

International Journal of Education Policy & Leadership

TESTING THE TESTING: VALIDITY OF A STATE GROWTH MODEL

KIM TRASK BROWN

University of North Carolina - Asheville

Possible threats to the validity of North Carolina's accountability model used to predict academic growth were investigated in two ways: the state's regression equations were replicated but updated to utilize current testing data and not that from years past as in the state's current model; and the updated equations were expanded to include additional socio-economic, financial, and demographic variables. The updated equations were found to explain a very low proportion of the variance in growth statewide, with R^2 values ranging from .054 to .135. This suggests that the state's model is extremely unproductive of academic growth. The expanded equations were found to offer almost twice as much predictive power as the updated equations and, hence, the state's model.

Brown, K. T. (2008). Testing the Testing: Validity of a State Growth Model. *International Journal of Education Policy and Leadership* 3(6). Retrieved [DATE] from <http://www.ijepl.org>.

The state of North Carolina did not have far to go to align its existing testing program with No Child Left Behind (NCLB) guidelines. The North Carolina ABCs of Public Education Program (ABCs) was implemented in 1996 (NCDPI, 2002). The ABCs is a program that was designed to reorganize the public schools of North Carolina. According to the program's philosophy, the intended outcomes of the ABCs include the following:

- invoking strong accountability standards for students and teachers.
- ensuring an emphasis on the basics in the public schools, with a particular concentration on reading, writing, and mathematics, especially in the early grades.
- bringing about maximum control of education at the local level.

Consequently, "ABC" in the ABCs program title stands for "Accountability, Basics, and local Control" (NCDPI, 1997a).

The alignment of the ABCs to NCLB guidelines has brought about high stakes for North Carolina schools and students, with high-stakes tests defined as those with results that impact student placement, promotion, retention, graduation, diploma certification or endorsement, and access to higher education; the distribution of dollars to students or to schools; school accreditation,

certification, or ranking; and the issuance of school or district report cards (Maylone, 2002). In North Carolina, student academic proficiency is considered not only in terms of actual assessment scores but also in terms of growth with regard to assessment scores from one year to the next. A standardized growth model was created by the North Carolina Department of Public Instruction (NCDPI) and is used to calculate the expected growth of all North Carolina schools each year. Schools are now placed into an ABCs awards and recognition category each year based on student score growth on End-of-Grade (EOG) Tests in reading and math. A pattern of unmet growth expectations can lead to schools' being placed in one of a number of corrective action categories, leading to such sanctions as categorical labeling, publicly distributed and published progress reports, loss of funding flexibility, and takeovers in the name of ensuring student achievement. North Carolina students can be retained in their grade level and face the possibility of graduating with a Certificate of Attendance and not a diploma based on their individual EOG scores.

Consequently, student academic achievement and high-stakes test performance is affected by socioeconomic, financial, and demographic factors that are not within the control of schools. However, these factors are not taken into consideration in the NCDPI growth model.

Since the NCDPI's model utilizes student EOG scores as the basis for calculating school growth, if relationships exist between the above-mentioned factors and these growth scores, validity could be compromised. The purpose of this study was to investigate possible predictive relationships between these socioeconomic, financial, and demographic factors and EOG scores to shed light on whether the North Carolina model being used is valid for the purpose of accurately determining expected growth for all North Carolina schools, regardless of their population. Given what is at stake for students and schools due to federal and North Carolina state mandates, it seems reasonable that evidence concerning the validity of the model used to calculate the growth of North Carolina schools should be investigated. Additionally, although the state has provided some validity evidence for the EOGs themselves, no publicly available research exists with regard to the validity of the NCDPI's growth model.

High-Stakes Test Performance

The same factors discussed above, which correlate with student achievement in general, have also been found to impact student performance on high-stakes tests such as the EOGs.

Socioeconomic status. Student socioeconomic status (SES) has a profound impact on performance on high-stakes tests. It has been found that students with lower SES do not perform as well on high-stakes tests as their more affluent peers (Amrein & Berliner, 2002; Ediger, 2000; Holman, 1995; Madaus & Clarke, 2001; Reeves & Grubb, 1997). Given this, it stands to reason that with today's education initiatives at state and federal levels that rely so heavily on the use of high-stakes testing, students who are economically disadvantaged are being negatively affected throughout their school career and future. Some suggest that if high-stakes testing continues to be the vehicle through which lawmakers at all levels make decisions about students, SES should be controlled for as part of the accountability formula, in order to put all students on a more even level (Madaus & Clarke, 2001; Reeves & Grubb, 1997). A very clear case is made for taking positive steps toward using high-stakes testing results as a means for determining ways to meet the needs of lower SES students instead of for simply imposing consequences upon them (Holman, 1995; Madaus & Clarke, 2001; Reeves & Grubb, 1997).

In addition to personal socioeconomic factors, community financial factors must also be considered with

regard to student achievement on high-stakes tests. In fact, it has been found that community financial factors can have an influence on students despite their individual and family characteristics (Holloway, 2004). In particular, researchers have reported the following findings:

- There is a correlation between per-pupil expenditures and scores on standardized high-stakes tests, including the American College Test and the SAT Reasoning Test (SAT) (Bolon, 2001; Herring, 1997; Richardson, 2000).
- High district wealth, hence high per capita income, is associated with high student performance (Caldas, 1999; Clore, 1991; Maylone, 2002).
- School district resources and financial appropriations influence the quality of teachers and types of students attracted to school systems and, therefore, impact academic achievement (Caldas, 1999; Schiller, Khmelkov, & Wang, 2002).

Gender. Another widely studied topic is the difference between the academic achievements of the two genders on high-stakes tests. Research shows that males have a slight advantage over females in math but that this difference is decreasing and close to nonexistent (ERIC Clearinghouse on Urban Education, 2001; University of Minnesota, 2002). However, it has been found consistently that males score lower on assessments of reading, writing, and language skills (Cole, 1997; ERIC Clearinghouse on Urban Education, 2001; Freedman, 2003; University of Minnesota, 2002).

Overall, researchers indicate that the performance difference that does exist between males and females on high-stakes tests is mostly very small (Kleinfeld, 1998; Willingham & Cole, 1997). However, the research frequently suggests that even these small differences have a substantial impact on the future for students. For example, the SAT is the test most commonly used by colleges and universities to make decisions about college admission because it is purported to be a good indicator of the performance of first-year college students. Males tend to score much higher on this test, thus obtaining more college admissions and scholarships. However, females far surpass males in terms of grades during the first year of college (Kleinfeld, 1998; National Coalition for Women and Girls in Education, 1997). This is just part of a widespread finding by researchers that claims that greater variability in scores on high-stakes tests works much more to the advantage of males and especially those scor-

ing at the top (Kleinfeld, 1998; Willingham & Cole, 1997). The NCDPI reports that on the reading EOGs, females score at least two scale points above males at each grade level (NCDPI, 1997b), and on the math EOGs, females score a half point to one point above males in all grade levels (NCDPI, 1997a).

Race. Student race is also a very clear predictor of performance on high-stakes tests. In general, it has been found that black and Hispanic students score lower on high-stakes tests than do white and Asian students (Amrein & Berliner, 2002; Biemesderfer, 2001; Kober, 2001; Madaus & Clarke, 2001). The NCDPI indicates that on the reading EOGs, black students perform one grade level below white students (NCDPI, 1997b), and on the math EOGs, black students perform one and one-quarter grade levels below white students (NCDPI, 1997a).

Parent Education Level. It has been found that scores on high-stakes tests are affected by such home factors as parent education level (Livingston & Livingston, 2003; Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998). In fact, parent education level is thought to be the most crucial family indicator influencing student academic performance on high-stakes tests, according to Holman (1995).

Student Academic Growth

Evidence does exist which points to a relationship between the socioeconomic, financial, and demographic factors defined in the previous section and student academic growth. However, as shown in the following paragraph, research on this topic was found to be limited.

Muthen and Siek-toon (1998) advocate for researchers to include student background variables in the data set when growth modeling is performed. They assert that this background information provides for a more accurate prediction of individual growth. Consequently, a growth curve model is widely recognized as an appropriate tool to use for considering differences when studying individual growth (Rogosa, 1987; Stone & Lane, 2003; Willet & Sayer, 1994). This is important because researchers have cited the following findings resulting from large-scale studies of student growth:

- Schools with a higher SES demonstrated a higher rate of growth than those with a lower SES on the high-stakes testing program in the state of Maryland (Stone & Lane, 2003).

- Marked gender differences occurred with regard to the growth rate in math given a study of scores on the standardized National Assessment of Educational Progress (NAEP), with females having a greater mean growth score (Muthen & Siek-toon, 1998).
- Growth rates in reading and math for minority students were found to be lower than those for white students in a study utilizing data from the statewide testing programs in 43 states (Olson, 2005).
- Higher levels of education obtained by the mother of students were found to have a positive influence on student growth rates in the Muthen and Siek-toon (1998) study.

Method

NCDPI's Prediction Model

A strong element of the ABCs is accountability with regard to the performance of individual schools in reading and math, as measured by the growth of each school's students, taken as cohorts, over time. The School-Based Management and Accountability Program was developed by the NCDPI as a means for holding schools accountable for this growth, with the use of the formulas described below. With these formulas, the expected growth of each cohort of students is calculated and compared with its actual growth to determine if it reached acceptable growth levels. The figures used in the growth calculations come from student raw scores on the EOGs in reading and math converted to developmental scale scores. The regression equation used by the NCDPI for calculating the growth of schools with regard to student EOG scores from one year to the next is as follows, where EG is expected growth:

$$EG = b_0 + (b_1 \times ITP) + (b_2 \times IRM)$$

In this equation,

- b_0 = The North Carolina Average Rate of Growth. This figure is a coefficient generated from regression analysis using school-level growth as the dependent variable. For grades 4 through 8, growth is defined as a school's average scale score in one year (for a given grade level and subject) minus the average scale score from the same cohort of students the previous year. For grade 3, growth is calculated by taking the average score on the 3rd grade EOG (given

in the spring) and subtracting the average score on the 3rd grade EOG pretest (given in the fall), since there is not an EOG test in 2nd grade. The prediction formulas used by the state through the 2004–05 school year for determining schools' expected growth were generated using data from the 1992–93 and 1993–94 school years for grades 4 through 8 and from the 2000–01 school year for grade 3.

- b_1 = A value used to weight the Index of True Proficiency (see below) of students. Coefficients differ at each grade level and for reading and math (i.e., separate prediction models were generated for each of these). The models currently used by the state to determine expected growth were derived using data from the year 2000–01 for grade 3 and from the year 1994–95 for grades 4 through 8.

- ITP = The Index of True Proficiency. The formula for establishing ITP is (average scale score in reading for the school from year one + average scale score in math for the school from year one) – (average scale score in reading for the state + average scale score in math for the state). In this formula, average scale score in reading for the state and average scale score in math for the state are coefficients using the average scale scores for the state in the respective subjects from the year 2000–01 for grade 3 and from the year 1994–95 for grades 4 through 8.

- b_2 = A value used to weight the influence of regression to the mean of scores. The coefficients currently used by the state to represent b_2 were derived using data from the year 2000–01 for grade 3 and from the year 1994–95 for grades 4 through 8.

- IRM = The Index of Regression to the Mean. The formula for establishing a school's IRM in reading is (average scale score in reading for the school for year one – average scale score in reading for the state). The formula for establishing IRM in math is (average scale score in math for the school for year one – average scale score in math for the state). (NCDPI, 2004)

Analysis for This Study

Step 1: Updating of the state's prediction model. As stated above, coefficients used in the state's growth formula are

from years past. The first step in the analysis for this study was to use 2002–03 and 2003–04 data, the most current years available, to reestimate the coefficients that yield the best prediction of growth based on current data. This was done based on the assumption drawn from the review of literature that as societal issues change, so too must student high-stakes test performance.

Expected growth models were generated for each public noncharter school (1,186 elementary and 501 middle) in North Carolina for both reading and math to determine the growth students would be expected to experience between their 3rd grade year in 2002–03 (year one) and their 4th grade year in 2003–04 (year two), and also between their 6th grade year in 2002–03 (year one) and their 7th grade year in 2003–04 (year two). These models used independent variables identical to those used in the state's model, with the exception of their being taken from current school years as described above.

Step 2: Expanding the updated state prediction model. In this step, I generated an expanded regression model using additional predictor variables that I believe affect student high-stakes test performance. These variables included school socioeconomic level, school per-pupil expenditures, school community per capita income, school community county appropriations, school gender percentages, school racial percentages, and parents' education level.

All of these additional variables, with the exception of school gender percentages and school racial percentages, are, in a sense, proxies for the larger construct that is usually called SES. The percentage of students who receive free or reduced lunch is the most commonly used variable to represent the SES of a school. The other variables listed above are also limited but do represent some aspect of SES, as they are a reflection of social status or economic position. As discussed above, although per-pupil expenditures, per capita income, and county appropriations are not direct measures of the students' SES, they do in a sense relate to the SES of the school as a whole. Therefore, multiple measures were used to strengthen the research.

In developing the expanded regression model, the updated state expected growth model, as described in Step 1, was used as a basis. The expanded model included the state's independent variables (ITP and IRM), but

the additional predictor variables described above were placed into regression models for both reading and math.

The total R^2 for the expanded model was compared to the R^2 for the model generated using only the updated elements in the state's model (i.e., the model generated in Step 1). This was done to demonstrate the extent to which the prediction of expected growth was improved when the additional variables were added to the model.

Results

Step 1: Updating of the State's Prediction Model

The original state regression equations for the elementary grade data are as follows, where EG is expected growth:

Reading:

$$EG = 5.200 + (.220 \times ITP) + (-.600 \times IRM)$$

Math:

$$EG = 7.300 + (.260 \times ITP) + (-.580 \times IRM)$$

The updated regression equations obtained in this study for the elementary grade data are as follows:

Reading:

$$EG = 4.485 + (.166 \times ITP) + (-.445 \times IRM)$$

Math:

$$EG = 5.888 + (.319 \times ITP) + (-.702 \times IRM)$$

Table 1 (page 9) illustrates the regression analysis results for the updated equation regarding the elementary reading data, and Table 2 (page 9) illustrates the regression analysis results for the updated equation regarding the elementary math data.

The original state regression equations for the middle grade data are as follows:

Reading:

$$EG = 3.300 + (.220 \times ITP) + (-.600 \times IRM)$$

Math:

$$EG = 6.500 + (.260 \times ITP) + (-.580 \times IRM)$$

The updated regression equations obtained in this study for the middle grade data are as follows:

Reading:

$$EG = 2.661 + (.157 \times ITP) + (-.390 \times IRM)$$

Math:

$$EG = 2.788 + (.274 \times ITP) + (-.466 \times IRM)$$

Table 3 (page 9) illustrates the regression analysis results for the updated equation regarding the middle grade reading data, and Table 4 (page 10) illustrates the regression analysis results for the updated equation regarding the middle grade math data.

Step 2: Expanding the Updated State Prediction Model

Table 5 (page 10) illustrates the regression analysis results for the expanded equation regarding the elementary reading data, and Table 6 (page 11) illustrates the regression analysis results for the expanded equation regarding the elementary math data.

Table 7 (page 11) illustrates the regression analysis results for the expanded equation regarding the middle grade reading data, and Table 8 (page 12) illustrates the regression analysis results for the expanded equation regarding the middle grade math data.

Summary of Results

The following includes the most important points that arise from the results of this study.

- With regard to the results in Step 1 (i.e., the updated model), it is important to note that the coefficients are different from those in the state's model, which includes coefficients from several years past. Additionally, the independent variables in each regression analysis in the updated model are all significant, even though the R^2 and adjusted R^2 values are small for each analysis.
- The small R^2 values for both the updated model, and the results from Step 2 (i.e., the expanded model) show that they are not strong predictors of expected growth. However, it should be noted that the expanded model offers almost twice as much predictive power as does the updated model.

Discussion

Most Significant Findings

The most noteworthy findings of this study, which incidentally were unanticipated, are the extremely small R^2 values that resulted from the regression analyses of the updated model. These mean that the four updated equations explain only 13.5 percent, 5.6 percent, 8 percent, and 5.4 percent of the variance in school growth on their respective tests. Clearly, this indicates that the amount of variance in growth that is explained by the model is alarmingly low. All of this suggests that the state's model

is extremely unproductive of student growth on the EOGs, as the state's model was used as the basis for creating the updated model for this study. This is an alarming finding because, under the ABCs, schools are rewarded or sanctioned based on comparisons of their actual growth to expected growth values that are generated by the state's model.

There are possible explanations for the low R^2 values resulting from the regression analyses of the updated model. One is that there were some differences between the way growth was computed in this study and the way it is computed by the state for the ABCs. In this study, growth was computed by subtracting the year one average scale scores from the year two average scale scores for each school and both subject areas. The NCDPI's growth calculations include data from charter schools, but charter school data were not included in this study because charter school scale score data for the years examined were not uniformly available. Also, the NCDPI includes alternative assessment scores in growth calculations, but alternative assessment scores were not included in the analyses for this study because these scores are not reported. Finally, students who are not in attendance at a school for 91 days are not counted in the NCDPI's growth calculations, nor are students who do not have an EOG score from the previous year. Since this study utilized data available from individual schools, it is likely that scores for students in these categories were included in the analyses. This explanation for the low R^2 values seems improbable, however. In a very large sample, such as the 1,687 used in this study, these factors would be unlikely to affect the outcome dramatically enough to warrant a different conclusion about the predictive power of the model.

A second explanation is that the state's accountability model is simply no longer appropriate for predicting growth—assuming it ever was appropriate. The data may have drifted during the time between when the state's coefficients were set and the years included in this study. In the intervening years, societal, curricular, or other changes may have occurred that affected test scores and rendered the previous model invalid. Since R^2 values for the state's original model do not seem to be available to the public, it is not known how well the model fits the original data set from which it was generated.

Another significant finding to note is that, at each grade level, the expanded model offers almost twice as much predictive power than does the updated model. Even if the above-described differences in and explana-

tions for calculation of growth had a major impact on the outcome of the analyses of this study and, for instance, quadrupled the R^2 values, the resulting R^2 values would still be astonishingly low. Because the expanded model is created from the updated model, it stands to reason that the R^2 values resulting from its analysis would be low, but it is astonishing to see that the expanded model is almost twice as good at predicting student growth than is the state's model.

Other Important Findings

The NCDPI uses a growth model as part of the ABCs that operates by comparing students to themselves. This raises the question as to whether the rate at which students grow, and hence the amount of growth made in a year, relates to the collective characteristics of the students in a school. A few studies have addressed this. Stone and Lane (2003) found that on the Maryland state high-stakes assessments, schools with a higher SES were found to have a higher rate of growth. Also, Muthen and Siek-toon (1998) uncovered that females obtained a greater mean growth score than did males in their study of scores on the NAEP, which speaks to possible negative implications for schools with higher concentrations of males. Additionally, the data from statewide testing programs in 43 states indicated that white students experienced higher growth rates than did minority students (Olson, 2005). Therefore, schools with high percentages of minority students may experience lower growth rates overall. Finally, Muthen and Siek-toon (1998) found that the higher the education level attained by the mother of students, the more positively student growth was affected. This affects schools in communities where parent education levels are low.

Additionally, the beta weights of the individual additional variables, found as a result of the regression analyses of the expanded model, were relatively small. However, it is important to remember that even though the beta weights were small, several of the variables were found to be statistically significant. In particular, socioeconomic level, per-pupil expenditures, percent male, and percent white were statistically significant most often. The beta weights provide the *unique* relative contribution of each variable. Thus, though the betas for most individual variables were small, their combined contribution may be considerable. Therefore, it can be concluded that these particular additional variables did contribute somewhat to the prediction of student growth, and this is consistent with the literature.

Implications for Practice

Implications for validity. Construct validity relates to whether the construct being measured by a test is what is purported to be measured and combines aspects of both content and criterion-related validity. Regarding the ABCs, the constructs of interest are reading and math achievement and growth as evidenced by EOG scores. Therefore, the construct validity evidence needed regarding the ABCs would be that which shows that what is measured by the EOGs is, in fact, reading and math achievement and growth. The focus of this study was not to question whether the EOGs are actually measuring reading and math achievement and growth. Rather, this study looked at whether the NCDPI's growth model predicts reading and math growth *equivalently* for different schools given their status with regard to the additional variables considered in this study.

It is interesting to note Oosterhof (1994), who states that construct validity can include evidence that irrelevant qualities are not being measured by a test. In the case of the ABCs, in order to establish construct validity, evidence should exist to show that performance and growth on the EOGs is not impacted by factors such as those presented as additional variables in this study. Existing EOG validity verification presented by the NCDPI does not include this type of construct-irrelevant variance evidence. This is important to understand, as this study found that EOG scores vary at least partly because of these irrelevant factors, rather than just because of students' abilities in the trait being measured by the EOGs.

Another significant point is that a measure that is valid for measuring student achievement may or may not also be valid for measuring whether or not *schools* are doing a good job. With the EOGs, two different constructs are said to be considered. The first of these is student achievement, and the other is school growth. Since the inception of the ABCs, however, the primary purpose of the EOGs has been to measure school growth. Viewed in this light, it is clear that the only construct measured by the EOGs is school growth. Therefore, construct validity evidence is needed for that construct. To establish this evidence, the accountability model that uses the growth prediction formula must be studied. Additionally, doing so necessitates that school-level data, with the school as the unit of analysis, be observed. It is important to note that the current study ensured that, in measuring school growth performance, the EOGs were not measuring factors that should have been irrelevant, such as those represented by the additional variables used in this study.

I can conclude that if the consideration of the additional variables used in this study had played a huge role with regard to predicting student growth, then this could threaten construct validity. Since the consideration of these additional variables was not found to contribute considerably to growth predictions, I can assume that construct validity was not threatened. Overall I can say that, although the conclusion resulting from this study that construct validity was not threatened due to socioeconomic, financial, or demographic factors cannot necessarily be proven, it is acceptable to consider it fact since evidence to the contrary is not present. In other words, it is appropriate to conclude that the model used by the NCDPI to predict student growth does adequately control for the additional variables since it is free of construct-irrelevant variance.

Implications for educational policy. The results of this study suggest that North Carolina has been implementing an accountability program based on a statistical model that is ineffective for predicting school growth. Though from this study nothing about how well the model predicted growth for school years other than 2003–04 can be concluded, it seems unlikely that this was the first year for which the model did not fit the actual data.

Ironically, during the course of the final stages of this research, the North Carolina State Board of Education conducted an evaluation of the validity of the ABCs, in accordance with instructions from the North Carolina General Assembly. The result of this evaluation was that the following points were recognized:

- Changes in growth expectations were necessary given test changes that resulted from curriculum revisions.
- Statewide growth figures showed a “saw-toothed” pattern with regard to the percent of schools meeting and/or exceeding growth goals. This means that the percent of schools meeting and/or exceeding growth goals fluctuated randomly from year to year.

Given the above information, it was recommended by the North Carolina State Board of Education that a new measure for student growth be developed that:

- uses past student performance as a starting point,
- eliminates the aforementioned “saw-tooth” pattern,
- accounts for changes in student performance due to curriculum revision,

- is accurate despite the use of differing test editions, and
- provides valid and reliable results. (NCDPI, 2005)

Interestingly enough, these points are exactly those pointed to as a result of this research. Thus, the state's decision to change the model provides further support for this study's conclusion that the accountability model is not predictive of school growth.

Test developers and policymakers in North Carolina, and in states where similar test results are found or growth models are used as part of the testing system, would benefit from looking at the results of this study—particularly at the benefit of accounting for individual student socioeconomic, financial, and demographic variables as a means for determining true academic growth. This is also true for federal education policymakers as they are considering recommendations related to the reauthorization of NCLB and, as a result, are considering espousing a testing system containing growth models. If such a growth model type of testing system is adopted as a result of the reauthorization of NCLB, policymakers and test developers in each state will need to look at growth models and the data available concerning them. As such, the implications of this study regarding individual student characteristics and their impact on academic growth must be noted. Along these lines, said policymakers and test developers should also consider other data that came to light as a result of this study, such as the benefits of the use of a growth curve model, as recommended by Stone and Lane (2003), Willet and Sayer (1994), and Rogosa (1987), and adjusting testing starting points to allow for individual student characteristics that affect growth, as suggested by Muthen and Siektoon (1998).

Interestingly enough, the findings of this research are counter to those of William Sanders, a leading contributor to the academic growth debate occurring in the United States. In essence, Sanders purports that student academic growth is affected by teacher effectiveness above and beyond any other factors, including individual student socioeconomic, financial, and demographic variables. In fact, Sanders calls teacher effectiveness “the single biggest factor influencing gains in achievement, an influence many times greater than poverty or per-pupil expenditures” (quoted in Hill, 2000, p. 46). Hill believes Sanders's statement “challenges long-held assumption about the influence of a child's socioeconomic background on his or her learning” (Hill, 2000, p. 46). Because many policymakers and test developers at the federal level are looking to Sanders and his research, it is

imperative that the results of this study, and others like it, be considered as a basis for questioning Sanders's findings and effectively creating testing mandates.

Suggestions for Future Research

As a result of this study, several suggestions for future research came to light. These are as follows:

- Replicate the study and include data for more grade levels.
- Replicate the study using technical considerations, as the state does.
- Replicate the study using data from several years back in an effort to see how long the state has been using a model that predicts next to nothing.
- Expand the current study to determine how many North Carolina schools would have been placed into different ABCs awards and recognition categories if the updated and/or expanded model had been used.
- Conduct policy studies regarding the impetus and outcomes of the aforementioned North Carolina State Board of Education evaluation of the validity of the ABCs.
- Investigate what political or other factors facilitate or impede investigations into the validity of accountability models.

Conclusion

Some surprising findings were uncovered as a result of this study. The state's accountability model was discovered to be virtually invalid. This is true since it was found that the updated equations (in effect, replicas of those contained in the model used by the state through the 2004–05 school year) were essentially useless in predicting school growth on the EOGs. Additionally, it was found that the expanded equations created with the consideration of individual student socioeconomic, financial, and demographic variables offered almost twice as much predictive power as the updated equations and, hence, the state's model.

Many implications regarding validity and educational policy were realized in light of the findings of this study, despite the acknowledged limitations. The merit of the type of research undertaken by this study is realized by many who claim that researchers from different specialty areas should make it their goal to study high-stakes test results, such as those resulting from NCLB requirements, in the interest of providing information to policymakers and other professionals.

Table 1. Summary of Regression Analysis for Updated Elementary Grade Reading Growth Data			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	4.485**	.041	--
Index of True Proficiency	.166**	.041	.631
Reading Regression to M	-.445**	.072	-.972
Note. $R^2 = .135$; adjusted $R^2 = .133$. ** $p < .001$.			

Table 2. Summary of Regression Analysis for Updated Elementary Grade Math Growth Data			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	5.888**	.048	--
Index of True Proficiency	.319**	.038	1.070
Math Regression to M	-.702**	.085	-1.046
Note. $R^2 = .056$; adjusted $R^2 = .055$. ** $p < .001$.			

Table 3. Summary of Regression Analysis for Updated Middle Grade Reading Growth Data			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	2.661**	.053	--
Index of True Proficiency	.157**	.032	.866
Reading Regression to M	-.390**	.067	-1.030
Note. $R^2 = .080$; adjusted $R^2 = .076$. ** $p < .001$.			

Table 4. Summary of Regression Analysis for Updated Middle Grade Math Growth Data			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	2.788**	.075	--
Index of True Proficiency	.274**	.052	1.086
Math Regression to M	-.466**	.095	-1.015
Note. $R^2 = .054$; adjusted $R^2 = .050$. ** $p < .001$.			

Table 5. Summary of Regression Analysis for Expanded Growth Model for Elementary Grade Reading			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	5.753***	.848	--
Index of True Proficiency	.123**	.040	.470
Socioeconomic Level	-.016***	.003	-.243
Per-Pupil Expenditures	.000***	.000	.117
Per Capita Income	.000	.000	.020
County Appropriations	.000	.000	-.035
Percent Male	-4.173**	1.368	-0.079
Percent White	.769**	.231	.147
Percent College Graduate	.563*	.221	.089
Reading Regression to M	-.519***	.070	-1.136
Note. $R^2 = .215$; adjusted $R^2 = .209$. * $p < .05$. ** $p < .01$. *** $p < .001$.			

Table 6. Summary of Regression Analysis for Expanded Growth Model for Elementary Grade Math			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	4.924**	1.026	--
Index of True Proficiency	.223**	.040	.746
Socioeconomic Level	-.008*	.004	-.109
Per-Pupil Expenditures	.000**	.000	.152
Per Capita Income	.000	.000	.069
County Appropriations	.000*	.000	-.109
Percent Male	-2.207	1.657	-.037
Percent White	1.207**	.280	.203
Percent College Graduate	.084	.267	.012
Math Regression to M	-.627**	.085	-.935
Note. $R^2 = .113$; adjusted $R^2 = .107$. * $p < .05$. ** $p < .001$.			

Table 7. Summary of Regression Analysis for Expanded Growth Model for Middle Grade Reading			
Variable	<i>B</i>	<i>SE</i>	β
(Constant)	5.069**	.956	--
Index of True Proficiency	.136**	.032	.749
Socioeconomic Level	-.008	.004	-.139
Per-Pupil Expenditures	.000	.000	.064
Per Capita Income	.000*	.000	-.164
County Appropriations	.000	.000	.087
Percent Male	-5.086**	1.300	-.171
Percent White	1.323**	.297	.291
Percent College Graduate	.120	.211	.026
Reading Regression to M	-.468**	.065	-1.234
Note. $R^2 = .186$; adjusted $R^2 = .171$. * $p < .05$. ** $p < .001$.			

Table 8. Summary of Regression Analysis for Expanded Growth Model for Middle Grade Math			
Variable	B	SE	β
(Constant)	5.343***	1.398	--
Index of True Proficiency	.194**	.056	.770
Socioeconomic Level	-.023**	.007	-.276
Per-Pupil Expenditures	.000*	.000	.104
Per Capita Income	.000	.000	-.054
County Appropriations	.000	.000	-.008
Percent Male	-4.527*	1.902	-.109
Percent White	.230	.435	.036
Percent College Graduate	.137	.308	.022
Math Regression to M	-.434***	.095	-.946
Note. $R^2 = .101$; adjusted $R^2 = .084$. * $p < .05$. ** $p < .01$. *** $p < .001$.			

References

- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved July 11, 2002, from <http://epaa.asu.edu/epaa/v10n18/>
- Biemesderfer, S. (2001). *Building on what we have learned: A national dialogue on standards-based education* (Contract Nos. ED-01-CO-0011 & ED-01-CO-0006). Washington, DC: Office of Educational Research and Improvement.
- Bolon, C. (2001). Significance of test-based ratings for metropolitan Boston schools. *Education Policy Analysis Archives*, 9(42). Retrieved January 29, 2002, from <http://epaa.asu.edu/epaa/v9n42>
- Caldas, S. J. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *Journal of Educational Research*, 93(2), 91-100.
- Clore, W. P. (1991). The relationship of superintendent instructional leadership behavior and school district demographics to student achievement (Doctoral dissertation, The University of Texas at Austin, 1991). *Dissertation Abstracts International*, 52, 04A.
- Cole, N. S. (1997). *The ETS gender study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.
- Ediger, M. (2000). *High stakes testing and student achievement*. (Report No. TM032082). Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED447193)
- ERIC Clearinghouse on Urban Education. (2001). *Gender differences in educational achievement within racial and ethnic groups* (Report No. EDO-UD-01-3). Washington, DC: Office of Educational Research and Improvement.
- Freedman, B. (2003, April 21). *Boys and literacy: Why boys? Which boys? Why now?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Herring, D. S. (1997). The relationship between education achievement and educational variables when controlling for demographics (Doctoral dissertation, The University of Southern Mississippi, 1997). *Dissertation Abstracts International*, 58, 08A.

- Hill, D. (2000). He's got your number. *Teacher Magazine*, 11(8), 42–47.
- Holloway, J. H. (2004). Research link: How the community influences achievement. *Educational Leadership*, 61(8), 89–90.
- Holman, L. J. (1995, April 21). *Impact of ethnicity, class, and gender on achievement of border area students on a high-stakes examination*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Kleinfeld, J. (1998). Why smart people believe that schools shortchange girls: What you see when you live in a tail. *Gender Issues*, 16(1/2), 47–63.
- Kober, N. (2001). *It takes more than testing: Closing the achievement gap*. A report of the Center on Education Policy (Report No. UD034289). Washington, DC: Center on Education Policy. (ERIC Document Reproduction Service No. ED454358).
- Livingston, D. R., & Livingston, S. (2003, March). *Unbraiding the rhetoric about student achievement and teacher quality in Georgia*. Paper presented at the annual meeting of the Georgia Association of Teacher Educators, St. Simons Island, GA.
- Madaus, G., & Clarke, M. (2001). The impact of high-stakes testing on minority students. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high stakes testing in public education* (pp. 85–106). New York: Century Foundation Press.
- Maylone, N. J. (2002). The relationship of socioeconomic factors and district scores on the Michigan educational assessment program tests: An analysis. Retrieved July 24, 2003, from <http://www.pdkintl.org/edres/ddwind5.htm>
- Muthen, B., & Siek-toon, K. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences*, 10(2), 73–101.
- National Coalition for Women and Girls in Education. (1997). *Title IX at 25: Report card on gender equity* (Report No. EA028430). Washington, DC: National Women's Law Center.
- North Carolina Department of Public Instruction [NCDPI]. (1997a). Research on end-of-grade testing: Assessment of mathematics. *Assessment Brief*, 3(4), 1–2.
- North Carolina Department of Public Instruction [NCDPI]. (1997b). Research update on end-of-grade testing: Assessment of reading comprehension. *Assessment Brief*, 3(2), 1–2.
- North Carolina Department of Public Instruction [NCDPI]. (2002). Testing started with the ABCs and other myths about testing and accountability in North Carolina. Retrieved July 11, 2002, from <http://www.ncpublicschools.org/parents/myths.html>
- North Carolina Department of Public Instruction [NCDPI]. (2004). Accountability brief. Retrieved April 14, 2008, from <http://www.ncpublicschools.org/docs/accountability/reporting/2004memo/standards2004.pdf>
- North Carolina Department of Public Instruction [NCDPI]. (2005). The ABCs model for 2005–06. Retrieved May 1, 2005, from http://www.ncpublicschools.org/docs/accountability/reporting/growth/th_eabcsmodelfor2005.pdf
- Olson, L. (2005). Achievement up, growth rate down. *Education Week*, 24(32), 10–14.
- Oosterhof, A. (1994). *Classroom applications of educational measurement* (2nd ed.). New York: Macmillan.
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 103–145). Washington, DC: Brookings Press.
- Reeves, E., & Grubb, D. (1997, November). *Spatial analysis of Kentucky assessment results, social and economic indicators*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis, TN.
- Richardson, S. E. (2000). A relationship study of instructional costs and student performance variables among public elementary and secondary schools of the United States (Doctoral dissertation, University of Arkansas, 2000). *Dissertation Abstracts International*, 61, 09A.
- Rogosa, D. R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics*, 12, 185–195.
- Schiller, K. S., Khmelkov, V. T., & Wang, X. Q. (2002). Economic development and the effects of family characteristics on mathematics achievement. *Journal of Marriage and the Family*, 64(3), 730–742.
- Stone, C., & Lane, S. (2003). Consequences of a state accountability program: Examining relationships between school performance gains and teacher, student, and school variables. *Applied Measurement in Education*, 16(1), 1–26.

- University of Minnesota. (2002). *The Minnesota basic skills test: Performance gaps for 1996 to 2001 on the reading and mathematics tests, by gender, ethnicity, limited English proficiency, individual education plans, and socio-economic status* (Report No. UD035348). Minneapolis, MN: Office of Educational Accountability.
- Willet, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of change. *Psychological Bulletin*, 116, 363–381.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.