

# Predicting Student Performance in Statewide High-Stakes Tests for Middle School Mathematics Using the Results from Third Party Testing Instruments

Rusen Meylani<sup>1</sup>, Gary G. Bitter<sup>1,2,3</sup> & Rene Castaneda<sup>1</sup>

<sup>1</sup> Mary Lou Fulton Teachers College, Arizona State University, Tempe, Arizona, USA

<sup>2</sup> Learning Sciences Institute, Arizona State University, Tempe, Arizona, USA

<sup>3</sup> Technology Based Learning and Research, Arizona State University, Scottsdale, Arizona, USA

Correspondence: Rusen Meylani, Mary Lou Fulton Teachers College, Tempe, Arizona, USA. Tel: 1-480-249-4218. E-mail: rmeylani@asu.edu

Received: June 19, 2014

Accepted: July 22, 2014

Online Published: August 14, 2014

doi:10.5539/jel.v3n3p135

URL: <http://dx.doi.org/10.5539/jel.v3n3p135>

## Abstract

In this study regression and neural networks based methods are used to predict statewide high-stakes test results for middle school mathematics using the scores obtained from third party tests throughout the school year. Such prediction is of utmost significance for school districts to live up to the state's educational standards mandated by the No Child Left Behind Act by helping them take the necessary measures in a timely manner and avoid penalties such as decreased funding, salary cuts, job losses, the state taking over the school administration, etc. Although the predictive analyses were performed in the context of middle school mathematics, the suggested models can readily be applied to other grade levels and content areas as well.

**Keywords:** regression, neural networks, prediction, statewide high-stakes tests

## 1. Introduction

The No Child Left Behind Act (NCLB, 2001) supports standards-based education reform, based on the idea that setting high standards and establishing quantifiable goals are likely to improve individual student outcomes in education. NCLB requires states to develop basic skills assessments to be administered to all students in certain grades every year, if those states are to receive federal funding for schools. These yearly standardized tests are a major part of the research used to determine whether or not schools are living up to the standards that they are required to meet (Braden, 2002; Linn, 2000). If these standards are not met, the schools face decreased funding, salary cuts, job losses, and/or the state taking over a school district's administration, along with other punishments that contribute to the increased liability brought by the NCLB.

Conforming to the NCLB, the state of Arizona uses a standardized testing system entitled Arizona's Instrument to Measure Standards (AIMS) to track how well students are performing compared to state standards. Students in grades 3 through 8 and 10 take the AIMS in mathematics, reading and writing. The AIMS test is based on the Arizona state standards, which define what students should be learning each year. AIMS results show the level of proficiency a student demonstrates in each of the subject areas tested. For each student taking the AIMS test on a particular content area, the raw score, the scaled score, the placement rating and a pass/fail index is reported.

In order to keep track of the students' progress, many schools employ third party testing instruments developed in accordance with the state's educational standards and known to be highly correlated with the AIMS test. In this study, the scores obtained from such a third party testing system throughout the school year are used to predict the actual AIMS mathematics test results at grades 6, 7 and 8, using regression based methods, for an urban middle school district in Arizona.

Organization of this paper is as follows: First, the related theoretical framework and literature review will be given followed by the rationale and the research questions answered in this paper. In the methodology section, the regression and neural networks based algorithms, the prediction process and the participant profile will be elaborated. The analyses and results will separately display the findings while predicting the scaled score, placement rating and the pass/fail rating in the AIMS Mathematics test. Finally, the results will be discussed in

detail and suggestions for future research will be given.

## 2. Theoretical Framework

Statewide tests, which are conducted near the end of the school year, are considered “high-stakes” tests, single assessments that have a predetermined cut score used to distinguish those who pass from those who fail, with direct consequences associated with passing and failing. For example, major decisions such as retaining students, terminating teachers, and removing funding, accreditation, or administrative control from schools are based on the outcomes of statewide tests. Given that the scores of all students in a school determine the school’s success and schools’ scores are used to determine state performance, there is substantial pressure on teachers to raise students’ test scores.

Although these standardized tests are designed to measure overall academic achievement and are used to make high-stakes decisions, they typically provide too little information too late (McGlinchey & Hixson, 2004). It can be argued that decisions such as retention, which can result in detrimental consequences for students (Jimerson, 2001), should not be solely based on a one-shot assessment. Rather, students and teachers should be assessed and given performance feedback throughout the year, which can improve the probability of schools continuing effective practices and modifying or eliminating ineffective instructional procedures (Good, Simmons, & Kame’enui, 2001). Ensuring effective instruction is being provided during the school year not only prevents individual students from failing but also entire schools from performing poorly. Furthermore, identifying in advance, those students at the risk of not passing the statewide test can initiate taking the necessary measures in a timely manner (Fuchs et al., 2002) reducing a significant amount of pressure experienced by teachers and students as test dates approach.

## 3. Rationale and Research Questions

A third part testing system employed by an urban school district in Arizona was designed to monitor student progress by: a pretest administered at the beginning of the school year (Math Pretest); a posttest administered at the end of the school year (Math Posttest); and three additional tests administered at the end of each of the first, second and third quarters (Math1, Math2 and Math3). The system was developed for Arizona’s AIMS test in mathematics at the middle school level. The strong positive correlation at each of the grade levels 6, 7 and 8, between the scores obtained from these tests and the actual AIMS test (Table 1) motivated the authors to go one step further in an attempt to predict the actual AIMS test results using the scores obtained from these five tests.

Table 1. Correlations between the results from third party tests and the AIMS test in mathematics

Grade		Correlation with AIMS Mathematics Test Scaled Score				
		Math Pretest	Math1	Math2	Math3	Math Posttest
6	R	0.72	0.73	0.80	0.84	0.84
	N	583	655	666	705	703
7	R	0.64	0.76	0.79	0.73	0.78
	N	600	634	658	682	655
8	R	0.71	0.82	0.83	0.82	0.82
	N	594	626	653	671	654

*Note.* All correlations are significant at the 0.01 level.

Considering the decisions based on statewide test outcomes, having the ability to identify students who are unlikely to pass the test in advance is certainly considered highly desirable by school and district administrations. In light of this critical issue, the authors seek to answer the following research questions in this paper: (1) Can the scaled score, the placement rating and the pass/fail rating in the AIMS Mathematics Test be predicted using the Math Pretest, Math1, Math2, Math3 and Math Posttest results obtained from a third party testing system? (2) How do the results of prediction differ based on the method selected for performing the prediction?

## 4. Methodology

In this study, regression and neural networks based models are used to predict the scaled score, pass/fail rating and the placement rating for the AIMS Mathematics Test using the Math Pretest, Math1, Math2, Math3 and Math Posttest scores. In the AIMS test students receive one of the four placement ratings in the AIMS test for each

subject tested: 1-falls far below standards, 2-approaches the standards, 3-meets the standards, or 4-exceeds the standards; the placement ratings of 1 and 2 are considered as failing and the placement ratings of 3 and 4 are considered as passing. Due to the continuous nature of the scaled score, it will be predicted only by linear regression and neural networks based models, whereas the pass/fail and placement ratings will be predicted by all models which will be described later in this section.

#### *4.1 Regression*

In statistics, regression analysis includes techniques for modeling and analysis of several variables, when attention is focused on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us understand how the typical value of the dependent variable changes when any of the independent variables is varied, while the other independent variables remain fixed. Usually, regression analysis estimates the conditional expected value of the dependent variable given the independent variables (i.e. the mean (average) value of the dependent variable when independent variables are kept fixed). Regression analysis is widely used for estimation and prediction. It is also used to explore and comprehend the causal relationships that exist among the independent variables in relation to the dependent variable. Regression analysis is widely used for estimation and prediction (Kutner et al., 2005).

##### *4.1.1 Regression Based Predictive Algorithms*

**Regression Model 1-The Multiple Linear Regression Model.** Multiple linear regression model (Kent, 2001; Kutner et al., 2005) assumes that a linear relation exists between the dependent and the independent variables where the random errors are assumed to be independent and normally distributed random variables with zero mean and constant standard deviation, (i.e., assumptions of normality, linearity, and homogeneity of variance are met).

**Regression Model 2-The Multinomial Logistic Regression Model.** Multinomial logistic regression (Kent, 2001; Kutner et al., 2005; Peng, 2002) does not require any assumptions of normality, linearity, and homogeneity of variance for the independent variables. Because this regression model is less stringent it is often preferred to discriminant analysis when the data does not satisfy these assumptions. Suppose the dependent variable has M nominal (unordered) categories. One value of the dependent variable is chosen as the reference category and the probability of membership in each of the other categories is compared to the probability of membership in the reference category. For the dependent variable with M categories, this requires the calculation of M-1 equations, one for each category relative to the reference category, in order to describe the relationship between the dependent and the independent variables. Please note that multinomial logistic regression (MLR) model ignores the ordinal nature that might exist within the levels of the dependent variable and treats each category in a similar manner

**Regression Model 3-The Cumulative Odds (CO)-Ordinal Logistic Regression Model.** The CO-ordinal regression model (Kent, 2001; Kutner et al., 2005; Peng, 2002) calculates the probability of being at or below category m of an ordinal dependent variable with M categories. Ordinal logistic regression (OR) is different from multinomial logistic regression in that it takes into account the ordinal nature inherent within the levels of the dependent variable, which might be useful in some cases.

Please note that due to the continuous nature of the AIMS Mathematics Scaled Score, Multinomial Logistic Regression and the CO-Ordinal Logistic Regression models cannot be used in the prediction process. On the other hand, the Multinomial Logistic Regression and The CO-Ordinal Logistic Regression models are likely to produce identical results when there are two categorical levels in the data (such as pass/fail rating).

#### *4.2 Neural Networks*

Neural-Networks (NN) are a powerful alternative to linear and nonlinear regression especially for predicting and forecasting but not widely used in the field of education. A neural-network is a mathematical model that imitates the structure and/or the functional aspects of biological neural networks. It is an interconnected group of artificial neurons, which processes information through a series of connections as a means of computing. The modern neural networks are generally used as tools for non-linear modeling of statistical data which may reveal the complex nonlinear relationships between the inputs and outputs better than nonlinear regression methods so as to the patterns in the data.

Neural networks have successfully been employed in a variety of applications that include system identification and control (vehicle and process control), quantum chemistry, game-theory and decision making processes, pattern recognition (radar systems, speech, face and object recognition), sequence recognition (gesture, handwritten text recognition), medical diagnoses, financial applications (automated trading systems), data

mining (prediction, forecasting and modeling), visualization and e-mail spam filtering.

Despite the well-established theory of and various applications regarding neural networks, their capabilities are yet to be discovered and applied to analyses in the context of education. In this paper, we utilize two commonly used Neural-Networks models: the Multilayer Perceptron (MLP) Model and the Radial Basis Function (RBF) Model (Kutner et.al. 2005) to make predictions.

#### 4.2.1 Neural Networks (NN) Based Predictive Algorithms

**The MLP Model.** An MLP neural network contains multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (i.e. a processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training the network and is a modification of the standard linear perceptron having the ability to capture the nonlinearities that may exist in the data.

**The RBF Model.** An RBF neural network uses radial basis functions as activation functions. Such a network is a linear combination of radial basis functions used in function approximation, time series prediction, and control.

#### 4.3 Participants

The participants in this study are the 6th, 7th and 8th grade students from a southwestern urban school district in Phoenix Arizona. The population and ethnic composition of the participants are given in tables 2 and 3 where 94% of the student population are of Hispanic origin and 99% of the students qualified for free lunch as an indicator of their low SES.

Table 2. The participants in this study by school and grade level

School	Grade 6	Grade 7	Grade 8
School 1	60	66	56
School 2	113	105	97
School 3	259	236	277
School 4	211	232	190
School 5	98	92	71

Table 3. Ethnic composition of the participants

Ethnicity	Percentage
White	2.50%
Hispanic	94.00%
African American	3.00%
Other	0.50%

## 5. Analyses and Results

The analyses performed investigated the correlation between the actual and predicted values as well as the percentage of predicted values that correctly matched the actual values. The results will be displayed by grade level.

### 5.1 Prediction of the AIMS Mathematics Scaled Score

Table 4 displays the results of predicting the AIMS Mathematics Scaled Score using multiple linear regression as well as the two neural networks based models. All correlations (R values) are significant at the 0.01 level and indicate strong positive correlation between the actual and predicted values for each of the grade levels. It can be seen that based on the correlational analyses, multiple linear regression and multilayer perceptron (MLP) models perform comparably and both models outperform the radial basis function (RBF) model.

Table 4. Correlations pertaining to the prediction of the AIMS Mathematics scaled score

Grade	Attribute	Regression	Neural Networks	
		Linear Regression	MLP	RBF
6	R	0.91	0.90	0.85
	N	543	535	534
7	R	0.87	0.88	0.82
	N	533	529	529
8	R	0.92	0.90	0.85
	N	553	548	550

Note. All correlations are significant at the 0.01 level.

Figure 1 displays the mean squared errors associated with predicting the AIMS Mathematics Scaled Score using multiple linear regression as well as the two neural networks based models. It can be seen that the mean squared error is greater for the radial basis function model than the multiple linear regression and the multilayer perceptron models which performed comparably at all grade levels.

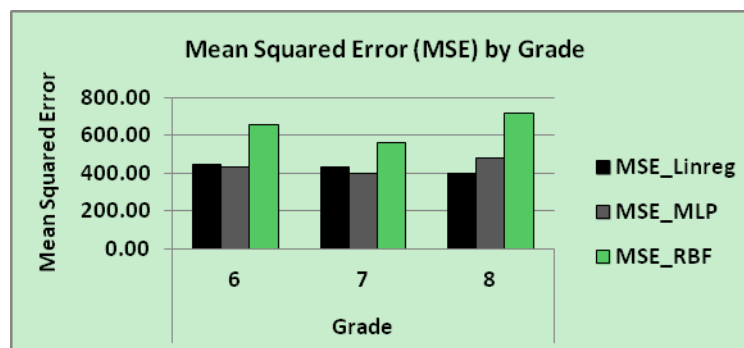


Figure 1. Mean squared errors associated with the prediction of the AIMS Mathematics scaled score

### 5.2 Prediction of the AIMS Mathematics Placement Rating

Tables 5 and 6 as well as Figure 2 display the results of predicting the AIMS Mathematics Placement rating using all models. All correlations (R values) depicted in table 5 are significant at the 0.01 level and indicate strong positive correlation between the actual and predicted values for each grade level and model. The regression models given in the order of decreasing performance are as follows: Best-Regression Model 2 (Multinomial Logistic Regression—MLR); Second Best-Regression Model 3 (CO—Ordinal Logistic Regression—OR); Third Best-Regression Model 1 (Linear Regression); since the predicted value is categorical in nature, it is not surprising that Multinomial Logistic Regression and The CO—Ordinal Logistic Regression outperform Multiple Linear Regression. Please also note that Multinomial Logistic Regression and the CO—Ordinal Logistic Regression produced identical results because the data had two categories (i.e. pass or fail). Both Neural Networks based methods of Multilayer Perceptron (MLP) and Radial Basis Function (RBF) are outperformed by regression based methods.

Table 5. Correlations pertaining to the prediction of the AIMS Mathematics placement rating using all models

Grade	Attribute	Regression		Neural Networks		
		Linear Regression	MLR	OR	MLP	RBF
6	R	0.87	0.96	0.90	0.83	0.84
	N	543	543	543	534	537
7	R	0.82	0.88	0.84	0.80	0.78
	N	533	533	533	528	531
8	R	0.87	0.98	0.91	0.89	0.85
	N	553	553	553	550	549

Note. All correlations are significant at the 0.01 level.

Table 6. Prediction rates as percentages for the AIMS Mathematics placement rating using all regression models

		Regression						Neural Networks					
		Linear Regression		MLR		OR		MLP		RBF			
		N	Percent	N	Percent	N	Percent	N	Percent	N	Percent		
Grade	6	F	1	93	90.32	135	96.30	132	89.39	122	77.05	84	80.95
			2	180	52.22	127	81.89	126	68.25	160	51.88	194	43.81
		P	3	196	66.84	181	88.40	199	73.87	248	54.03	196	63.78
			4	73	83.56	100	100.00	86	89.53	4	100.00	63	87.30
	7	F	1	101	83.17	147	87.07	143	80.42	140	72.14	85	82.35
			2	226	45.13	142	71.13	146	55.48	166	44.58	235	42.98
		P	3	163	66.26	188	78.72	188	71.28	222	57.66	175	57.71
			4	43	86.05	56	91.07	56	82.14	NA	0.00	36	52.78
	8	F	1	140	91.43	198	100.00	198	89.90	165	89.09	144	86.11
			2	182	34.07	82	100.00	83	54.22	135	39.26	171	29.24
		P	3	149	67.79	175	91.43	177	76.27	139	73.38	161	57.14
			4	82	81.71	98	91.84	95	88.42	111	72.07	73	73.97

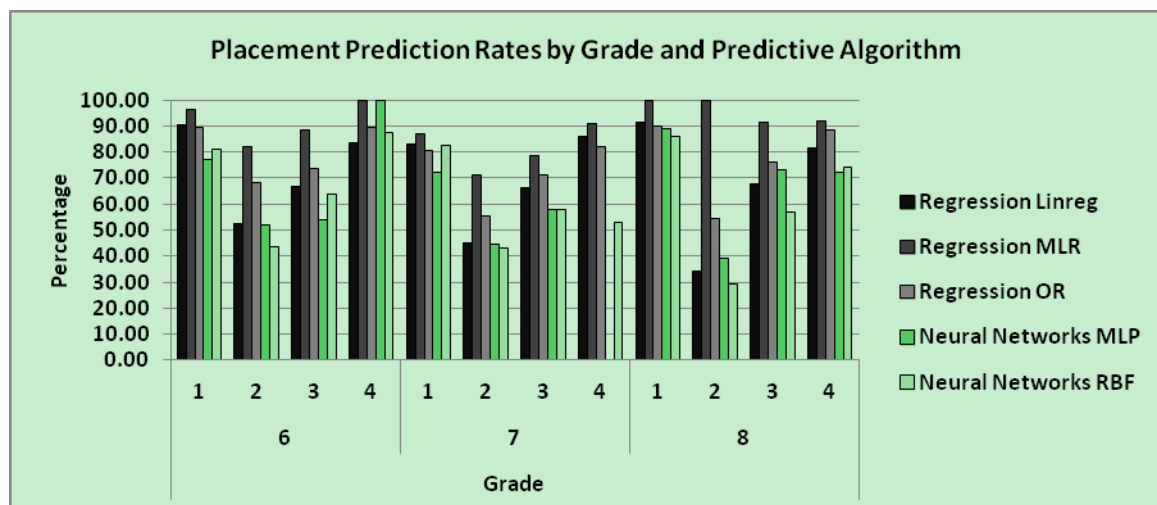


Figure 2. Prediction rates as percentages for the AIMS Mathematics placement rating using all regression models

5.3 Prediction of the AIMS Mathematics Pass/Fail Rating

Tables 7 and 8 as well as Figure 3 display the results of predicting the AIMS Mathematics Pass/Fail rating using all regression models. All correlations (R values) are significant at the 0.01 level and indicate strong positive correlation between the actual and predicted values for each grade level and regression model. Since pass fail rating is a variable with two categorical levels, Regression Model's 2 (MLR) and 3 (OR) produce identical results and performed better than Regression Model 1 (Linear Regression).

Table 7. Correlations pertaining to the prediction of the AIMS Mathematics pass/fail rating using all regression models

Grade	Attribute	Regression			Neural Networks	
		Linear Regression	MLR	OR	MLP	RBF
6	R	0.75	0.85	0.85	0.79	0.76
	N	543	543	543	537	540
7	R	0.69	0.82	0.82	0.64	0.66
	N	533	533	533	531	527
8	R	0.78	0.94	0.94	0.75	0.77
	N	553	553	553	549	548

Note. All correlations are significant at the 0.01 level.

Table 8. Prediction rates as percentages for the AIMS Mathematics pass/fail rating using all regression models

Grade	Attribute	Regression						Neural Networks			
		Linreg		MLR		OR		MLP		RBF	
		N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
6	F	254	87.80	265	92.45	265	92.45	277	87.36	264	87.88
	P	289	85.12	278	92.45	278	92.45	260	91.54	276	87.68
7	F	303	83.17	298	90.27	298	90.27	299	82.27	323	80.50
	P	230	83.91	235	91.49	235	91.49	232	81.47	204	87.25
8	F	287	88.15	278	97.12	278	97.12	263	89.73	284	88.03
	P	266	89.85	275	96.36	275	96.36	286	85.66	264	89.39

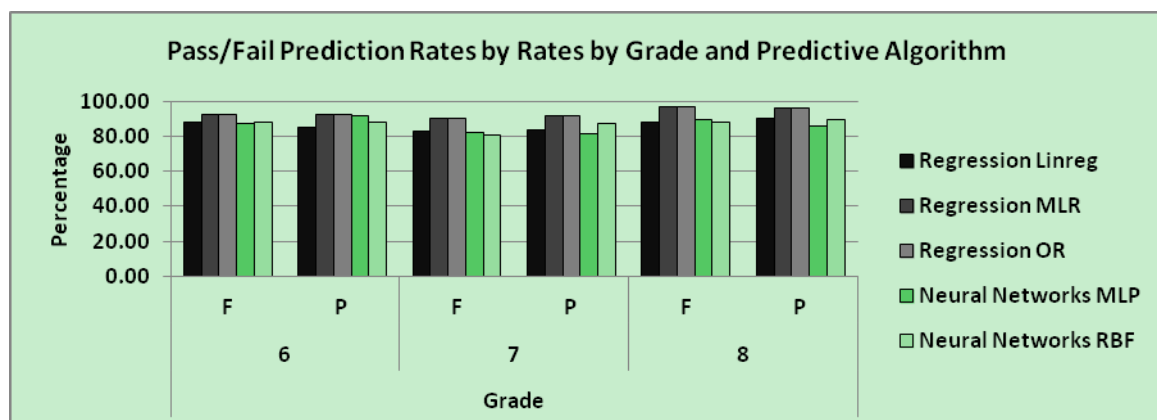


Figure 3. Prediction rates as percentages for the AIMS Mathematics pass/fail rating using all regression models

## 6. Discussion of Results

A considerable amount of time and taxpayer money is spent every year on the schooling system in the U.S. Therefore, students, parents and educators naturally desire to see favorable outcomes in the statewide high-stakes tests. Thus, an early detection system could be of immense assistance to students, parents and teachers to take the necessary measures when there is adequate time to prepare for the statewide high stakes tests so as to help improve the performance of students in such tests. Many schools employ third party testing instruments to monitor student progress throughout the school year. Using the results of such instruments that are already available, it is possible to predict student performance in the statewide high-stakes tests and take the necessary measures in a timely conduct.

In this paper we have shown that the scaled score, the placement rating and the pass/fail rating in the AIMS Mathematics Test can all be predicted using the Math Pretest, Math1, Math2, Math3 and Math Posttest results obtained from a third party testing system. It should be noted that the ultimate goal in making predictions is to use the data and predictive algorithms for the *current school year* as a means to predict how students will perform in the *next school year*. This is possible if and only if the content of the statewide high - stakes tests stay the same and the school district decides to use the same third party testing system throughout the next school year.

Prediction naturally becomes more accurate when more data is available. However, by changing the predictors used as they become available, it is possible to make a reasonable prediction throughout the school year. For instance, when the Math Pretest is available, a predictive model can be created to predict the AIMS results from the Pretest results only; or, when the Math Pretest and the Math1 test results are available, the AIMS results can be predicted from the Math Pretest and the Math1 test results; or, when the Math Pretest, the Math1 and the Math2 test results are available, the AIMS results can be predicted from the Math Pretest, the Math1 and the Math2 test results, and so on. Ultimately, the purpose is to pinpoint the students who are at the risk of failing the high - stakes test and introduce intervention programs targeting such students before it is too late.

In this study, we also showed that the results of prediction differ based on the method selected for performing the prediction and the nature of the predicted value. When, for instance, the AIMS scaled score, which is continuous in nature, is predicted, multiple linear regression and multilayer perceptron (MLP) models performed comparably and both models outperformed the radial basis function (RBF) model. On the other hand, when the predicted value is categorical in nature, Multinomial Logistic Regression (MLR) and The CO—Ordinal Logistic Regression (OR) outperformed Multiple Linear Regression. Please note that regardless of the nature of the predicted value or the method of prediction, we have observed strong positive, and statistically significant Pearson correlation values between the actual and predicted values of the statewide high stakes test scores in middle school mathematics.

In this study we also introduced Neural Networks as an alternative to regression based methods in order to predict statewide high-stakes test results for middle school mathematics using the scores obtained from third party testing instruments throughout the school year. The neural network models employed are the two commonly used ones, namely, the Multilayer Perceptron (MLP) and the Radial Basis Function (RBF) models. Simulations yielded strong, positive and statistically significant Pearson correlation values, between the actual and predicted values of the statewide high stakes test scores in middle school mathematics.

Please note that Neural Networks and regression based methods are expected to perform comparably when data satisfies the assumptions required by linear regression; this has indeed been the case depicted through the findings of this study. The assumptions required by linear regression are, normality, linearity, and homogeneity of variance which, in a perfect world, would be fulfilled by the nature of the data available and to be analyzed. In this study, data satisfied the assumptions of regression and as expected regression based methods performed comparably or slightly better. However it must be noted that Neural Networks do not necessitate any of the assumptions of linear regression to be met. As a matter of fact, data in general cannot be expected to meet the assumptions of normality, linearity, and homogeneity of variance. When it does not, i.e. when the assumptions of linear regression are not met, Neural Networks are expected to capture the nonlinear nature in the data in a much superior manner and thereby give rise to better results than those to be yielded by regression based algorithms.

Last but not the least, although the predictive analyses were performed in the context of middle school mathematics, the suggested regression models can be applied readily to other grade levels and content areas as well. The only requirement is that there should be strong positive correlation between the predicted values (such as AIMS scores) and the predictors used (such as the results of third party testing instruments) and that the student profile, instruments, and content of high stakes tests do not change dramatically from one school year to



the next.

### References

- Braden, J. P. (2002). Best practices for school psychologists in educational accountability: High-stakes testing and educational reform. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, pp. 301-320). Bethesda, MD: National Association of School Psychologists.
- Fuchs, L. S., & Fuchs, D. (2002). *What is scientifically-based research on progress monitoring*. National Center on Progress Monitoring.
- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency - based indicators of foundational reading skills for third grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288. [http://dx.doi.org/10.1207/S1532799XSSR0503\\_4](http://dx.doi.org/10.1207/S1532799XSSR0503_4)
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for Practice in the 21st century. *School Psychology Review*, 30, 420-437.
- Kent, R. (2001). *Data Analysis and Data Construction for Survey Research* (1st ed.). PALGRAVE, N.Y.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). McGraw-Hill Irwin.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29, 4-16.
- Laffey, J., Tupper, T., Musser, D., & Wedman, J. (1998). A computer-mediated support system for project-based learning. *Educational Technology Research and Development*, 46(1), 73-86. <http://dx.doi.org/10.1007/BF02299830>
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Review*, 33, 193-203.
- No Child Left Behind (NCLB). (Unknown). Act of 2001. Pub. L. No. 107-110, Psychology H.R. 1, 115 Stat. 1425.
- Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3-14. <http://dx.doi.org/10.1080/00220670209598786>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).