RMLE Online
Research in Middle Level Education

David C. Virtue, Ph.D., Editor
University of South Carolina
Columbia, South Carolina

2014 • Volume 38 • Number 3 ISSN 1940-4476

Interim Testing, Socio-Economic Status, and the Odds of Passing Grade 8 State Tests in New Jersey

Gerard Babo, Ed.D.
Seton Hall University, South Orange, NJ

Christopher H. Tienken, Ed.D.
Seton Hall University, South Orange, NJ

Maria A. Gencarelli, Ed.D.
Woodbridge Township (NJ) School District

Abstract

A review of the literature pertaining to the effect and influence of commercially-prepared interim assessments in mathematics and language arts literacy reveals a lack of quantitative data to determine the value of these products for school reform. This study examined the ability of commercially-prepared interim pretest and posttest assessments in language arts literacy (LAL) and math to predict student achievement on the state-mandated summative assessment in those subjects. Analyses were conducted using binary logistic regression models. Data for this study included results from the state-mandated grade 8 assessments (NJ ASK 8) for 291 eighth grade students enrolled in two middle level schools located in a suburban/urban central New Jersey community during the 2009–2010 academic school year. The findings suggest that the predictive value of the students' pretest results is very similar to that of the posttest results and call into question the efficacy of implementing both interim pretests and posttests.

Introduction

Effective middle level schools promote the use of varied, ongoing assessments that measure and advance student learning (National Middle School Association, 2010), however the proliferation of federal- and state-mandated testing policies in recent decades has tended to emphasize narrowly focused, standardized systems of assessment and evaluation. The No Child Left Behind Act of 2001 (NCLB) and the reforms brought about with the creation of the Common Core State Standards have perpetuated the practice of using standardized test results as the deciding factor to evaluate student achievement and the quality of school personnel. The 2014 proposal for the reauthorization of NCLB and the NCLB waivers granted to various states include provisions for testing of students each year in third through eighth grade and once during high school. Thus, it seems the pressure on school administrators to raise test scores will remain the focus of education policy and practice for the near future.

Exacerbating the testing-related pressures on educators are the persistent achievement gaps among demographic groups. In no state have students categorized as economically disadvantaged, as a group, had a higher mean scale score on their states' tests of mathematics and language arts than the group of students not categorized as economically disadvantaged, and the scores from economically disadvantaged students more frequently cause school personnel to be cited as ineffective by state education bureaucrats (Tienken, 2011). Middle level education leaders continue to search for scientifically based interventions to address issues related to improving achievement on state-mandated tests of mathematics and language arts, as they are required to do by state law, and some studies have sought to determine factors that correlate with student success (e.g., Hunley, Davies, & Miller, 2013) to strategically focus interventions. Interventions targeted toward historically low-achieving groups, such as economically disadvantaged students, are needed due to the rise in the percentage of students living in poverty. For example, 48.1% of all K–12 students were eligible for free or reduced price lunches during the 2010–2011 school year compared to 38.3% during the 2000–2001 school year (U.S. Department of Education, 2011). In the South, more than half of all students qualify for free or reduced lunch and are considered economically disadvantaged.

Purpose of the Study

One school-wide academic intervention some middle level administrators use in an attempt to raise student achievement on state-mandated standardized tests is the administration of commercially prepared standardized interim assessments in mathematics and language arts literacy (LAL) (Perie, Marion, Gong, & Wurtzel, 2007; Sloane & Kelly, 2003). School administrators direct teachers to administer the assessments as pretests and posttests or on regular intervals such as every six weeks or every marking period (Christman et al., 2009). Interim assessment tools widely used in the U.S. include Measures of Academic Progress (MAP) produced by Northwest Evaluation Associates, Learnia Formative Assessment created by the Pearson Corporation, and interim assessments offered by the Smarter Balanced Assessment Consortium designed to align with the Common Core State Standards.

One prevailing claim of companies that market commercially-prepared interim assessment products is that the more frequently students take interim tests that mimic state-mandated tests, the better they will perform on the state-mandated tests. Another purported reason to use interim assessments is that they provide school administrators and teachers with data to make instructional decisions, enabling formative, data-driven decision making. A third claim by vendors is that posttest interim assessments can predict with a high degree of accuracy which students might score proficient, or not, on the state test before the date of the state-mandated standardized tests and thus allow educators to apply interventions that will increase the chances of those students passing the state tests.

A lack of empirical research exists that explains how well interim pretest and posttest assessments predict future achievement on state-mandated standardized tests of LAL and mathematics, especially when controlling for student demographic factors (Brown & Coughlin, 2007; Goertz, Olah, & Riggan, 2009). Our purpose for conducting this study was to determine how well results from one pretest and posttest interim assessment product predicted performance on state-mandated tests. We present results from more than 290 eighth grade students from an economically diverse school district in central New Jersey.

This study was guided by two essential research questions:

1. How well does a commercially prepared pretest and posttest interim assessment predict eighth grade student performance on the New Jersey state-mandated LAL and mathematics tests?

2. How well does a commercially prepared pretest and posttest interim assessment predict performance of eighth grade students eligible for free lunch on the New Jersey state-mandated LAL and mathematics tests?

Theoretical Framework

We used the lens of production function theory to investigate the use of interim assessments in schools (Greenwald, Hedges, Laine, 1996). A reason school administrators use interim assessments is to increase student production or output via test scores. In fact, some commercially prepared interim assessment products claim that their use will improve scores on state-mandated tests. Thus, the theoretical framework for using interim assessments flows from the idea that interim assessments provide important data to teachers and school administrators that they can use to improve instructional practice and student

learning at the school level via more accurate student level achievement data. The improved practices are supposed to increase teacher productivity, and those increases in productivity should yield increases in test scores. While we view this linear perspective of teaching as problematic, it is a view that has gained greater acceptance in the era of accountability (Paulson & Marchant, 2009).

Paulson and Marchant (2009) emphasized those who use standardized tests as a tool in the production function process believe, among other things, that (a) the collective scores of teachers' students reflect the output of their instruction; (b) the scores of schools and districts reflect the output of their educational programs; and (c) the scores of test takers from a state represent the output of education and educational policies of the state. Interventions or inputs that school administrators put into the system, such as interim assessments, are supposed to yield greater teacher and student output.

Methodology

Variables

Results from various studies since the mid-1960s have suggested there are a host of student demographic variables and school variables that influence outputs. Variables that influence student achievement are generally categorized as either pertaining to the school, the student, or the teacher.

The dependent variables in this study were eighth grade student proficiency ratings on the LAL and M sections of the New Jersey assessment of skills and knowledge (NJ ASK). For this study, a student was rated either "not-proficient" if he or she did not at least achieve the mandated minimum scale score of 200 or "proficient" if he or she achieved a scale score of 200 or greater. The independent variables found in the literature with a demonstrated history of influencing achievement on such tests included (a) student eligibility for free lunch, (b) student

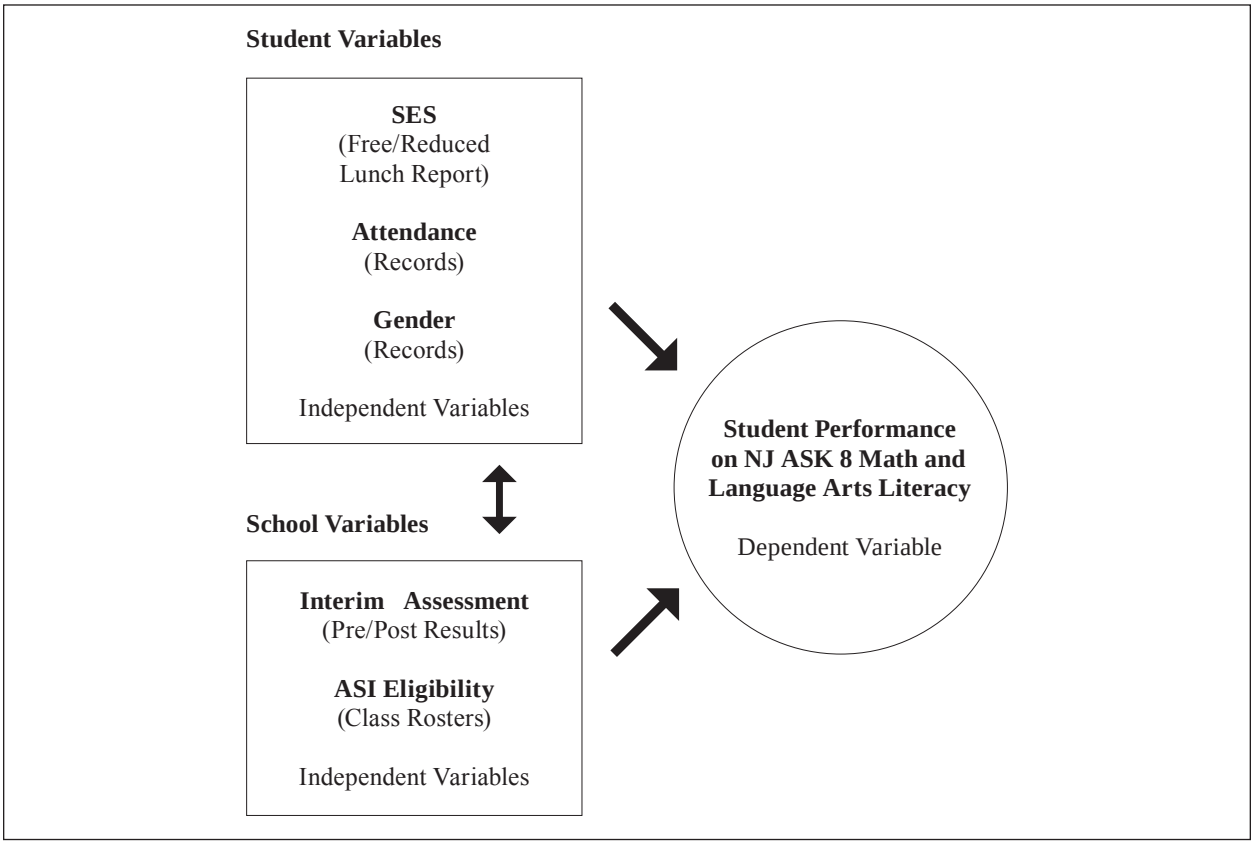

Figure 1. Production Function Framework with Variables

attendance, (c) gender, and (d) eligibility for Title I basic skills (ASI). Pretest and posttest interim assessment scores for LAL and M were included because the state commissioner of education at the time stated that the commercially prepared interim assessment tool was in fact a variable that could raise student achievement.

The school district administration in the study district also mandated the use of the commercially prepared assessment product as a way to increase achievement, based on the information the vendor provided to them. All teachers had to administer a pretest in the fall, use the results to guide teaching, and then administer a posttest in the late winter before the NJ ASK. Due to the limited number of content-specific degrees held by the middle level teachers in this study and the lack of variance in teacher quality as measured by annual summative evaluations, we removed the variable of teacher quality from the theoretical model (see Figure 1).

Design

Because educational researchers are often unable to perform randomized experiments and quasi-experiments in school settings, "non-experimental research is frequently an important and appropriate mode of research in education" (Johnson, 2001, p.3). We conducted a relational, cross-sectional, predictive study that used collected data from one point in time—the 2009–2010 school year.

We collected data from two middle level schools in the district (N=448). The schools were similar in demographic makeup of students, student achievement on the grade 8 state tests in mathematics, and teacher degree status. Data from eighth grade students who met the following criteria were included in the final sample: (a) attended school in the district at least two years, (b) were not eligible for special education services during the last two years, (c) achieved a valid score on the commercially prepared interim pretest and posttest in eighth grade, and (d) achieved a valid score on the LAL and math section of the NJ ASK 8 state-mandated test.

We excluded data from students eligible for special education services due to the individualized nature of their programs and also due to the fact that the NJ ASK 8 scores derived from students eligible for special education services have never been validated empirically by the New Jersey Department of Education (NJDOE) or the maker of the NJ ASK. NJDOE has neither conducted nor commissioned any validation studies to determine the effect of accommodations and modifications on test results for students who receive special education services. The final sample included data from 291 eighth grade students. Approximately 34% of the students in the sample qualified for free lunch and 61% were non-white.

We used binary logistic regression to determine the predictive ability, if any, of interim assessments on student achievement when controlling for other student and school variables. Researchers typically use a logistic regression analysis strategy to explore and maximize prediction of a dichotomous outcome, such as "proficient" or "not proficient" (Leech, Barrett & Morgan, 2011). The use of logistic regression requires dichotomous dependent variables. We used 0,1 to represent "not proficient" (0) or "proficient and above" (1) on the grade 8 state tests of LAL and math (NJ ASK 8). Each case (n=291) was represented only once, thereby satisfying the assumption of independence.

We exceeded the minimum requirement of cases for the total number of predictors for each model. In general, binary logistic regression requires large sample sizes to maintain the accuracy and integrity of the analyses, approximately 20 cases per predictor (Leech et al., 2011). In our case, six predictor variables were used in the models requiring a minimum of 120 cases, which we more than satisfied for a majority of the analyses.

We created scatter diagrams of residuals and normal probability plots of residuals to further test data assumptions. Additionally, standard simultaneous regression analyses were run on all models to check for multicollinearity. All tolerances and variance inflation factors (VIFs) were within the normal ranges. Overall, the data exhibited normal characteristics. Then we constructed predictive models for both LAL and math results for the following groups: (a) all students, (b) students eligible for free lunch, and (c) students not eligible for free lunch. Thus, our strategy produced six models.

Results

Grade 8 LA All students

We used binary logistic regression to assess whether the six predictor variables—gender, attendance, SES, ASI, Pre LA (interim assessment) and Post LA (interim assessment)—significantly predicted whether a student achieved a passing score on the

NJ ASK 8 LAL test. The results of a test of the full model with all six variables entered compared to the constant-only model was statistically significant, $\chi 2\ (6) = 62.413$, N = 291, $p < .001$. Approximately 19% to 42% of the variance associated with student pass rate on the NJ ASK 8 LAL could be explained by the model with Cox and Snell $R^2 = .193$ and Nagelkerke's $R^2 = .419$, although the Cox and Snell R^2 routinely underestimates variance (Leech et al., 2011). Two of the six variables were statistically significant in the model: (a) Pre LA interim results and (b) Post LA interim results.

Table 1 contains results for the raw score binary logistic regression coefficients, Wald statistic, and the 95% confidence interval for only those variables found to be statistically significant in the model, which agrees with the suggested reporting guidelines for binary logistic regression (Tabachnick & Fidell, 2007, as cited in Warner, 2008). The results suggest that a one unit increase in a student's Pre LA score equates to an expected increase of 1.06, or 6%, in the odds of a student passing the NJ ASK 8 LAL. The confidence interval for Pre LA indicates a possible range of increase in the odds of 1.01 to 1.12. A one unit increase in the student's Post LA score equates to an expected increase of 1.10, or 10%, in the odds of a student passing the NJ ASK 8 LAL. The confidence interval for Post LA indicates a possible range of increase in the odds of 1.05 to 1.15.

The results from Model I indicate that the odds of a student receiving a passing score on the NJ ASK 8 LAL are greater as a student's performance increases on both the Pre LA and Post LA. The results suggest that the predictive value of the students' Pre LA results is very similar to that of the Post LA results and call into question the value of pretesting and posttesting. Because all teachers used the product and made instructional adjustments based on the results, it seems as if the value of posttesting might be overstated if the pretest provides almost the same amount of predictive power.

Grade 8 Math All Students

We conducted another test of the full model for mathematics with eight variables entered and compared with the constant-only model. The model was statistically significant, $\chi 2\ (8) = 118.4$, N = 291, $p < .001$. Approximately 33% to 63% of the variance associated with the student pass rate on the NJ ASK 8 Math was explained by the model with Cox and Snell $R^2 = .334$ and Nagelkerke's $R^2 = .627$. Three of the eight variables were statistically significant in the model: (a) Pre Math interim assessment, (b) Post Math interim assessment, and (c) Post LA interim assessment.

We present results in Table 2 for the raw score binary logistic regression coefficients, Wald statistic, and the 95% confidence interval for only those variables found to be statistically significant in the model. The

Table 1
Binary Logistic Regression Analysis: All Students NJ ASK 8 LAL Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Pre LA	.062	4.862	.03	1.06	1.01	1.12
Post LA	.093	15.72	<.001	1.10	1.05	1.15
Constant	2.28	127.35	<.001	9.78		

**= 95% Confidence Interval for exp(B)*

Table 2
Binary Logistic Regression Analysis: All Students NJ ASK 8 Math Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Pre LA	.121	13.19	<.001	1.13	1.06	1.21
Post Math	.071	9.85	.002	1.07	1.03	1.12
Post LA	.065	7.78	.005	1.07	1.02	1.12
Constant	1.93	119.85	<.001	6.87		

**= 95% Confidence Interval for exp(B)*

results suggest that a one-unit increase in a student's Pre Math score equates to an expected increase of 1.13, or 13%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Pre Math indicates a possible range of increase in the odds of 1.06 to 1.21. A one unit increase in the student's Post Math score equates to an expected increase of 1 .07, or 7%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Post Math indicates a possible range of increase in the odds of 1.03 to 1.12. A one unit increase in the student's Post LA score equates to an expected increase of 1.07, or 7%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Post Math indicates a possible range of increase in the odds of 1.03 to 1.12.

These results suggest that the odds of a student achieving a passing score on the NJ ASK 8 Math increase as a student's performance increases on the Pre Math, Post Math, and Post LA. Additionally, these results suggest that student performance on the Pre Math is a slightly stronger predictor of the odds of passing the NJ ASK 8 Math than the Post Math and the Post LA. Interestingly, it appears that the Pre Math interim assessment result is a better predictor of being proficient on the NJ ASK 8 Math than Post LA or Post Math. As with the LAL results, it seems as if the Posttest Math product is unnecessary in terms of predicting future achievement because the results from Pretest Math product is the stronger predictor.

Students Eligible for Free Lunch

We created two models to examine any discrete differences in the odds ratio for passing the NJ ASK 8 Math and LAL that might associate with students' SES status as measured by free lunch eligibility. The first model included data for only those students eligible for free lunch to assess whether the five predictor variables of gender, attendance, ASI, Pre LA, and Post LA predicted with statistical significance whether a student received a passing score on the NJ ASK 8 LAL test. A test of the full model with the five variables entered compared with the constant-only model (see Table 3) was statistically significant, $\chi 2\ (5) = 26.319$, $N = 112$, $p < .001$. Approximately 21% to 41% of the variance associated with student pass rate on the NJ ASK 8 LAL for students eligible for free lunch could be explained by the model with Cox and Snell $R^2 = .209$ and Nagelkerke's $R^2 = .409$. Post LA was the only statistically significant variable in the model.

Table 3 displays the raw score binary logistic regression coefficients, Wald statistic, and the 95% confidence interval for only those variables found to be statistically significant in the model, Post LA. The results from the model indicate that a one-unit increase in a student's Post LA score relates to an expected increase of 1.09, or 9%, in the odds of a student passing the NJ ASK 8 LAL. The confidence interval for Post LA indicates a possible range of increase in the odds of 1.02 to 1.17.

Table 3
Binary Logistic Regression Analysis: Students Eligible for Free Lunch NJ ASK 8 LAL Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Pre LA	.090	6.917	.009	1.09	1.02	1.17
Constant	2.03	47.362	<.001	7.62		

**= 95% Confidence Interval for exp(B)*

Table 4
Binary Logistic Regression Analysis: Students Eligible for Free Lunch NJ ASK 8 Math Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Pre LA	.156	8.017	.005	1.12	1.05	1.30
Post LA	.077	4.829	.03	1.08	1.01	1.12
Constant	1.59	39.793	<.001	4.90		

**= 95% Confidence Interval for exp(B)*

We performed another binary logistic regression on the data for students eligible for free lunch to assess whether the five predictor variables of gender, attendance, ASI, Pre LA, Pre Math, Post Math, and Post LA predicted with statistical significance whether a student received a passing score on the NJ ASK 8 Math. A test of the full model compared with the constant-only model (See Table 4) was statistically significant, $\chi 2$ (5) = 55.3, N = 112, p < .001.

Approximately 39% to 65% of the variance associated with the student pass rate on the NJ ASK 8 Math for students eligible for free lunch could be explained by the model with Cox and Snell R^2 = .209 and Nagelkerke's R^2 = .409. Pre Math and Post LA were statistically significant in the model. However, it should be noted, the samples size for this single analysis (N = 112) did not meet the requirements for the recommended sample size (20 x k or 20 x 7 = 140) suggested by Leech and associates (2011) and, consequently, the results from the model are tentative.

Table 4 presents the raw score binary logistic regression coefficients, Wald statistic, and the 95% confidence interval for only those variables found to be statistically significant in the model. The results indicate that a one-unit increase in a student's Pre Math score equates to an increase of 1.12, or 12%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Pre Math indicates a possible range of increase in the odds of 1.05 to 1.3. A one-unit increase in a student's Post LA score equates to an expected increase of 1.08, or 8%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Post LA indicates a possible range of increase in the odds of 1.01 to 1.12. The odds for a student who is eligible for free lunch of receiving a passing score on the NJ ASK 8 Math increase as a student's performance increases on the Pre Math and Post LA interim assessments. Once again, the pretest is the best predictor of passing the NJ ASK, calling into question the vendor's claim that school personnel must use both the pretest and posttest.

Students Not Eligible for Free Lunch

We performed a final set of two binary logistic regressions on the data for students not eligible for free lunch to assess whether the five predictor variables of gender, attendance, ASI, Pre LA, and Post LA predicted with statistical significance whether a student received a passing score on the NJ ASK 8 LAL. A test of the full model compared with the constant-only model (see Table 5) was statistically significant, $\chi 2$ (5) = 37.041, N = 179, p < .001. Approximately 19% to 44% of the variance associated with the student pass rate on the NJ ASK 8 LAL could be explained by the model with Cox and Snell R^2 = .187 and Nagelkerke's R^2 = .443. Only Post LA was statistically significant in the model.

The raw score binary logistic regression coefficients, Wald statistic, and the 95% confidence interval for the only variable found to be statistically significant in the model (Post LA) is presented in Table 5. The model indicates that a one-unit increase in a student's Post LA score equates to an expected increase of 1.11, or 11%, in the odds of a student who is not eligible for free lunch passing the NJ ASK 8 LAL. The

Table 5
Binary Logistic Regression Analysis: Students Not Eligible for Free Lunch NJ ASK 8 LAL Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Pre LA	.150	9.186	.002	1.11	1.04	1.19
Constant	2.47	78.534	<.001	11.79		

**= 95% Confidence Interval for exp(B)*

Table 6
Binary Logistic Regression Analysis: Students Not Eligible for Free Lunch NJ ASK 8 Math Pass Rate

Predictor Variables	B	Wald Chi-Square	P	exp(B)* Odds Ratio	Lower	Upper
Post Math	.087	4.829	.03	1.08	1.01	1.12
Constant	2.19	39.793	<.001	4.90		

**= 95% Confidence Interval for exp(B)*

confidence interval for Post LA indicates a possible range of increase in the odds of 1.04 to 1.19.

Our final model tested how well the seven predictor variables of gender, attendance, ASI, Pre LA, Pre Math, Post LA, and Post Math predicted with statistical significance whether a student not eligible for free lunch received a passing score on the NJ ASK 8 Math. A test of the full model was statistically significant, $\chi 2\ (5) = 61.825$, $N = 179$, $p < .001$ (See Table 6). Approximately 29% to 61% of the variance associated with student pass rate on the NJ ASK 8 Math for only those students eligible for free lunch could be explained by the model with Cox and Snell $R^2 = .292$ and Nagelkerke's $R^2 = .609$. Only Post Math was statistically significant in the model.

The results indicate that a one-unit increase in a student's Post Math score is related to an expected increase of 1.09, or 9%, in the odds of a student passing the NJ ASK 8 Math. The confidence interval for Post Math indicates a possible range of increase in the odds of 1.02 to 1.17. These results are markedly different than those found for students on free lunch where the odds of passing the NJ ASK 8 Math increased based on their Pre Math and Post LA performance.

Discussion

Table 7 summarizes statistically significant variables from each model and provides the odds ratio, or *exp (B)*. Post LA is a statistically significant predictor variable in five of the six models, including math models. This finding aligns with the publically available information found in state test technical manuals on the correlation between LAL and math scores on all state-mandated tests. Results from state-mandated mathematics tests that include open-ended questions, like those administered in New Jersey, generally have moderate to strong correlations with results from LAL tests.

Pre Math was the strongest statistically significant predictor in two models: (a) All Students NJ ASK 8 Math and (b) Students on Free Lunch NJ ASK 8 Math. This finding seemed to be counter-intuitive based on the marketing materials for the product, which suggest that pretesting and posttesting are necessary and that the posttest is a reliable predictor of future achievement on the state-mandated tests.

Results here bring into question these pre- and post-assessments' predictive ability with regards to student performance on state-mandated assessments. At best, the odds ratio gains reported here, although statistically significant, are minimal in relation to some of the monies being spent on the products. This is just one study done in one school district in New Jersey; however despite these limitations, the data analyses reported here should at the very least urge school administrators and district curriculum coordinators to question the efficacy and overall value of these pre-packaged interim assessments. In addition, the results beg the question: Could well-trained teachers predict—with greater accuracy than test publishers and without the added costs of purchasing a commercial assessment product—which students might need more intervention prior to the state test and those that might not?

Table 7
Model Significant Predictor Variables for NJ ASK 8 Math & LAL and Respective Odds Ratios

Model & # of Predictors ( )	Significant Predictors	Odds Ratio – exp(B)
All Students LAL (2)	Post LA	1.10
	Pre LA	1.06
All Students Math (3)	Pre Math	1.13
	Post Math	1.07
	Post LA	1.07
Eligible Free Lunch LAL (1)	Post LA	1.09
Eligible Free Lunch Math (2)	Pre Math	1.12
	Post LAL	1.08
Not Eligible Free Lunch LAL (1)	Post LAL	1.11
Not Eligible Free Lunch Math (1)	Post Math	1.09

We suggest that school administrators and district curriculum coordinators use the results reported here to provide a rationale and support to begin exploring the possibility of providing district- and/or school-specific professional development in the area of formative assessment (Strahan & Rogers, 2012). Valid, reliable, subject-specific assessments that are criterion-referenced and constructed locally by teachers might prove more educationally valuable than any generic pre-packaged interim or benchmark assessment for the purpose of diagnosing specific school and/or district remediation needs for a unique student population. At the very least, middle level administrators should request independent, empirical evidence that demonstrates that the interim assessment interventions marketed to them are appropriate for the specific populations they administer, lead, and serve. Corporate claims of one-size-fits-all interventions might be overstated, thus wasting precious resources like professional time, taxpayer money, and students' academic futures.

References

Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs

Christman, J., Neild, R., Bulkley, K., Blanc, S., Liu, R. Mitchell, C., & Travers, E. (2009). *Making the most of interim assessment data: Lessons from Philadelphia.* Philadelphia, PA: Research for Action.

Goertz, M. E., Olah, L. N., & Riggan, M. (2009, December). Can interim assessments be used for instructional change? Consortium for Policy Research in Education Policy Brief. Retrieved from http://www.cpre.org/sites/default/files/policybrief/894_role-policy-brief-final.pdf

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396.

Hunley, S. A., Davies, S. C., & Miller, C. R. (2013). The relationship between curriculum-based measures in oral reading fluency and high-stakes tests for seventh grade students. *Research in Middle Level Education Online, 36*(5), 1–8.

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

Leech, N, Barrett, K, & Morgan, G. A. (2011). *IBM SPSS for intermediate statistics* (4th ed.). New York, NY: Routledge.

National Middle School Association. (2010). *This we believe: Keys to educating young adolescents.* Westerville, OH: Author.

Paulson, S. E., & Marchant, G. J. (2009). Background variables, levels of aggregation, and standardized test scores. *Education Policy Analysis Archives, 17*(22). Retrieved from http://epaa.asu.edu/epaa/v17n22/

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). The role of interim assessments in a comprehensive assessment system: A policy brief. Retrieved from http://www.achieve.org/files/TheRoleofInterimAssessments.pdf

Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory into Practice, 42*(1), 12–17.

Strahan, D., & Rogers, C. (2012). *Research summary: Formative assessment practices in successful middle level classrooms.* Retrieved from http://www.amle.org/TabId/180/ArtMID/780/ArticleID/301/Research-Summary-Formative-Assessment-Practices.aspx

Tabachnick, B. G. & Fidell, L. S. (2007) *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.

Tienken, C. H. (2011). High school exit exams and mismeasurement. *The Educational Forum, 4*(75), 298–314.

U.S. Department of Education. (2011, August). *Number and percentage of public school students eligible for free or reduced-price lunch, by state: Selected years, 2000–01 through 2009–10* [Data file]. Retrieved from http://nces.ed.gov/programs/digest/d11/tables/dt11_045.asp

Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques.* Los Angeles, CA: Sage Publications.