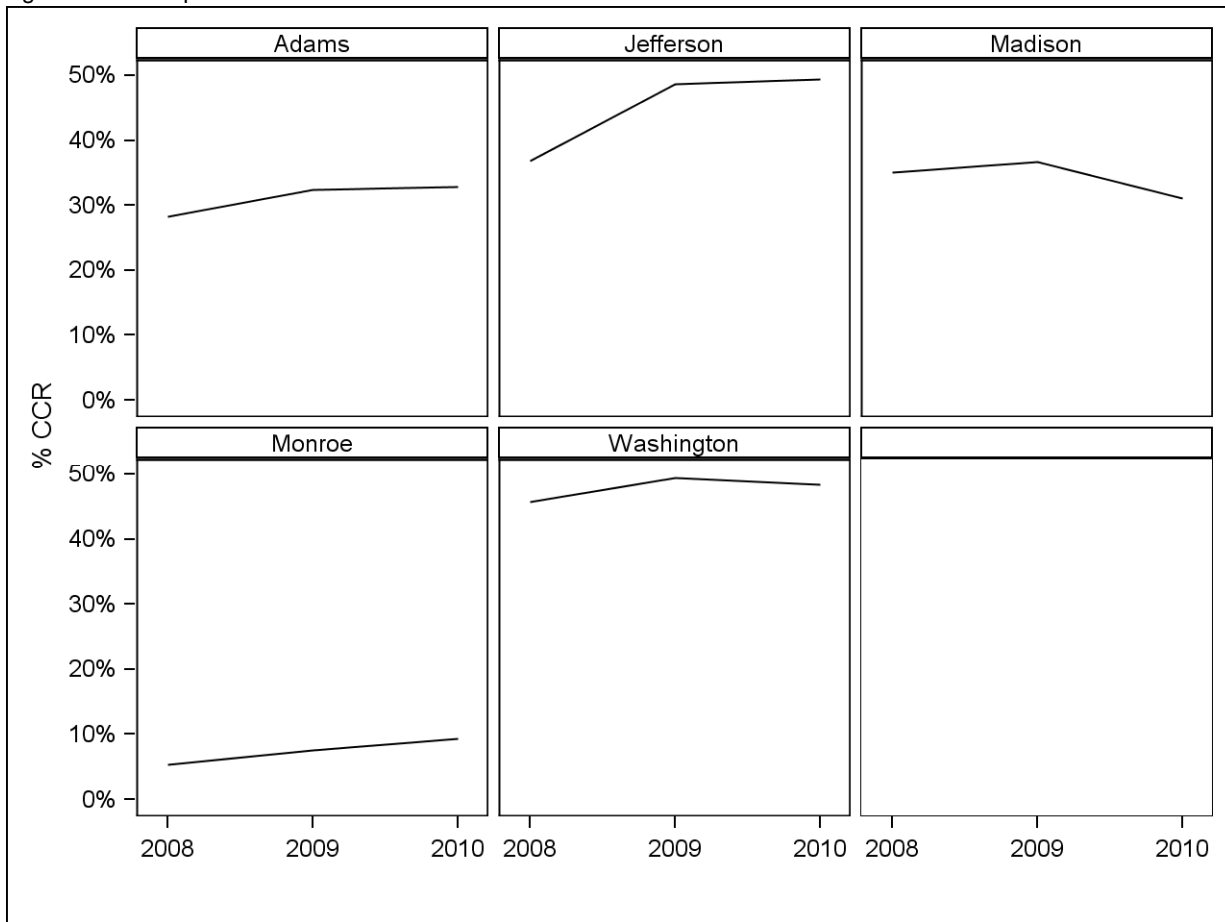
Cohort	2007	2008	2009	2010	2011
1	7	8			
2	6	7	8		
3		6	7	8	
4			6	7	
5				6	

%CCR 2008
% CCR 2009
% CCR 2010
Projected % CCR 2011

- Contributes to the improvement measure
- Contributes to the logistic regression

Performance trends show that all schools but Madison improved their percentage of students reaching the CCR Target between 2008 and 2010 (Figure 1). Jefferson Middle School appears to have been improving more quickly than the other schools and should fare well under the improvement model. Figure 1 was produced using the SG PANEL procedure (Code Sample 3). A logistic regression model was fit to the CCR rate over time. Logistic regression was chosen because it aligns with the constraint of the CCR rate between 0 and 1 (Equation 1; Hosmer and Lemeshow, 1989). In this study, time is the explanatory variable and no covariates are included in the logistic regression model although they could be. The results show that Jefferson Middle School is on track to distinguish themselves from Washington Middle School in 2011 (Table 10).

Figure 1: School performance trends



Equation 1: Logistic regression model

$$\ln\left(\frac{pCCR}{1-pCCR}\right) = \beta_0 + \beta_1(model_year)$$

Code Sample 3

```
proc summary data=arm_08to10 nway;
  class campus_id year;
  var ccr;
  output out=school_year (drop=_type_ rename=(freq=nStudents)) sum=nCCR mean=pCCR ;
run;

data scsug.school_year;
  set school_year;
  model_year = year-2009;
run;

proc sort data=scsug.school_year;
  by campus_name year;
run;

proc sgpanel data=scsug.school_year nocycleattrs; /* Figure 1 */
  panelby campus_name / columns=3 rows=2 novarname spacing=5;
  series x=year y=pCCR;
  rowaxis label="% CCR" min=0;
  colaxis label=" ";
  format pCCR percent8.;
run;

/* fit a logistic regression model to the CCR percentages */
proc logistic data=scsug.school_year outest=est;
  by campus_id ;
  model nCCR / nStudents = model_year ;
run;

/* create data set to score based on regression model */
data log_score;
  set scsug.middle_schools (keep=campus_id) ;
  intercept = 1;
  model_year = 2;
run;

proc sort data=log_score;
  by campus_id;
run;

/* Calculate projected CCR rates for 2011 based on the parameters from the logistic
regression model */
proc score data=log_score predict score=est type=parms out=log_predict;
  by campus_id;
  var intercept model_year;
run;

data scsug.log_predict;
  set log_predict;
  pCCR_predict = exp(nCCR) / (1 + exp(nCCR)); /* Convert back to probability scale */
run;

proc sort data=scsug.log_predict;
  by descending pCCR_predict;
run;

proc report data=scsug.log_predict nowd; /* Table 10, columns 1 and 5 */
  column campus_name pCCR_predict;
  define campus_name / "School" display;
  define pCCR_predict / " " display format=percent8.;
run;
```

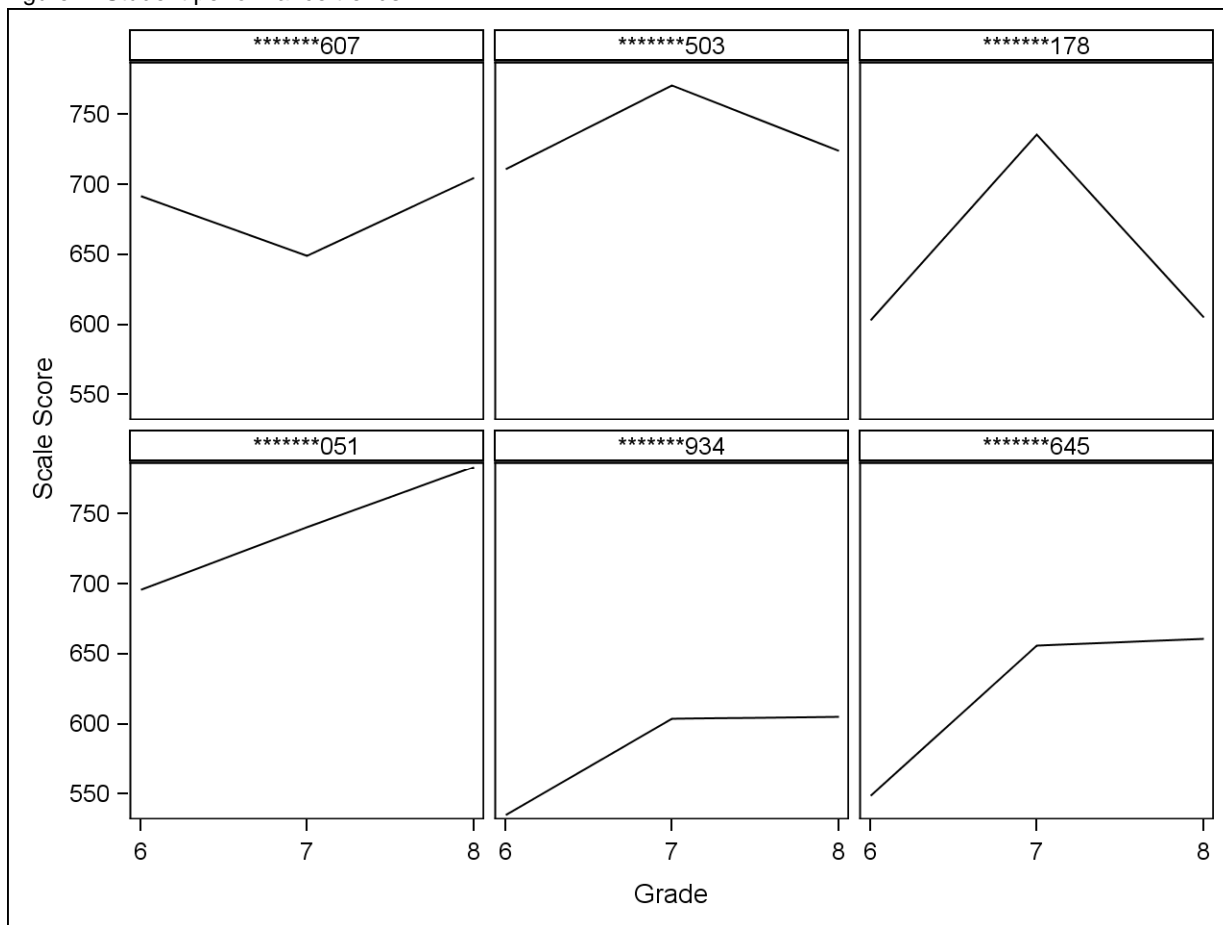
Table 10: Jefferson Middle School is projected to have 57% of students meeting the CCR Target based on the improvement measure

School	2008	2009	2010	Projected 2011
Jefferson	37%	49%	49%	57%
Washington	46%	49%	48%	50%
Adams	28%	32%	33%	36%
Madison	35%	37%	31%	30%
Monroe	5%	8%	9%	13%

GROWTH MODELS

In contrast to improvement measures which project a school's performance into the future, growth measures project scores for individual students. The key attributes of a growth model are that it includes an element of time and the purpose is to estimate future student performance (Wright, Sanders, and Rivers, 2006). School effectiveness can then be calculated from an aggregation of individual student growth. Some advantages of growth models include identifying whether students who are academically behind are growing quickly enough to catch up in a reasonable amount of time and identifying unusually effective schools (Dougherty, 2008a). Students vary in their progress across grade-levels (Figure 2), some improve and some decline while others do well one year and poorly the next.

Figure 2: Student performance trends



There are many types of growth measures. For example, growth can also be measured via a hierarchical linear model (HLM) with scores nested within students nested within schools (Singer & Willett, 2003). However, this study examined the Wright/Sanders/Rivers (WSR) growth methodology. This model predicts student test scores based on past test scores on the assumption that the student will have an average schooling experience in the future (Wright, Sanders, and Rivers, 2006). Some of the advantages of the WSR methodology include:

- The model accommodates missing values in the predictor variables.
- The test scores are not required to be measured on a continuous vertically-linked scale.
- The overall shape of the growth curve can be left unspecified.

Some drawbacks of the WSR methodology are (Wright, Sanders, and Rivers, 2006):

- At least two cohorts are required, one to calculate the parameters and one on which to apply the projections.
- When schools undergo rapid demographic shifts, the means and covariances in the first cohort may not apply as well to the second cohort.

In this study, Grade 8 test scores are projected from Grades 6 and/or 7 test scores (Equation 2). Cohort 3 results will be used to estimate the projection parameters which will be applied to Cohort 4.

Equation 2: WSR Projection

$$Projected\ Score_i = M_Y + \beta_1(X_{1i} - M_1) + \beta_2(X_{2i} - M_2)$$

Where *Projected Score_i* is the projected Grade 8 test score for student *i*; β_1 , and β_2 are the regression parameters; X_{1i} and X_{2i} are the student's Grade 6 and 7 test scores respectively; M_Y , M_1 , and M_2 are the average of the schools' average Grade 8, Grade 6, and 7 test scores respectively. The WSR model is implemented in PROC IML using code provided by Jeff Allen of ACT (Appendix 1). The effect of Grade 6 and 7 scores on Grade 8 scores for Cohort 3 will be used to project the Grade 8 scores of Cohort 4 (Table 11). Test data is rearranged so that each student in Cohort 3 has one record containing separate variables indicating scale scores in Grades 6-8. First scores for students in Cohort 3 who were enrolled in the same school for the full academic year during their Grade 8 year were gathered, and then Grades 6 and 7 scores were obtained in a similar manner and merged with the Grade 8 scores (Code sample 4). The data for Cohort 4 was then processed in a similar manner and appended to the Cohort 3 data while creating a field indicating that Cohort 4 should not be used to create the projection parameters. This was followed by the call to the WSR macro and the coding of the projected score as meeting the CCR Target or not. The projected CCR status is then aggregated by school for Cohort 4 just as we did with the Status measure. The difference in projecting results at the student level versus the school level is that Washington Middle School tops Jefferson Middle School based on this measure (Table 12).

Table 11: Cohort grade levels contributing to the growth measure

Cohort	2007	2008	2009	2010	2011
1	7	8			
2	6	7	8		
3		6	7	8	
4			6	7	Projected 8
5				6	

- ☐ Contributes to the growth measure
- Contributes to the WSR projection

Code Sample 5

```

/* Retrieve Grade 8 scores from 2010 for students enrolled in middle schools for the entire
year. */
data test_g8(drop=grade subject tested);
  set araggr.arm_student_10 (keep=grade subject sid campus_id ssc_10 FAY tested
  rename=(ssc_10=ssc_g8)
  where=(grade=8 and subject="MATHEMATICS"
  and campus_id in (&MS.) and FAY=1 and tested=1)) ;
run;

/* Retrieve Grade 7 scores from 2009. */
data test_g7(drop=grade subject tested);
  set araggr.arm_student_09 (keep=grade subject sid ssc_09 tested
  rename=(ssc_09=ssc_g7)
  where=(grade=7 and subject="MATHEMATICS" and tested=1)) ;
run;

```



```

/* Retrieve Grade 6 scores from 2008. */
data test_g6(drop=grade subject tested);
    set araggr.arm_student_08 (keep=grade subject sid ssc_08 tested
                             rename=(ssc_08=ssc_g6)
                             where=(grade=6 and subject="MATHEMATICS" and tested=1)) ;
run;

proc sort data=test_g8;
    by sid;
proc sort data=test_g7;
    by sid;
proc sort data=test_g6;
    by sid;
run;

data test;
    merge test_g8 (in=g8) test_g7 test_g6 ;
    by sid;
    if g8;

    /* wsr macro fails if the predictor variables are all missing */
    if missing(ssc_g6) and missing(ssc_g7) then delete;
run;

/* ... */

data test_all;
    set test          /* Cohort 3 */
        test_wsr(in=wsr); /* Cohort 4, Grade 7 scores from 2010 for students enrolled in
                           middle schools merged with Grade 6 scores from 2009. */
    use_in_project = 1;
    if wsr then use_in_project = 0;
run;

%wsr(test_all, campus_id, use_in_project, ssc_g7 ssc_g6, ssc_g8);

data scsug.growth_projection;
    set test_all;

    ccr_projection = 0;
    if projection1 >= 760 then ccr_projection = 1; /* 760 is the Grade 8 CCR Target */
run;

/* Calculate the growth measure */
proc summary data=scsug.growth_projection nway ;
    where use_in_project=0;
    class campus_id;
    var ccr_projection;
    output out=scsug.growth mean=pCCR ;
run;

proc sort data=scsug.growth;
    by descending pCCR ;
run;

proc report data=scsug.growth nowd; /* Table 12 */
    column campus_name pCCR ;
    define campus_name / display "School";
    define pCCR / display "Projected Grade 8 % CCR" format=percent8.;
run;

```

Table 12: Washington Middle School is projected to have the highest Grade 8 % CCR for Cohort 4 using the Wright/Sanders Rivers model

School	Projected Grade 8 % CCR
Washington	54%
Jefferson	52%
Madison	33%
Adams	30%
Monroe	9%

VALUE-ADDED MODELS

Concerns over fairness cause researchers to look for school accountability models that are not correlated with incoming student achievement or demographics. This allows performance to be compared among schools with similar characteristics (Goldschmidt et. al., 2005). However, this adjustment is incompatible with NCLB measures which must be the same for all demographic groups. Most importantly, value-added models account for the nesting of students within schools and allow covariates to be introduced at appropriate levels in the model. These models are HLM models as mentioned in the growth section. There are many potential models in the value-added family. Phan (2008) described a value-added model for mathematics data on an international assessment. In our case, the outcome was test scores controlled for test score in the prior grade at the student level (Equation 3). This is a simpler version of the value-added model used by NCEA to select higher performing schools.

Equation 3: Value-added Model (Student-level)

$$Y_{ij} = \beta_{0j} + \beta_1 X_{1i} + \varepsilon_{ij}$$

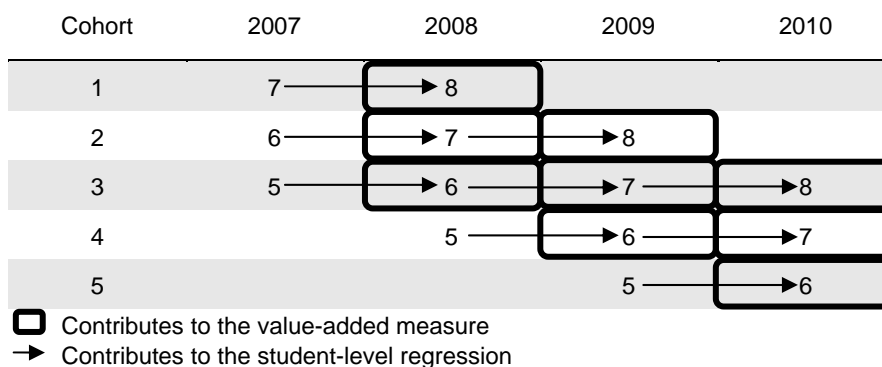
Where Y_{ij} is the test score for student i in school j ; β_{0j}, β_1 are the regression parameters; X_{1i} is the student's prior grade test score; and ε_{ij} is a normally distributed error term with mean 0 and standard deviation σ_y . At the school-level, the percentage of students qualifying for free and reduced lunch for each school is controlled for, and the remaining unexplained variance yields a term representing the value-added effectiveness each school. (Equation 4).

Equation 4: Value-added Model (School-level)

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_j$$

Where γ_{00}, γ_{01} are the school-level regression parameters; W_j is the percentage of students qualifying for free and reduced lunch in school j ; and u_j is a school-level random effect with mean 0 and standard deviation σ_u . An additional assumption of the model is that the terms ε_{ij} and u_j are uncorrelated. The effect of prior grade math score on current year score is calculated for students present in the schools in 2008, 2009, and 2010 (Table 13). Only students with a prior year math score and who were enrolled at the school for the full academic year are included in the analysis.

Table 13: Cohort grade levels contributing to the value-added measure



The value-added model is implemented in PROC MIXED (Code Sample 6). A 95% confidence interval for the school effect is calculated and the data is restructured for SAS/GRAPH. A table of the results is created using PROC REPORT, and shows that Jefferson Middle School added 20 points above typical to student mathematics scores (Table 14). The HILO interpolation is used with PROC GPLOT to produce a graph showing the confidence intervals (Figure 3). It is clear that Jefferson Middle School is providing improvement above that expected to the student mathematics scores compared to other schools.

Code sample 6

```
ods output solutionr=rand_effect ; /* ods used to get the school effects */
ods exclude solutionr; /* prevents entire list of residuals from printing */
proc mixed data=va_input noclprint covtest;
  class campus_id ;
  model ssc = pssc_mt pschlow / solution ddfm=bw;
  random intercept/type=un sub=campus_id s;
run;

data scsug.va_rand_effect;
  set rand_effect;

  estimate_low = estimate - 2*stdErrPred;
  estimate_high = estimate + 2*stdErrPred;
run;

proc sort data=scsug.va_rand_effect;
  by descending estimate ;
run;

data rand_bounds(drop=estimate:);
  set scsug.va_rand_effect;
  array est [3] estimate:;

  do i = 1 to 3;
    bound_value = est[i];
    output;
  end;
run;

proc report data=scsug.va_rand_effect nowd; /* Table 14 */
  column campus_name estimate ;
  define campus_name / display "School" ;
  define estimate / display "Estimated School Effect on Scale Score" format=3.0;
run;

/* see Cranford (2009), figure 24 */
axis1 label=none color=gray value=(color=black height=1) minor=none order=(-10 to 25 by 5);
axis2 label=(color=black 'School') value=(color=black height=1) color=gray major=none
  minor=none offset=(5pct, 5pct)
  order=("Jefferson" "Madison" "Washington" "Monroe" "Adams");
symbol i=hiloc value=none color=black;

proc gplot data=rand_bounds;
  plot bound_value * campus_name / noframe vaxis=axis1 haxis=axis2 vref=0;
  format bound_value 4.0;
run; quit;
```

CORRELATIONS BETWEEN MEASURES

An investigation of the correlations between school effectiveness measures for the 39 middle schools found that the calculated status, improvement, and growth measures are very similar (Table 15). This is consistent with previous findings (Yu et. al., 2007). The value-added measure, while positively related to the other measures was not as strongly related.

The relationship between the school effectiveness measures and school demographics is more varied (Table 16). The status, improvement, and growth measures are all strongly, negatively related with the percentages of low income, special education, and African-American students in the school. The value-added measure does not follow the same pattern only having a moderately negative relationship with the percentage of special education students in the school. If this relationship was of concern it would be simple to add percentage of special education students to the school level of the value-added model to zero out the relationship. No relationships were found between the school effectiveness measures and the number of students in the school and the percentages of limited English proficiency and Hispanic students in the school.

Table 14: Jefferson Middle Schools ranks highest on value-added.

School	Estimated School Effect on Scale Score
Jefferson	20
Madison	3
Washington	2
Monroe	-3
Adams	-4

Figure 3: Estimated school effects and 95% confidence intervals based on the value-added model

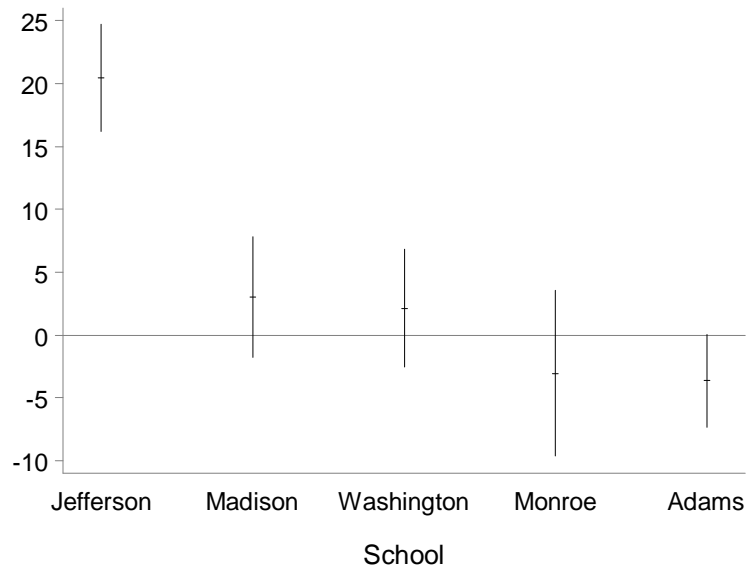


Table 15: Correlations between school effectiveness measures

Pearson Correlation Coefficients, N = 39

	Improvement	Growth	Value-Added
Status	0.98	0.96	0.54
Improvement		0.94	0.52
Growth			0.47

■ Statistically Significant at < 0.01 level

Table 16: Correlations between school effectiveness measures and school demographics

Pearson Correlation Coefficients, N = 39

	Status	Improvement	Growth	Value-Added
Number in school	0.10	0.10	0.01	-0.12
% students receiving free and reduced lunch	-0.77	-0.77	-0.79	-0.00
% students of limited English proficiency	-0.15	-0.16	-0.19	0.08
% students identified as special education	-0.61	-0.56	-0.61	-0.36
% White	0.61	0.61	0.65	0.11
% African American	-0.68	-0.66	-0.71	-0.16
% Hispanic	0.06	0.05	0.02	0.19

■ Statistically Significant at < 0.01 level

■ Statistically Significant between 0.1 and 0.01 levels

CONCLUSION

Four models for school effectiveness and their implementation in SAS were described. The relative rating of a school was dependent on the measure used. It is important to isolate and select the question you want answered and understand the full context surrounding your target audience before choosing a measure. The calculated status, improvement, and growth measures for 39 middle schools were found to be closely related to the percentage of students receiving free and reduced lunch while the value-added measure, by design, was not. Concern for fairness may play into your decision of what model to choose.

REFERENCES

- Allen, Jeff, Dina Bassiri, and Julie Noble. (2009). *Statistical Properties of Accountability Measures Based on ACT's Educational Planning and Assessment System (ACT Research Report Series 2009-1)*. Retrieved August 19, 2010 from http://act.org/research/researchers/reports/pdf/ACT_RR2009-1.pdf.
- Allen, Jeff and Jim Sconing. (2005). *Using ACT Assessment Scores to Set Benchmarks for College Readiness (ACT Research Report Series 2005-3)*. Retrieved September 21, 2010 from http://act.org/research/researchers/reports/pdf/ACT_RR2005-3.pdf
- Cranford, Keith. (2009). *How to Produce Excellent Graphs in SAS* (from the Proceedings of the 2009 South Central SAS Users Group Conference). Retrieved August 24, 2010 from http://www.scsug.org/SCSUGProceedings/2009/Keith_Cranford.pdf.
- Dougherty, Chrys. (2008a). *The Power of Longitudinal Data: Measuring Student Academic Growth*. Retrieved August 19, 2010 from http://www.nc4ea.org/files/dqc_academic_growth-10-09-08.pdf.
- Dougherty, Chrys. (2008b). *They Can Pass, but Are They College Ready?: Using Longitudinal Data To Identify College and Career Readiness Benchmarks on State Assessments*. Retrieved September 21, 2010 from http://www.nc4ea.org/files/dqc_state_ccr-10-09-08.pdf.
- Goldschmidt, Pete, Pat Roschewski, Kilchan Choi, William Auty, Steve Hebbler, Rolf Blank, and Andra Williams. (2005). *Policymaker's Guide to Growth Models for School Accountability: How do Accountability Models Differ?* Retrieved August 23, 2010 from http://www.wera-web.org/pages/activities/WERA_6_2_-6/Growth%20ModelsGuide%202005.pdf.
- Hosmer, David W. and Stanley Lemishow. (1989). *Applied Logistic Regression*. John Wiley & Sons, Inc.
- Lissitz, R. W., editor. (2005). *Value Added Models in Education: Theory and Applications*. Maple Grove Minnesota: JAM Press.
- Mulvenon, Sean W., Ronna C. Turner, Barbara J. Ganley, Antionette R. Thorn, Kristina A. Fritts Scott. (2000). *SAS Procedures for Use in Public School Accountability Programs* (paper 216-25 in the Proceedings of the 25th SAS Users Group International Conference). Retrieved August 19, 2010 from <http://www2.sas.com/proceedings/sugi25/25/po/25p216.pdf>.
- Phan, Ha T. (2008) *Building Mathematics-Achievement Models for Developed and Developing Countries: An Application of SAS® to Large-Scale, Secondary Data Management and Analysis* (paper 239-2008 from the 2008 SAS Global Forum Proceedings). Retrieved August 19, 2010 from <http://www2.sas.com/proceedings/forum2008/239-2008.pdf>.
- Singer, J.D. & Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press, Inc.
- Stevens, J.P. (2007). *A Modern Approach to Intermediate Statistics*. New York: Lawrence Erlbaum Associates.
- Wright, S. Paul, William L. Sanders, and June C. Rivers. (2006). *Measuring Academic Growth of Individual Students toward Variable and Meaningful Standards* in R. W. Lissitz (ed.) *Longitudinal and Value Added Models of Student Performance*. Maple Grove Minnesota: JAM Press.
- Yu, Fen, Eugene Kennedy, Charles Teddlie, and Mindy Crain (2007). *Identifying Effective and Ineffective Schools for Accountability Purposes: A Comparison of Four Generic Types of Accountability Models*. Retrieved August 23, 2010 from http://www.learningpt.org/sipsig/Yu_Kennedy_Teddlie_Crain.pdf.

ACKNOWLEDGMENTS

The author would like to recognize Jeff Allen, Dina Bassiri, and Julie Noble of ACT for their paper which inspired this work. The Research and Outreach teams of NCEA also greatly influenced this work. Special thanks to Leland Lockhart, Teresa Shaw, Charly Simmons, and Chrys Bouvier for their helpful comments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Steve Fleming
Enterprise: National Center for Educational Achievement
Address: 8701 North MoPac Expressway, Suite 200
City, State ZIP: Austin, TX 78759
Work Phone: 512-320-1827
Fax: 512-320-1877
E-mail: sfleming@nc4ea.org
Web: www.nc4ea.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX 1

```
/*Macro sent by Jeff Allen, ACT
```

```
This macro can be used to produce projected test scores using the methodology developed by Wright/Sanders/Rivers as described in Lissitz (2005) pp. 389-390
```

```
The inputs to the macro are:
```

1. The name of the SAS dataset that contains the predictor and response variable(s).
2. The SAS variable name for the school identifier.
3. The SAS variable name for whether the observation (student) is to be included in the estimation sample. Typically, students who have one or more observed response variables in the most recent year are included in the estimation sample. The variable indicating whether or not a student should be included in the estimation sample must be coded as 1=in estimation sample, 0=not in estimation sample. This must be a numeric variable.
4. The list of SAS variable names for the predictor variables. These must be numeric variables.
5. The list of SAS variable names for the response variables. Note that one can specify between 1 and 20 response variables. These must be numeric variables.

```
*/
```

```
%macro wsr(dataset, schoolid, estimationindicator, pv, dv);
```

```
data &dataset;  
set &dataset;  
line+1;
```

```
/* Obtain mean of school means for each predictor and response variable */  
proc sort data=&dataset;  
by &schoolid;
```

```
proc means data=&dataset noprint mean;  
where &estimationindicator=1;  
by &schoolid;  
var &pv &dv;  
output out=m mean=&pv &dv;
```

```
proc means data=m noprint mean;  
var &pv &dv;  
output out=m mean=&pv &dv;
```

```
/* Apply school-level mean centering to data per WSR instructions on page 390 */  
proc stdize data=&dataset out=st method=mean;  
where &estimationindicator=1;  
by &schoolid;  
var &pv &dv;
```

```
/* Obtain sample covariance matrix. Note that we are ignoring missing data and not using EM  
Algorithm */  
proc corr data=st cov outp=c noprint;  
var &pv &dv;
```

```
data c;  
set c;
```

```

if _TYPE_="COV";
drop _TYPE_ _NAME_;

proc iml;

/* Read in predictor variables into x, means of means into m and my and covariances into c */
use &dataset;
read all var{&pv} into x;

use m;
read all var{&pv} into m;
read all var{&dv} into my;

use c;
read all into c;

/* n is the sample size, npv is the number of predictors, ndv is the number of response
   variables */
n=nrow(x);
npv=ncol(m);
ndv=ncol(my);

/* Apply centering to predictors using means of means as given on page 389 */
do i=1 to npv;
  x[,i]=x[,i]-m[i];
end;

projections=j(n,ndv);

id=I(npv+ndv);

/* Since missing data patterns vary by student, projection parameters also vary by student.
   The code below obtains each students parameters based on their subset of the covariance
   matrix corresponding to their set of predictors */
do i=1 to n;
  counter=0;
  do j=1 to npv;
    if x[i,j] <> . then do;
      if counter=0 then do;
        xi=x[i,j];
        ci=id[j,];
      end;
      else if counter=1 then do;
        xi=xi||x[i,j];
        ci=ci//id[j,];
      end;
      if counter=0 then counter=1;
    end;
  end;
  ci=ci//id[(npv+1):(npv+ndv),];
  ci=ci*c*t(ci);
  npvi=ncol(xi);
  beta=inv(ci[1:npvi,1:npvi])*t(ci[(npvi+1):(npvi+ndv),1:npvi]);

/* Obtain projections for response variables using equation given on page 389 */
  projections[i,] = my + xi*beta;
end;

cname={"projection1" "projection2" "projection3" "projection4" "projection5" "projection6"
"projection7" "projection8" "projection9" "projection10" "projection11" "projection12"
"projection13" "projection14" "projection15" "projection16" "projection17" "projection18"
"projection19" "projection20"};
cname=cname[1:ndv];

/* Create a SAS dataset named work.projections that has the projected response variables */
create projections from projections [colname=cname];
append from projections;

```

```
quit;

/* Append projected response variables to input dataset */
data &dataset;
  merge &dataset projections;

/* Sort the input dataset so that it is sorted exactly as it was upon input */
proc sort;
  by line;
data &dataset;
  set &dataset;
  drop line;
run;

%mend;
```