Predicting Middle Level State Standardized Test Results Using Family and Community Demographic Data

Christopher H. Tienken
Seton Hall University
South Orange, NJ
christienken@gmail.com

Anthony Colella
Seton Hall University
South Orange, NJ

Christian Angelillo
Boonton Township School District
Boonton, NJ

Meredith Fox
Nanuet Union Free School District
Nanuet, NY

Kevin R. McCahill
George W. Miller Elementary School
Nanuet, NY

Adam Wolfe
Peoria Unified School District
Peoria, AZ

Abstract

The use of standardized test results to drive school administrator evaluations pervades education policymaking in more than 40 states. However, the results of state standardized tests are strongly influenced by non-school factors. The models of best fit ($n = 18$) from this correlational, explanatory, longitudinal study predicted accurately the percentage of middle school students scoring proficient or above on the New Jersey state-mandated standardized tests in mathematics and language arts for grades 6–8 during the years 2010, 2011, and 2012 for 70% to 78% of the schools in the statewide samples ($n = 292$ to 311), using only family and community demographic variables from the U.S. Census. Just three demographic variables, (a) percentage of families in a community with income over $200,000 a year, (b) percentage of people in a community in poverty, and (c) percentage of people in a community with bachelor’s degrees, predicted results accurately in
14/18 of the models. The findings suggest that state standardized test results are not as objective and transparent as advertised by state and federal department of education officials. Some middle level school administrators might be getting rewarded or punished based on factors that they do not influence.

**Keywords:** principal evaluation, high stakes testing, education reform, accountability, middle school assessment

The federally implemented Race to the Top (RTTT) grant program and the No Child Left Behind Act of 2001 (NCLB, 2002) waiver initiative increased the use of student results from state-mandated standardized tests of mathematics and language arts to evaluate the effectiveness of school administrators in most states. The RTTT competitive grant program, funded under the 2009 American Reinvestment and Recovery Act (ARRA), required states to submit applications for funding that were evaluated based on six broad categories of criteria. The Great Teachers and Leaders (GTL) category was worth the most credit in the application with 138 possible points (United States Department of Education, 2009, p. 3).

One subcategory of GTL, Improving Teacher and Principal Effectiveness Based on Performance, carried the most weight within the category and was worth 58 points (United States Department of Education, 2009, p. 3). “Students’ results from standardized state assessments to make determinations about principal effectiveness” was a key component of the category (United States Department of Education, 2009, p. 9). The student results from state-mandated tests aligned to college- and career-ready standards formed the basis for school administrator evaluations in many states, and these evaluations could trigger career-changing decisions about school administrators’ compensation, retention, promotion, tenure, and certification (United States Department of Education, 2009).

Similar conditions applied to public school administrators in states granted an NCLB waiver by the Secretary of Education under the waiver program. The parameters for teacher and principal evaluation found in NCLB waivers tracked closely to the criteria set forth in RTTT as many of the previously granted NCLB waiver policies used student results from state-mandated tests as the linchpin of school administrator accountability.

Although the rule changes in the Every Student Succeeds Act of 2015 (ESSA) renewal legislation nullified the Secretary of Education’s ability to grant waivers, new rules maintain a focus on school administrator effectiveness. State education agency officials have the latitude to continue with test-based school administrator accountability policies and may continue to do so in the foreseeable future.

The existing literature about evaluating middle level administrators based on student test results suggests the practice is tenuous at best. There exist a myriad of out-of-school factors that influence the lives of middle level students that require administrators to divide their time among initiatives aimed at cognitive, social, and emotional factors (NMSA, 2010; Wiles and Bondi, 1981). Standardized test results cannot capture the complexities of the influence middle level administrators may have on the lives of students.

**Common Practice**

Education officials and governors from more than 40 states essentially volunteered their public school students, parents, teachers, and school administrators to participate in various standardized testing programs that met the requirements set forth in the RTTT grant application and other state-developed accountability guidelines. The mandated tests must align to the Common Core State Standards (CCSS) (National Governors Association Center for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010) in mathematics and English language arts, or other state-adopted curriculum standards that conformed to the college and career readiness definitions and mandates set forth in RTTT.

According to the RTTT guidelines, former NCLB waiver requirements still in effect, and requirements set forth in ESSA, data generated from the student results on state-mandated assessments of math and English language arts will be used to evaluate school administrator performance in the states granted either NCLB waivers or RTTT grants and as part of ESSA (U.S. Department of Education [USDOE], 2015).

**The Study**

New Jersey was awarded a RTTT grant and was the recipient of an NCLB waiver. State officials decided to continue to evaluate school administrators based (in part) on the results from state standardized tests as a component of New Jersey’s ESSA application. Middle level administrators in New Jersey, like school administrators in approximately 40 other states, found themselves being...
judged in part by student test results from state-mandated tests of mathematics and language arts.

New Jersey is home to over 300 middle schools or schools that house middle level grades 6–8. Thus, middle level principals and assistant principals constitute a large portion of the school administrators working in New Jersey and form an interesting convenience sample from which to examine various influences of RTTT and ESSA evaluation guidelines on the evaluation of middle level administrator effectiveness.

**Research Problem**

We were struck by provisions embedded in two of the RTTT grant requirements related to the procedures that must be used to grant tenure and remove school administrators:

(c) Whether to grant tenure and/or full certification (where applicable) to teachers and principals using rigorous standards and streamlined, transparent, and fair procedures; and (d) Removing ineffective tenured and untenured teachers and principals after they have had ample opportunities to improve, and ensuring that such decisions are made using rigorous standards and streamlined, transparent, and fair procedures. (p. 9)

Furthermore, the language about principal evaluation embedded in ESSA seems to have been influenced by the RTTT guidelines. For example, Section 2103, Local Use of Funds, subsection (a)(3)(A)(i) states that ESSA funds can be used by states for “developing or improving a rigorous, transparent, and fair evaluation and support system for teachers, principals, or other school leaders that is based in part on evidence of student achievement.”

We took interest in the fact that the RTTT guidelines and ESSA language included the requirement for school administrator evaluation procedures that are transparent and fair. The requirements for transparent and fair procedures conflict with some results from existing literature that call into question the use of standardized test results to make important decisions about children and educators (e.g., Nichols & Berliner, 2007). Specifically, results from previous predictive studies suggest that the percentage of students in a school or district who will score proficient or above on state standardized tests in language arts and mathematics at the district level can be predicted, with a good deal of accuracy, by using only community- and family-level demographic variables found in U.S. census data (e.g., Darnell, 2015; Maylone, 2002; Sackey, 2014; Tienken, 2016).

Results from most of the previous quantitative studies that used algorithms to predict the percentages of students who scored proficient or above on state tests were cross-sectional in nature, explaining only one year in time, and looked at the district-level data (e.g., Sackey, 2014). Although district-level predictive studies provide important insights into how non-school factors can reflect on ratings and rankings of teachers, school administrators, and school districts, they have limitations. Such studies do not provide insights into how predictable state test results are at the individual school level over time or how non-school factors can influence evaluations of school administrator effectiveness.

There has been increased use of test results to judge school-level administrators, yet little is known about the influence of family and community demographic variables on the percentage of students who score proficient or above on state tests, at the school level of analysis, over the course of multiple years in the middle level grades. Middle level administrators constitute a substantial proportion of the public school administration corps in New Jersey and the United States, yet there is no quantitative, longitudinal literature since the inception of the CCSS that has directly addressed this issue at the middle level.

**Middle Level Context**

The middle grades are characterized by many social, emotional, and cognitive changes in children. According to information presented in *Turning Points 2000* (Jackson & Davis, 2000), middle level students undergo a host of physical and environmental experiences.

It is a time when young people experience puberty, when growth and development is more rapid than any other developmental stage except that of infancy. Dramatic physical changes are accompanied by the capacity to have sexual relations and reproduce. It is a time, too, of emotional peaks and valleys. (p. 7)

Given the focus on developmentally appropriate practices and social and emotional development found in the classic and recent middle level literature, it seems to us that middle level administrators, more so than administrators in elementary school and high school, must spend a considerable amount of
leadership time on tasks not directly related to academics, but no less important (Beane, 1990; Lounsbury, 1991; Mann, 2013).

The middle level administrator is expected to be more than a top-down manager who is single-mindedly focused on a narrow sliver of test-based content in the language arts and mathematics. In sum, the middle level administrator position is a multifaceted job that requires multifaceted success criteria, not a single standardized test score (NMSA, 2010; Wiles and Bondi, 1981).

Purpose and Questions

Our purpose for this study was to explain the accuracy of family and community demographic variables as predictors of the percentage of students who scored proficient or above, at the school level, on state-mandated tests of mathematics and language arts in grades six, seven, and eight over the three-year testing cycles of 2010 through 2012.

We guided the explanatory study with two questions: (1) How well do family- and community-level demographic variables, found in the 2010 U.S. census data, predict the percentage of students scoring proficient or above on New Jersey state tests of mathematics and language arts for the testing years 2010 through 2012? (2) How accurate a measure are state standardized test results, at the school level, of principal or assistant principal effectiveness?

Conceptual Framework

Family characteristics, such as wealth factors and family structure, and the characteristics of the community in which a student lives, such as the percentage of community members who are unemployed or the percentage of people in a community with a high school diploma, are important factors that can and have been used in past studies to predict academic achievement on standardized tests (e.g., Darnell, 2015; Maylone, 2002). In general, there exists a broad category of family demographic factors that influence and predict levels of achievement. Some factors are families headed by two parents, family income levels that exceed the federal poverty line, and a parent employed securely and full time (College Board, 2012; Davis-Kean, 2005; Dawson, 1991; Weinberg, 2001).

Some researchers use eligibility for free and reduced lunch as the proxy for socioeconomic status (SES) of individual students or a proxy for the overall SES of the students served by an entire school or school district. Sirin (2005) conducted one of the largest meta-analyses of the influence of SES on student outcomes and found that approximately 20% of the studies conducted between 1990 and 2000 used free and reduced lunch as the only variable to describe SES at the individual student, school, or school district levels.

Although student eligibility for free and reduced lunch is one variable that can help provide a general description of student SES or overall family human capital, Harwell and LaBeau (2010) presented an argument against using the variable of free and reduced lunch status as the sole indicator of student family capital, school SES, or school district SES. They suggested that student eligibility for free and reduced lunch is not as precise an indicator of overall student or community SES as some researchers might think. To overcome this potential limitation, we used multiple variables from the U.S. census data to create a multidimensional view of the family and community characteristics that influence student achievement on standardized tests.

Some researchers use the term “human capital” to refer to the broad collection of people’s skills, experiences, and abilities that allow them to potentially become more economically successful and act with greater skill (Becker, 1993; Coleman, 1988). In the context of our study, the student’s family is the closest level of human capital that the student experiences and that directly influences him or her. We synthesized from the existing literature that students who live in families with more human capital more frequently have access to academically-oriented life experiences, collateral learning experiences that extend school learning, and more supports to help them connect to and capitalize on academic content embedded in formal and collateral education opportunities (Scherrer, 2014). Thus, students with more family human capital often perform better on standardized tests of traditional academic achievement (Tienken, 2011). Family demographic factors, such as income, living in a single- versus a double-parent household, or the percentage of female households in poverty, are factors that help describe the level of human capital in a family. Those demographic factors can be used as proxies for the human capital experienced by the student.

The community in which a student lives also plays an important complimentory function, and in some ways
has a reciprocal relationship with the human capital of a family. We connect the community demographics to “social capital” as described by Coleman (1988):

Social capital is defined by its functions. It is not a single entity but a variety of different entities, with two elements in common: they all consist of some aspect of social structures, and they facilitate certain actions of actors—whether persons or corporate actors—within the structure. Like other forms of capital, social capital is productive, making possible the achievement of certain ends that in its absence would not be possible. . . . A given form of social capital that is valuable in facilitating certain actions may be useless or even harmful for others. (p. 98)

Formal and informal interactions and relationships within a community create social capital at the community level (Coleman, 1988; The World Bank, 2011). The types of professionals who live in a community, the community groups that exist for adults, structured community recreation programs for children, religious groups, libraries, services for senior citizens and disabled residents, arts commissions, local social advocacy groups, quality and affordable daycare and preschool opportunities, and other similar resources intersect to contribute to the overall social capital of a community (Becker, 1993; Putnam, 2000).

When children grow up in communities with access to high levels of social capital, it may increase the chances those children and their families will interact with and develop formal and informal relationships with people who have high levels of human capital. Therefore, children have the potential to be exposed to more academic ideas and life experiences that influence their learning in school directly and indirectly through the people that they interact with on a daily basis.

Children living in communities with higher levels of human and social capital are more likely to have access to varied life experiences that build academic background knowledge. They are more likely to come to school with existing academic knowledge that they can use to connect their life experiences to new content and effectively transfer ideas from school to other situations (Tanner & Tanner, 2007). Access to varied types and quality of social and human capital influences student learning in traditional classroom situations where children must connect the content of the classroom to their life experiences. Therefore, community social capital and family human capital may play a significant role in ultimate achievement on standardized tests (Scherrer, 2014).

Our connecting of family and community demographic variables to human and social capital as a tool to predict standardized tests results is situated within the broader theoretical framework of ecological systems theory, as described by Bronfenbrenner (1979). He posited that children exist in an ecological system, and various layers of the system exert influence upon them. Family, school, peer groups, and community intersect to directly and indirectly influence behaviors and outcomes of children.

Specifically, ecological systems theory provides support for our thinking because it suggests that key circles of family human capital and community social capital surround children. We conclude that the ecological systems theory helps explain how those circles of influence can hinder or encourage academic achievement and thus influence results on standardized tests. We hypothesized that a combination of the indicators related to family human capital and community social capital can predict student test results at the school level because the school is within the ecological system of children and is thus influenced by the other factors within the system.

Ecological systems theory also comports with research-based perspectives of poverty that suggest the importance of not only providing children the formal education resources necessary for learning at high levels, but also ensuring the appropriate social supports are in place so that children can make full use of the resources they encounter in formal learning environments. Known as the “capabilities perspective,” the line of research suggests students can have varying capabilities to “convert resources into their intended benefits” (Scherrer, 2014, p. 203). The realized output from the resources provided is influenced by the capability of the student to fully utilize the resources as intended.

Factors such as health, stable living situations in which basic needs are met, and access to academically oriented life experiences are necessary to support students connecting to formal curricular content; seeing the lessons learned in school being used in real world situations by others to whom they aspire; applying formal academic content in their everyday lives; and participating in life experiences that extend, enhance, and encourage more learning and increase a child’s capability to make the most of school (Kelly,
2010). According to Bronfenbrenner’s ecological systems theory, an important way one can gain an accurate understanding of a child is by considering the various layers of the system in which a child lives and is raised.

The capabilities perspective is another layer that influences learning, and it should be considered when analyzing student achievement patterns. When we view the issue of middle level administrator evaluation via test results through Bronfenbrenner’s ecological systems theory (EST) and the capabilities perspective, we are left with questions about how the criteria in RTTT, ESSA, and state-level accountability policies related to the use of standardized test scores to evaluate middle level administrators can be transparent and fair.

We proposed to use variables found in the U.S. census data that represent various aspects of human and social capital at the family and community levels, and the capabilities perspective embedded in EST to predict, within a margin of error, the percentage of students in New Jersey schools serving the middle levels who will score proficient or above on state-mandated assessments in English language arts and mathematics. We drew upon previous works on the topic (Darnell, 2015; Maylone, 2002; Sackey, 2014; Tienken, 2016; Turnamian, 2012) to guide our work.

Significance

We expanded upon previous studies on this topic in three specific areas. First, we constructed a series of hierarchical regression models that extended beyond the district level, to the individual school level, closer to students. Therefore, we could draw conclusions and recommendations related to individual schools (and hence, their principals and assistant principals) regarding how factors within the students’ ecological systems related to family human capital and community social capital acted upon student test results, and thus influenced the school administrators’ evaluation ratings. Earlier studies were unable to reach to this level of specificity with the school as the unit of analysis.

Second, whereas earlier studies were cross-sectional, using only one year of data, our study was longitudinal and covered three years of data for English language arts and mathematics. We used three years of data at the middle levels in grades sixth through eighth to explain the predictive accuracy of family and community demographic variables.

Finally, this was the first longitudinal study to use family and social capital factors to predict state test results at the middle level since the inception of the CCSS. The results from this study reflect a first look at this issue at this level of detail for middle level administrators.

Methodology

We used a correlational, longitudinal, explanatory design with quantitative methods to complete our study (Johnson, 2001). Such a design is appropriate when the research aims are to (a) identify relationships among independent and dependent variables, and (b) to explain and predict outcomes at one period in time or over time.

Variables

The percentages of students who scored proficient or above at the school level for each of the New Jersey mathematics and language arts state tests in grades six, seven, and eight during the 2010, 2011, and 2012 test administrations constituted the dependent variables. We located 18 independent variables in the census data consistently found in the extant literature related to family human capital and community social capital. We found eight variables related to family human capital and 10 variables related to community social capital (see Table 1).

Sample

The final samples for our study ranged from 292 to 311 schools serving the middle grades, depending on the number of schools that met the sampling requirements (see Table 2). The state of New Jersey consists of 21 counties with approximately 590 operating public school districts within those counties. The types of school categorizations in New Jersey include elementary schools, middle schools, comprehensive high schools, magnet schools, vocational schools, charter schools, and special education schools. The size and grade composition of schools within each district varies across the state. Some school districts house all students from pre-kindergarten (PK) to grade 12, whereas other school districts include only kindergarten through grade six or kindergarten through grade eight. Districts with PK–6 or K–8 do not have high schools within their districts. Instead, regional school districts house high schools that include students from various districts.

To ensure consistent matches between family and community demographic data and the characteristics of the students that attended the schools in grades six,
seven, and/or eight, the samples met the following criteria:

- Schools that only served grade six and/or grade seven and/or grade eight in one school building in the district during the years 2010, 2011, and 2012;
- Schools that serviced students within their home district only. Regional schools were excluded;
- Schools that were the only school in the district that served grades six, seven, and/or eight;
- Schools that had at least 25 students per middle level grade participate in the administration of the NJASK in English language arts and mathematics who received valid scores and whose town had complete U.S. census data for the 18 home and community variables included in this study.

Samples sizes varied somewhat due to violations of the sampling criteria. For example, in some years, some schools had fewer than 25 students in a grade level take a state test or receive valid data from the state. Small schools with small student populations are common in New Jersey, a state with almost 600 school districts. Variations in the number of schools within the same grade levels in our samples over different years are most often due to variations in student populations within some schools. The sampling criteria also excluded schools in large districts because there were multiple schools that serviced middle level students. Human and social capital demographic data from the community could not be matched precisely to different schools within the same district.

### Table 1

**Family Human Capital and Community Social Capital Independent Variables**

| Family human capital | Community social capital |
|----------------------|-------------------------|
| % families making less than $25,000 | % people employed |
| % families making less than $35,000 | % households making less than $25,000 |
| % families making more than $200,000 | % households making less than $35,000 |
| % families in poverty for 12 months | % households making more than $200,000 |
| % male only households, no females | % all people under poverty |
| % female only households, no males | % population with less than 9th grade education |
| % lone parent households (total) | % population with no high school diploma |
| % female households in poverty | % population with some college |
| % population with a bachelor’s degree | % population with an advanced degree |

### Table 2

**Number of Schools in Each Sample**

|        | Grade 6 ELA | Grade 6 M | Grade 7 ELA | Grade 7 M | Grade 8 ELA | Grade 8 M |
|--------|-------------|-----------|-------------|-----------|-------------|-----------|
| 2010   | 311         | 311       | 300         | 300       | 297         | 296       |
| 2011   | 308         | 308       | 298         | 299       | 294         | 294       |
| 2012   | 305         | 306       | 297         | 297       | 292         | 295       |
We conducted a priori sample size calculations to ensure the sample sizes were large enough to accommodate working with up to 18 variables. None of our final predictive models included all the variables. Field (2009) recommended the formula $50 + 8(k)$ for simultaneous multiple regression, with $k$ equaling the total number of predictor variables in the model, to determine an appropriate sample size to detect an effect size of at least .50 at the 95% confidence level and a $p$ value of at least .05. This study included 18 potential predictor variables to represent $k$. Using Field’s formula, we calculated $50 + 8(18) = 194$. Therefore, to reach an appropriate effect size and $p$ value equal to or less than .05, we must include at least 194 schools in the study at each grade level and subject. All our sample sizes exceeded the minimum sample sizes required.

Then, we calculated the required sample sizes for hierarchical multiple regression. Green (1991), as cited in Field (2009), recommended $104 + k$, where $k$ represents the number of predictor variables to be entered into the model for hierarchical multiple regression. Our sample sizes ranged from 292 to 311 schools and exceeded the minimum requirements for 122 cases to conduct hierarchical linear regression.

**Instrumentation**

We collected the dependent variables, the percentage of students proficient or above on the mathematics and language arts portions of the state tests from the New Jersey Assessment of Skills and Knowledge (NJASK) tests administered to grades six, seven, and eight during the 2010 through 2012 school years. State officials mandated the administration of NJASK as an operational assessment in the schools within our sample during the spring months of April and May during the 2010, 2011, and 2012 school years. The NJASK was the assessment New Jersey education officials used to measure student achievement and progress under the requirements of the NCLB Act prior to the first administration of the Partnership for the Assessment of Readiness for College and Careers (PARCC) assessment during the 2014–2015 school year.

State officials used the results from the NJASK as part of rating systems for teachers and school administrators. Teachers, administrators, schools, and school districts received ratings as “effective” or “ineffective” based partially on results from the NJASK tests. The NJASK test can be categorized as a high-stakes assessment, given how the results have been used to evaluate teachers and school administrators.

**Data Collection**

We located the data for the dependent variables from the New Jersey Department of Education’s website (NJDOE, 2015). The percentages of students rated proficient and advanced proficient were combined into one total percentage for each subject and year tested. Next, we matched the percentages to corresponding town demographic data from the U.S. Census for the communities served by each school that met our sampling requirements. We retrieved data for the family and community level variables from the American Community Survey section of the U.S. Census (2010) and localized the data with American Factfinder. Finally, we matched town demographic data from the U.S. Census to school assessment data.

**Data Analysis**

We used two forms of regression as the primary methods to analyze results for each subject area: simultaneous multiple regression (SMR) to narrow down variables and then hierarchical linear regression (HLR) to identify the most efficient statistically significant predictor variables and models of best fit for our predictive algorithms. Considering that the goal was to predict the aggregate performance of a school based on the best predictive model, hierarchical regression analyses was an appropriate strategy. Predictions require the identification of models of best fit, and hierarchical regression is an accepted method to determine such models (Field, 2013).

Prior to running the regressions, we inspected the skewness of the dependent variables to determine whether the data were normally distributed within the 1.00 to −1.00 ranges. All dependent variables except grade seven math 2011 met the assumption of normality. Grade seven math 2011 data had an initial negative skew of −1.167 due to two low, outlier percentages of 29 and 39. Both variables exceeded three standard deviations below the mean, and they met the definition of an outlier. In this case, we used the Winsorizing procedure to substitute the outlier scores with scores of 40 and 50, respectively, which (in our case) were one standard deviation higher than the original percentages (Field, 2013). Winsorizing “involves replacing outliers with the next highest score that is not an outlier” (Field, 2013, p. 198). Winsorizing resulted in normalizing the grade seven math 2011 data within acceptable limits of +/− 1.000.
We next conducted a series of layered analyses for each subject area in each year. First, we created correlation matrices and scatterplots to help develop more refined SMR and HLR models. We reviewed the correlation matrices for relationships between independent and dependent variables and among dependent variables to anticipate possible multicollinearity. We loaded the independent variables into an initial SMR model for each subject within each year. We began to remove variables from the model that were statistically insignificant above .10 or that exhibited initially high levels of multicollinearity above 7.000, then above 4.000, and finally above 3.000.

The process of removing variables was important because to obtain accurate measures of $R^2$, one must correctly identify predictor variables that correlate strongly with the dependent variable (Hinkle, Wiersma, & Jurs, 2003). We attempted to isolate predictor variables so the variance in the criterion could be accounted for only once to ensure that the predictor variables accounted for different proportions of the variance in the criterion variable. Therefore, we sought to refine each model so the predictor variables exhibited low correlations among themselves with variance inflation factors (VIF) below 3.000 (Hinkle et al., 2003). The process of factor elimination and substitution continued until we arrived at two to four predictor variables that maximized $R^2$-squared in each model.

We placed the statistically significant ($p \leq .05$) predictor variables in highest to lowest rank order based upon beta values to run the hierarchical regression models. The hierarchical regression models allowed us to identify how much influence each specific variable had on the dependent variable. We ran hierarchical models for all grades in both subjects and all three years, and sought the model of best fit in each case. The formal representation of our final regression equation for each model of best fit was $y_1 = b0 + (b1Xi) + (b2Xii) + (b3Xiii) + e$, with $b$ representing the unstandardized beta for the predictor variable, $X$ representing the percentage of the variable in the community, and $e$ representing the constant for each model (Field, 2013).

For example, the final predictive model of best fit for the 2010 grade six mathematics data for the Brookside School in the Allendale Boro School District included three variables that represented family human capital and community social capital, and most accurately predicted the percentage of students who achieved proficiency or above: (a) percentage of households in the community with annual income less than $35,000 [% HS<35K]$, (b) percentage of families in the community with annual income greater than $200,000 [%Fam>200K]$, and (c) percentage of people in the community with bachelor’s degrees [%BA]. We entered the values of those demographic predictors into the predictive algorithm with their unstandardized betas from the regression models, and the constant:

$$y_1 = 60.642[\text{constant}] + (0.653 * [\%\text{BA}]) + (0.194 * 37.6[\%\text{Fam} > 200K]) + (-0.463 * 8.6[\%\text{HS} < 35K]).$$

In this case, $y_1 = 89.90$.

The answer, 89.90, represented the percentage of grade six students at the Brookside School predicted to score proficient or above on the 2010 grade six New Jersey standardized mathematics assessment. The actual percentage of grade six students at the Brookside School who scored proficient or above on the New Jersey standardized mathematics assessment, as reported on the New Jersey State Report Card, was 89.20, a difference of .70%. The standard error of the estimate for the model was 9.60. The standard error of the estimate was used to make final determinations about the accuracy of each prediction. If the prediction was within the margin of error for the model, it was deemed accurate, as was the case of our example.

**Findings**

The mean percentage of students at the school level scoring proficient or above on the 2010 through 2012 language arts state standardized tests for our models ranged from 64.28 to 73.83. The standard deviations of the means ranged from 13.60 to 18.04. The mean percentage of students at the school level scoring proficient or above on the 2010 through 2012 mathematics state standardized tests ranged from 65.69 to 83.37, and the standard deviations ranged from 10.86 to 16.94 (see Table 3).

The hierarchical regression calculations resulted in 18 models of best fit: one for each grade level and subject in each year 2010 through 2012. We accounted for more than 50% of the variance in the percentages of students scoring proficient or above on the language arts and mathematics portions of the 2010 through 2012 state standardized tests in 16/18 (89%) of our models. The $R^2$-squared values ranged from .351 for the grade six 2012 math test to .709 for the grade eight 2012 language arts test (see Table 4). The $R^2$-squared...
values appear at the top of each row in Table 4, and the standard error of the estimates appears below each R-squared value. The standard error of the estimates ranged from 7.79 for the grade six 2011 mathematics model to 10.66 for the grade seven 2011 language arts model.

Each model of best fit produced a set of statistically and practically significant predictor variables. The final set of variables for each model was narrowed from the original 18 to between two and four variables. All the models of best fit included variables related to family human capital and community social capital.

The majority of the models of best fit, 14/18 (78%), excluding 2010 grade six English language arts, 2011 grade six mathematics, 2011 grade seven math, and 2012 grade seven English language arts, included the same three variables: (a) percentage of families in a community with income over $200,000 a year, (b) percentage of people in a community in poverty, and (c) percentage of people in a community with bachelor’s degrees. Those three variables accurately predicted results for 78% of our samples.

The 2012 grade seven language arts and 2011 grade six mathematics models excluded the percentage of families in a community with income over $200,000 a year. Thus, 16/18 (89%) of the models included the percentage of people in a community in poverty and the percentage of people in a community with bachelor’s degrees.

The 2011 grade seven math model of best fit included the (a) percentage of people in a community with advanced degrees, (b) percentage of people in a community without a high school diploma, and (c) percentage of lone parent households. The 2010 grade six language arts model of best fit included the (a) percentage of households in a community with income over $200,000 a year, (b) percentage of people in a community without a high school diploma, (c) percentage of families in a community living in poverty, and (d) percentage of families headed by a lone parent female.

Finally, we used the unstandardized betas and constants from the statistically significant variables identified in each hierarchical regression models of best fit as part of our predictive algorithms. We predicted accurately (within the standard error of the estimate) the percentage of students scoring proficient or above on the New Jersey mandated standardized tests in language arts and mathematics for 70% to 78% of the schools in our samples (see Table 5).

For example, our algorithms for the 2011 grade eight English language arts test and 2011 grade six math tests predicted the school-level percentage of students who scored proficient and above in 78% of the schools.

Table 3
Means and Standard Deviations for Students in Sample Scoring Proficient or Above

|          | Grade 6 ELA | Grade 6 M | Grade 7 ELA | Grade 7 M | Grade 8 ELA | Grade 8 M |
|----------|-------------|-----------|-------------|-----------|-------------|-----------|
| 2010     | 71.43 (14.37)| 76.61 (13.62)| 73.76 (15.01)| 68.49 (15.76)| 73.83 (14.90)| 66.88 (16.58) |
| 2011     | 72.41 (13.60)| 69.77 (15.67)| 68.82 (16.45)| 70.11 (15.20)| 66.57 (18.04)| 68.28 (16.47) |
| 2012     | 70.77 (15.67)| 83.37 (10.86)| 66.57 (16.27)| 67.85 (15.21)| 64.28 (17.96)| 65.69 (16.94) |

Table 4
R-squared Values for Each Model and the Standard Error of the Estimate

|          | Grade 6 ELA | Grade 6 M | Grade 7 ELA | Grade 7 M | Grade 8 ELA | Grade 8 M |
|----------|-------------|-----------|-------------|-----------|-------------|-----------|
| 2010     | 0.643 8.640 | 0.482 9.870 | 0.686 8.380 | 0.600 9.950 | 0.707 8.550 | 0.625 10.140 |
| 2011     | 0.674 7.800 | 0.514 7.790 | 0.584 10.66 | 0.504 10.790 | 0.669 10.290 | 0.611 10.260 |
| 2012     | 0.618 9.740 | 0.351 8.790 | 0.661 9.700 | 0.587 10.290 | 0.709 9.610 | 0.6480 9.880 |
in those samples. The algorithms for the 2010 grade six language arts, 2011 grade seven math, and the 2012 grade eight language arts predicted accurately the school-level percentage of students scoring proficient or above for 75% of the schools in those samples (the majority of the middle schools in the state excluding the large urban districts).

Essentially, the results suggest that if we have access to the U.S. Census data for the (a) percentage of families in a community with income over $200,000 a year, (b) percentage of people in a community in poverty, and (c) percentage of people in a community with bachelor’s degrees, the probability is high that we can predict the percentage of students in grades six, seven, and eight in each school who will score proficient or above on the New Jersey standardized language arts and mathematics tests.

### Table 5

|            | Grade 6 ELA | Grade 6 M | Grade 7 ELA | Grade 7 M | Grade 8 ELA | Grade 8 M |
|------------|-------------|-----------|-------------|-----------|-------------|-----------|
| 2010       | 75          | 74        | 73          | 72        | 73          | 72        |
| 2011       | 70          | 78        | 70          | 76        | 78          | 74        |
| 2012       | 74          | 72        | 73          | 75        | 75          | 73        |

Conclusions

We interpret the findings from this and prior studies to suggest that using the student results from standardized tests to rate, rank, judge, or evaluate middle level administrators is not “transparent and fair” as required by the RTTT guidelines (United States Department of Education, 2009, p.9) and ESSA requirements. The results raise important issues about fairness and transparency in terms of (a) using results that are influenced strongly by factors outside of school and out of control of the middle level administrator, and (2) using results that can be predicted with a good deal of accuracy by family and community demographic factors. We view the two issues presented above as serious challenges to policies and practices that rely on standardized tests results to rate, rank, judge, or evaluate middle level administrators. We view the issues raised by our results and those of others also as challenges to the claimed scientific objectivity of results from standardized tests.

Provisions within RTTT, ESSA, and hold-over requirements in some state’s NCLB waivers that require, encourage, or reward the use of standardized test scores to rank, rate, categorize, or judge the effectiveness of school-level administrators seem fatally flawed. Given the other pressing responsibilities that middle level administrators must attend to, such as the vast physical, social, and emotional changes that occur with students during their middle level years, the use of test results as a determining factor for middle level principal effectiveness seems to run counter to the overall middle level vision and philosophy as describe in *Turning Points 2000* (Jackson & Davis, 2000), *This We Believe* (NMSA, 2010), and other seminal middle level sources.

The results from our study and previous studies, considered in concert with Bronfenbrenner’s EST, suggest policies and programs that mandate test results be used as the deciding factor in the effectiveness of middle level administrators are not grounded in sound science. Standardized test results are simply too unstable, inherently prone to contamination from non-school factors, and not representative of the multifaceted job middle level administrators perform. The results do not accurately reflect the ecosystem that influences students’ social and emotional development or their direct or collateral learning (Bronfenbrenner, 1979).

The tests provide only blunt measures of a narrow set of skills and cannot capture the nuances of middle level leadership, some of which take years to produce noticeable results, such as social and emotional coaching. In many cases, the results of middle level leadership can never be measured by student test results. For example, do test scores measure students’, parents’, and teachers’ sense of safety and security, their sense of belonging, their hope for the future, or sense of community that exists in a middle level school (Beane, 1990)? Standardized tests cannot measure pride, self-efficacy, resilience, or compassion,
yet middle level administrators help foster those attributes.

Bronfenbrenner’s (1979) EST brings to the forefront the myriad factors that interact to influence learning and combinations of various factors that can influence learning in different and unanticipated ways. The human capital of the family and social capital of the community in which students live and grow are integral parts of the ecology of learning. Family and community capital provide opportunities for formal and collateral learning opportunities and interactions that can enhance or impede achievement on traditional measures of achievement, such as standardized tests (Tanner & Tanner, 2007). Community and family development is not within the purview of the school principal. Those are influenced in part by public policy and controlled by political bureaucrats and legislators.

The use of standardized test results to judge middle level administrators belies a simplistic, mechanistic view of the multifaceted worlds of education leadership and child development, and ignores the social, emotional, physiological, physical, and ecological influences that cannot be controlled by school personnel or policies. This type of evaluation scheme seems inherently unfair and opaque. We view those who mandate and enforce middle level administrator evaluation and rating schemes based on student test results (and who claim they are not aware of the issues raised by these results and the results of other studies) as somewhat disingenuous and in dereliction of their duty to children, and in violation of basic professional ethics. We view the continued use of students’ results from standardized tests as criteria to determine middle level administrator effectiveness education malpractice.

References

Beane, J. A. (1990). A middle school curriculum from rhetoric to reality. Columbus, OH: National Middle School Association.

Becker, G. S. (1964, 1993, 3rd ed.). Human capital: A theoretical and empirical analysis, with special reference to education. Chicago: University of Chicago Press.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

Coleman, J. S. (1988). Social capital in the creation of human capital. American Journal of Sociology, 94, S95–S120.

College Board. (2012). 2012 college-bound seniors: Total group profile report. New York, NY: Author. Retrieved from http://research.collegeboard.org/programs/sat/data/archived/cb-seniors-2012

Darnell, B. (2015). The value of iowa school district demographic data in explaining school district ITBS/ITED 3rd and 11th grade language arts and mathematics scores. Seton Hall University Dissertations and Theses (ETDs). Paper 2075.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. Journal of Family Psychology, 19(2), 294–304.

Dawson, D. A. (1991). Family structure and children’s health and well-being: Data from the 1988 national health interview survey on child health. Journal of Marriage and Family, 53(3), 573–584.

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). Thousand Oaks, CA: Sage Publications.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26, 499–510.

Harwell, M. R., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. Educational Researcher, 39(2), 120–131.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Applied statistics for the behavioral sciences. Belmont, CA: Wadsworth.

Jackson, A. W., & Davis, G. A. (2000). Turning points 2000. New York: Teachers College Press.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. Educational Researcher, 30(2), 3–13.

Kelly, E. (2010). Equal opportunity, unequal capability. In H. Brighouse & I. Robeyns (Eds.), Measuring justice: Capabilities and primary goods (pp. 61–80). Cambridge: Cambridge University Press.

Lounsbury, J. H. (1991). As I see it. Columbus, OH: National Middle School Association.

Mann, M. J. (2013). Helping middle school girls at risk for school failure recover their confidence and achieve school success: An experimental study. Research in Middle Level Education Online, 36 (9), 1–14.
Maylone, N. (2002). The relationship of socioeconomic factors and district scores on the Michigan educational assessment program tests: An analysis. Unpublished, Eastern Michigan University, Ypsilanti, Michigan.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common core state standards. Washington, DC: Authors.

National Middle School Association. (2010). This we believe. Westerville, OH: Author.

New Jersey Department of Education. (2015). Data. Author. Retrieved from http://www.state.nj.us/education/data/

Nichols, S. L., & Berliner, D. C. (2007). Collateral damage: How high-stakes testing corrupts America’s schools. Cambridge, MA: Harvard University Press.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C. A. § 6301 et seq. (2002).

Putnam, R. D. (2000). Bowling alone: The collapse and revival of American community. New York, NY: Simon & Schuster.

Sackey, A. N. L. (2014). The influence of community demographics on student achievement on the Connecticut mastery test in mathematics and English language arts in grade 3 through 8. Seton Hall University Dissertations and Theses (ETDs). Paper 2010.

Scherrer, J. (2014). The role of the intellectual in eliminating poverty: A response to Tierney. Educational Researcher, 43, 201–207.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417–453.

Tanner, D., & Tanner, L. (2007). Curriculum development: Theory into practice (4th ed.). Upper Saddle River, NJ: Pearson.

Tienken, C. H. (2011). Structured inequity: The intersection of socioeconomic status and the standard error of measurement of state mandated high school test results. NCPEA Yearbook, 257–271.

Tienken, C. H. (2016). Standardized test results can be predicted, so stop using them to drive education policymaking. In C. Tienken & C. Mullen (Eds.), Education policy perils: Tackling the tough issues (pp. 157–185). Philadelphia, PA: Taylor Francis Routledge.

Turnamian, P. (2012). The value of NJ school district demographic data in explaining school district NJ ASK Grade 3 language arts and mathematics scores. Unpublished doctoral dissertation, Seton Hall University.

United States Department of Education. (2009). Race to the top program: Executive summary. Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

United States Department of Education. (2015). ESEA flexibility. Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Weinberg, B. A. (2001). An incentive model of the effect of parental income on children. Journal of Political Economy, 109(2), 266–280.

Wiles, J., & Bondi, J. (1981). The essential middle school. Columbus, OH: Charles E. Merrill Publishing Co.

The World Bank. (2011). Social capital. Washington, DC: Author. Retrieved from http://go.worldbank.org/K4LUMW43B0