Using Neural Networks and Support Vector Machines to Predict School Performance in Oklahoma

Tom Curran

June 6, 2018

Abstract

The A-F Report Card is Oklahoma's school performance evaluation tool. The Report Card uses test subject grade test scores for end of the year tests to evaluat if a school's students are performing at the expected level. Schools with students not performing at expectations recieve a low letter grade (e.g. F) while schools with high performing students recieve a high letter grade (e.g. A). Because Oklahoma's quality of education and educational outcomes have been declining for the past several years, the importance of identifying schools at risk of reciving a low grade on the report card has become very important. However, because of beuacratic inefficiencies teachers, school, state and district administrators do not recieve the necessary information before the start of the school year. The timing ultimately how schools administer support services as well as what interventions that schools recieve, as well as student. In sum, lack of availability data induces and reinforce cylces of low quality education. As such, this paper examines the effectiveness of neural networks and support vector machines as predictive frameworks using socio-economic, demographic and workforce data to predict school's performance on the A-F Report Card.Developing these frameworks can induce more effect and timely interventions for both schools and students. We find that Support Vector Machines using linear kernels are the most effective predictive framework, especially for Middle Schools.

1. Introduction

In Oklahoma, education has become one of the most hotly debated topics in recent memory. The state consistently ranks as one of the worst performing and as having some of the worst educational outcomes in the U.S. Recently, as of Spring 2018, a statewide teacher strike took place for several weeks to protest the extremely low teacher pay and recent cuts to education funding. Meanwhile, 16.3% of the state is living in poverty (the 9th highest of all states and Washington, D.C.) and 22.6% of children in Oklahoma live in poverty (10th highest of all states and Washington, D.C.). [1] [2]

A battery of research shows the persistent and undeniable relationship between education and poverty. Economist and policy researchers have shown that more years of quality education induces a positive effect on lifetime earnings and individual well-being. While psychologist and developmental researchers have shown how poverty can influence the outcomes of students. All of this is to say that the education landscape in Oklahoma is facing a critical issue when it comes to using education as a means to alleviate poverty in the state. Continuous funding cuts induces lower quality schools which induces poor outcomes for students, which are

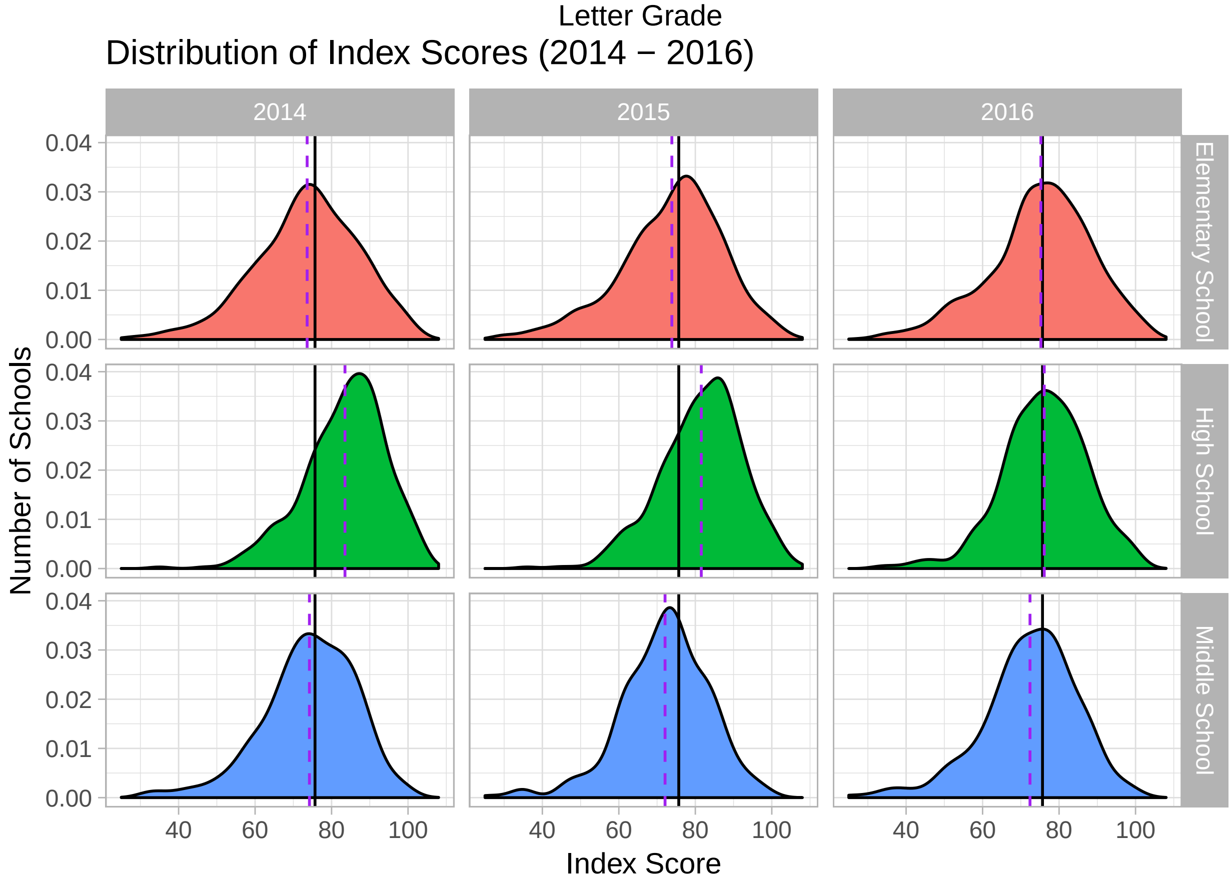
[1] "State-by-State Indicator Map." Talk Poverty, talkpoverty.org/indicator/listing/child_poverty/2011.
[2] "State-by-State Indicator Map." Talk Poverty, https://talkpoverty.org/indicator/listing/poverty/2017

amplified for students living in poverty.

In order to rectify poor school performance in the state, the Oklahoma State Department of Education developed their A-F Report Card. The A-F Report Card is an evaluation tool that grades an individual school's performance on the subject-grade test score performance of the previous year for students overall and the bottom quartile of students. Additionally, schools are awarded bonus points for different milestones, such as high attendance rates and (for High Schools) college entrance exams taken. Points are then totaled to calculate a school's Index Score, which is then assigned a letter grade like a traditional report card. The highest performing schools receive and "A" (Index Score above 90) and the lowest performing schools receive an "F" (Index Score below 60). Schools that consistently have low achievement are at risk of being taken over by the state but also fail to attract necessary teacher and administrator talent. It is the latter of these effects that compounds the cycle of poor educational out comes and poverty.

\

Distribution of Letter Grade Groups (2014−2016)
2014
2015
2016
Elementary School
High School
Middle School
Count of Letter Grade
0
100
200
300
A
B
C
D
F
Letter Grade



Distribution of Index Scores (2014 − 2016)
2014
2015
2016
Elementary School
High School
Middle School
Number of Schools
0.00
0.01
0.02
0.03
0.04
40
60
80
100
Index Score

1.1 Problem Statement

The Oklahoma State Department of Education has not been immune to the budget cuts imposed by the state legislature. The cuts to the education budget has induced a lack of governmental administrators and the bureaucratic infrastructure to manage data collection process required to calculate school's perforamnce on subject-grade level tests.

As a result of this, data collection and assessment are enormously time consuming a bordering on being cost prohibitive. The A-F grades and test score results often times are not received by schools till well into the school year. The timing becomes critical for schools because test scores are one (but certainly not the only) way that school administrators can make effective decisions when deciding which students are assigned to which class and teacher. These decisions are critical not only to impacting individual student performance, but also to improving the school's A-F Letter Grade. The core issue is simply lack of data to drive effective and timely decision making.

The bureaucratic issues revolving around data collection is a serious impediment for improving school performance, particularly those schools that serve high needs students and communities. Paradoxically, the Oklahoma State Department of Education retains a wealth of other historical socio-economic, demographic and workforce data. Therefore, the question must be asked can using socio-economic, demographic, and workforce data in a predictive frameworks (Neural Network and Support Vector Machine (SVM)) accurately forecast a school's Letter Grade on the A-F Report Card?

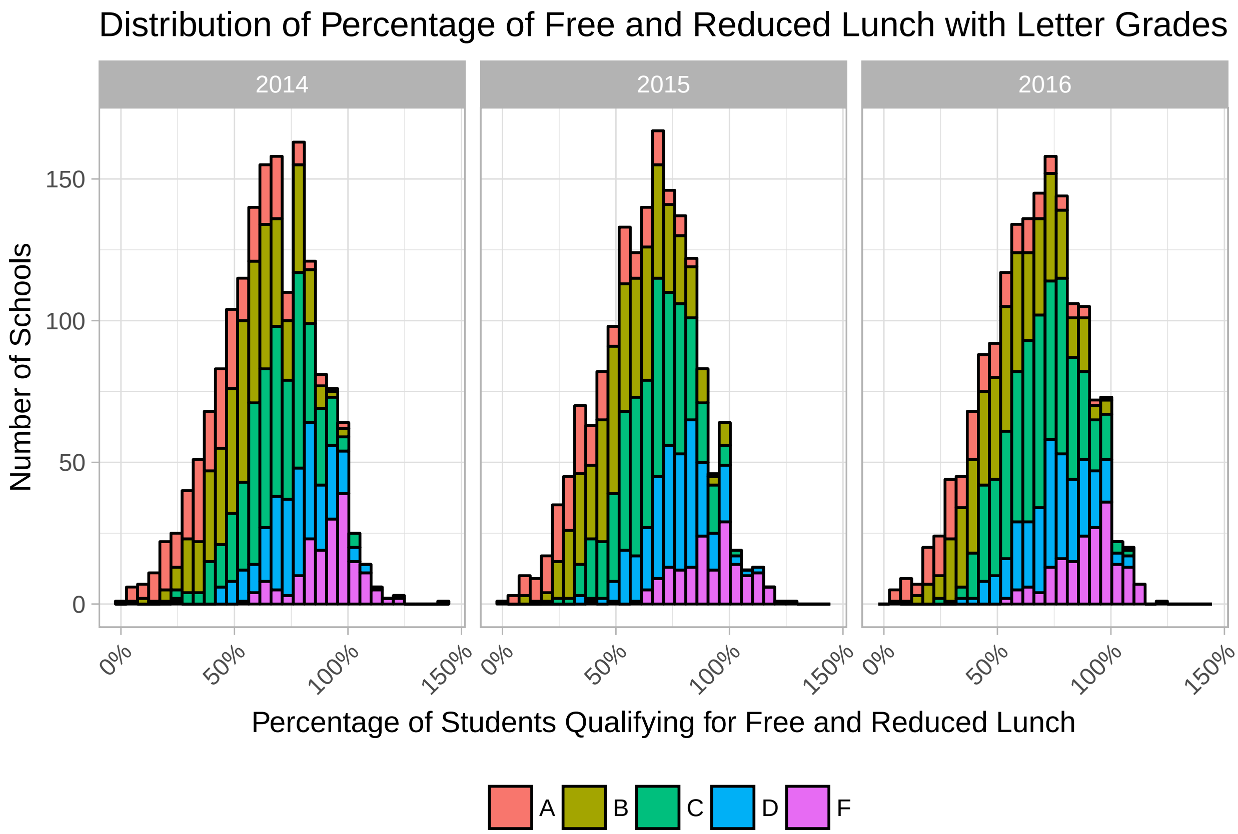
Based on the available literature and understanding of classroom, student and school dynamics the socio-economic, workforce and demographic data will

accurately predict a school's letter grade. Furthemore, I believe that neural networks will provide an effective predictive framework for using the data.

1.2 Motivation

The primary motivation for this research and excercise is to develop a methodology or framework that detects schools at risk of failing on the A-F Report Card. In being able to identify schools most at risk for failing district, state and school administration can not only be aware of the risk, but potentially provid timely and effective intervention.

Timely prediction becomes incredibly important because schools can vary so widely in their characteristics. Because of this variation, there is not always a panceea for improving school performance on the A-F Report Card. For example, a school with a high amount of English Language Students faces different challenges than a school with a lack of Special Education Teachers. However, by being aware that a school is at risk, can trigger needs assessments for the school and therefore provide impactful interventions. Furthermore, the benefit to using these data sources is that the are readily available from sources that do not suffer from the same beuacratic issues that test scores do. At its core, creating a predictive framework can help those students with the most to gain from improving education and schools. As you can see from thr graph below, schools that serve populations at risk, in this case students on Free and Reduced Lunch, are more often than not failing schools.



1.3 Goals:

The goal of this paper of this paper is 2 fold:

1. Develop a predictive framework and model that accurately predicuts a school's A-F Letter Grade
2. Use non-test data to build and apply the framework in order to be able to identify schools for necessary interventions in a timely manner.

2. Literature Review:

There is a wealth literature regarding student achievement and school performance from a variety of fields, including (but not limited to) economics, psychology and public policy program evaluation. The latter of these fields is perhaps the most prominent and rich source of literature. The nature of program evaluation literature and practice is to measure and evaluate the impact programs or policies have on intended audience. While there are many complicated and nuanced mathematical approaches that evaluations can take, a majority of them are based on regressions.

One of the primary issues with taking a purely program evaluation approach to predicting school grades is that the supporting literature cannot seem to agree on causal relationships between school characteristics and school outcomes. For example, the prevailing school of thought is that smaller student teacher ratio leads to greater gains in student achievement, particularly among low-income students. One of the earliest research papers on the relationship between class size and student achievement was conducted by Gene Glass and Mary Lee Smith in 1979 [Glass and Smith, 1979] [Huebner, 2013]. Research studies like the Tennessee STAR experiment furthered Glass and Smith's findings, pushing policy makers to focus on reducing class size as a mechanism for promoting student achievement. [Finn and Achilles, 1999] [Mosteller, 1995] [Grimmer, 1999]

However, there has also been significant literature that counter's the class size argument. The research conducted by Caroline M. Hoxby shows that natural variation in populations result in class size having little to no effect on student achievement [Hoxy, 2000]. Christopher Jepsen and Steven Rivkin find that there is a significant tradeoff between teacher quality and class size. While Jepsen and Rivkin find that class size does impact achievement the achievement is negatively affected when there the share of teachers in a school have little or no experience [Kepsen and Rivkin, 2009].

In addition to program evaluation literature, the rise of Education Data Mining (EDM) and Learning Analytics (LA) has provided a variety of new approaches and techniques to apply to education data. Combining these two approaches can greatly inform the research goals of this paper.

When it comes to EDM and LA, it is important to select the appropriate tools so that proper methodologies can be used in educational data. Slater et al suggest several such tools for each step of the EDM process, from data pre-processing, to analysis and visualization. Among the author's suggested set of tools are Python and Jupyter notebook. Slater et al share several insights as to the benefits of using Python and the Jupyter environment. Advantages include python's vast resources for manipulating data sets for pre-processing, python notebook's record of analyses conducted and perhaps most importantly, its computational power, especially when it comes to large data sets. Slater et al continue their article by summarizing the benefits and drawbacks to several tools when conducting text mining, network analysis or visualizations projects. Slater et al's review of EDM and LA tool kits provide a detailed landscape for tools to choose from that is best suited for the project at hand.

The survey of tools is complimented by Behdad Bakhshinategh et al's work Educational data mining applications and tasks: A survey of the last 10 years. Bakhshinategh breaks down EDM applications into distinct subgroups, among them are what he calls "Decision Support Systems". In essence, Bakhshinategh et al make the case for data mining and learning analytics to help educators and administrators make decisions in an education context. This subgroup's ultimate goal is to engage stakeholders by using data to provide feedback, creating alerts, generating recommendations, and enhancing courseware. In Bakhshinategh says that the target of the decision support systems is primarily focused on the teacher but can be applied to administrator. In the case of this research paper, the target population are policy makers. Since decision support is concerned with any and all stakeholders, policy makers are a critical component to that, it is reasonable to apply Bakhshinategh et al's framework to the question at hand for entire schools.

Modeling and predicting performance is particularly challenging in the context of education. Particularly because there are many opportunities to introduce bias into a model. Lou et al provide a statistical framework in which to reduce variables that they call "discrimination aware classifiers". The issue becomes prevalent when creating rule-based groups of students that necessarily depend on sensitive characteristics. For example, creating achievement levels that necessarily depend on student's gender and/or demographic identity. As Lou et al state, "It is desirable to keep the sensitive attribute during the training of a classifier to avoid information loss but decrease the undesirable correlation between the sensitive attribute and the class attribute when building the classifier". [?] In other words, it becomes increasingly necessary to use sensitive attributes in order to avoid information loss but creating decision rules from sensitive information can lead to discriminatory practices. As such, Lou et al suggest these sensitive attributes be used as an "information carrier and not a distinguishing factor". As such, Lou et al provide a method for developing a Discrimination Aware Association Rule classifier (DAAR) that measure the discrimination severity of a given rule. These DAAR rules are used to predict student achievement in a computer science class. They show that there is limited loss of predictive accuracy between non-discriminatory aware rules and the DAAR method. [?]

DAAR rules are of particular importance given the unit of observation for this research paper. Much like students in Lou et al's paper, districts and schools in Oklahoma vary widely in terms of their financial characteristics, teaching staff, and most importantly students along with many other characteristics. For example, in Oklahoma City Public Schools 31.9% of students are English Language Learners (ELL) [of Educational Quality and Accountability, 2016c] while the school district directly to the north of OKCPS has only 4.4% ELL students [of Educational Quality and Accountability, 2016a]. Even within districts there is great variation in populations. Within OKCPS there is great variation in school populations. Star Spencer High School has 2.5% ELL [of Educational Quality and Accountability, 2016d] students where as U.S. Grant High School has 30.8% ELL students [of Educational Quality and Accountability, 2016b]. Lou et al's methods for minimizing discriminatory variables becomes increasingly necessary to build the model, as well as produce a model that can be deployed in a policy setting.

Yorek and Ugulu provide a qualitative definition of artificial neural networks and their applicability in the education space. Their definitions prove especially useful since they attempt to create an artificial neural network (ANN) to measure student attitudes and qualitative data. An artificial neural network is a mathematical model that takes after the biological function and physical structure of neurons in the brain. Yorek et al concisely express the advantages of using ANN because they can be "used to model complex relationship without using simplifying assumptions which are commonly used in linear approaches" [Yorek and Ugulu, 2015]. Furthermore, ANN has the advantage of being able to represent linear and non-linear relationships that can be extrapolated directly from the data. Though York and Ugulu are using qualitative data, their outcome of interest is still a categorical variable. In the context of this research paper, the output is categorizing a student's "attitudes towards nature" using Kellert's typologies. The research approach in this paper provides a conceptual framework to predict the outcomes of school district performance on Oklahoma A-F Report

Card.

While Yorek and Ugulu provide a valuable conceptual framework for which to understand and apply ANN, Kash Barker et al provide an excellent example of how to apply ANN as well as Support Vector Machines to institutional decision making. The research conducting by Kash et al fall directly under the “Decision Support Systems” that Bakhshinategh et al describe in their assessment of application of EDM and LA. Kash et al use demographic and survey data collected from the University of Oklahoma and use ANN and SVM (support vector machines) to predict a student’s probability of graduating on time. [Barker et al., 2004]

In addition to providing a useful model for this research paper, Kash et al’s sampling method provides additional support and ideas for approach district level data. Kash et al run ANN and SVM (each data set is split into training and testing sets) on a combined data set where all student cohorts are randomized and split in ANN and SVM models, is run again on training and test data sets that are split between years (i.e. training on the 1995 cohort and testing on the 1996 cohort) and finally testing and training within years. In developing a predictive model for Oklahoma schools and school districts, training and testing the data in a similar way could provide additional useful insights. This is especially true due to the fact that the data sources to be used collect information on school districts annually. In sum, Kash et al provide a precedent for using ANN and SVM to make institutional decisions. The authors are a prime example of how combine proven socio-economic indicators of academic success with state-of-the-art procedures and methodologies. [Barker et al., 2004]

3. Data

3.1: Data Sources

3.1.1 A-F Report Card

The A-F Report Card was passed into law by the 2011 Oklahoma Legislature and implemented by the OKSDE in 2012. According to the 2013 documentation by the State Department of Education, the A-F Report Card "is designed to incentivize schools to strive for and reach high levels of college and career readiness". The A-F report card assesses school quality and performance in three main categories: overall grade-level performance on subject grade state assessments, grade-level performance of 'low performing students' (i.e. bottom quartile) on subject grade state assessments and school (i.e. High School, Middle School, Elementary School) characteristics. The latter of these areas can be metrics like graduation rate and college entrance exam participation for high schools, or attendance for Middle or Elementary Schools. It should be noted that there were several issues with the 2013 A-F Report Cards that altered the grading guidelines for 2014, 2015 and 2016. As such, the 2013 Report Card was not used in. this analysis. Furthermore, the 2017 A-F Report Cards have not been published due to ongoing legal issues.

The points earned on these fields are summed to create a school's Index Score and a letter grade is assigned to the school. For example, a high performing school that has an Index Score of 95 earns an "A" letter grade, a school that has a 70 Index Score earns a "C-" and so on. For our purposes, we are interested in predicting the letter grade. To simplify the letter grade variable, we removed the use of the "+" and "-", changing the outcomes from 15 possible categories to 5 possible categories. For example, any school earning greater than a 90 Index Score was now always an “A”, between 80 and 89 was a “B”, between 70 and 79 was a “C”, between 60 and 69 was an “D”, and any school below 60 was an “F”.

From the available data, we only used the Index Score and Letter Grade. The omission of the other data in the A-F report card was either because a more reliable source was found for the same variable or it was data that we believed could induce collinearity and inflate the training model. The information is freely available to the public through the Oklahoma State Department of Education’s Transparency Index. For complete source and citation please see the works cited list.

The table below provides a brief description of the variables from this data source as well as summary statistics.

Table 1: Description of Selected Variables from A-F Report Card

Variable	Description	Variable Code
Index Score	Ranging from 0 to 100, is the sum of all points earned on the A-F Report card.	IndexScore
Letter Grade	A categorical representation of the Index Score which follows the typical grading structure (i.e. Index Score >= 95, A+, Index Score between 89 and 85 is a B+)	LetterGrade
Full Code	A unique composite key used to identify a given school across multiple data sources	SchoolCode
Exam Count	The number of students taking a given subject test (the Exam Count applies to math and reading for all grade 3 through 8, and science, history and writing from grades 5 and 8)	ELA_ExamCount
<Subject> Performance Index	A 0 through 100 score representing the share of students passing the given subject tests	ELA_IndexScore
<Subject> Letter Grade	A categorical representation of the Performance Index representing the number of students passing the given grade level subject test	ELA_LetterGrade
Attendance Rate	The school’s overall attendance rate	AttendenceRate
Attendance Rate Bonus	The bonus points awarded for different levels of attendance rate	AttendenceRateBonus
Graduation Rate	The proportion of graduated students to the number of enrolled students. This is an example of bonus points that are only applied to High Schools. Elementary Schools and Middle Schools are automatically credited these points to correct for differences	GraduationRate

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
IndexScore	4,941	75.671	13.032	25	108

3.1.2: Office of Educational Quality and Assessment

The Office of Educational Quality and Assessment is a reporting organization housed within the Oklahoma State Department of Education that publishes yearly reports at the school and district level. The reports contain socio-economic, workforce and demographic information about each school, as well as reporting that percentage of students meeting satisfactory achievement on subject-grade level state assessments. This report differs from both the A-F Report Card data and the OKSDE Transparency Index in that it it sources its non-assessment data from the American Community Survey (and other data sources). As such, the report provides information such as average salary based on the school's census tract, not the average salary for the entire district. It should be noted further that this data is in no way connected to the A-F Report Card. Below are a selection of variables and their constructs sourced from the OEQA reports. The OEQA data can be acquired through two different channels. The first, more accessible channel, is through the pdf reports that are published by the OEQA on their website. The second channel is by requesting the flat file for a given year. The flat file contains the same information as each school's pdf report just in a tabular format that is more readily accessible by computational tools.

Table 3: Description of Selected Variables from OEQA Report

Variable	Description	Variable Code
School Code	A unique composite key used to identify a given school across multiple data sources	SchoolCode
Average Income	The average income for the census tract based on the address of the school. Sourced from the American Community survey	AvgIncome
Poverty Rate	The proportion of households considered "in poverty" in the school's census tract	PovertyRate
Average Teacher Salary	The average teacher salary for the given school. Each school typically follows the recommended salary schedule provided by OKSDE	AvgSalary
Average Years Experience	The Average number of years of experience each teacher has in each school	AvgExperience
Unemployment	The % of people unemploymed in school's census tract	Unemployed
% Free & Reduced Lunch	The porportion of students qualifying for Free & Reduced Lunch programs	FRL
Number of Special Ed Teachers	Number of Special Education Teachers	SpEdTeachersFTE
Number of Counselors	Number of Counselors employed at school site	CounselorsFTE
Number of Administrators	Number of principals, assistant prinicpals and other administrative position at a school site	AdminFTE

Table 4: Summary Statistics for Selected Variables from OEQA

Statistic	N	Mean	St. Dev.	Min	Max
Average Income	4,941	59,725.360	14,831.540	25,047.000	219,858.300
Average Property Value	4,941	52,587.960	47,835.550	0.000	618,737.100
Poverty Rate	4,941	0.170	0.070	0.020	0.560
Number of Teachers	4,941	21.402	16.186	0.000	171.870
Average Teacher Salary	4,941	44,265.840	3,014.855	0.000	64,204.370
Average Years Experience	4,941	12.933	3.361	0.000	27.460
Number of Special Ed Teachers	4,941	2.558	2.939	0.000	27.520
Number of Counselors	4,941	0.924	1.048	0.000	12.490
Number of Administrators	4,941	1.467	1.405	0.000	22.010

3.1.3: Oklahoma State Department of Education Transparency Index

The Oklahoma State Department of Education has published a "Transparency Index" on their website. The transparency index is a repository of education data that was used to collect demographic information about the enrolled students in Oklahoma public schools. As part of the ‘Student Count’ section of the transparency index, OKSDE has published several years of student demographics on the ‘Enrollment of Oklahoma State Public Schools’ section. The data is provided in a .csv format, wherein the document provides the count of students by demographic and gender for each school and grade.

Table 5:

Statistic	N	Mean	St. Dev.	Min	Max
White%	4,941	0.531	0.196	0.000	0.965
Hispanic%	4,941	0.127	0.149	0.000	0.906
AfricanAmerican%	4,941	0.069	0.130	0.000	0.945
Asian%	4,941	0.012	0.022	0.000	0.231
PacificIslander%	4,941	0.003	0.011	0.000	0.268
TwoRaces%	4,941	0.073	0.072	0.000	0.748
Total	4,941	385.909	345.729	16.000	3,778.000

3.4 Data Collection:

The data described in the previous section was collected and loaded into a MySQL database maintained by the Oklahoma Public School Resource Center (along with several other data sources). To collect the necessary data, MySQL queries were written that aggregated these different data sources and written into csv files. The data sources were compiled into three different files based on their sources. In other words, the A-F data was queried and written in one file, the OEQA file in another and the demographic information and the final.

The demographics data as sourced from the OKSDE Transparency Index was aggregated in MySQL as row wise percentages of the total number of students in each school. In other words, the query totals the number of each demographic for male and female, divided by the total number of students in the school (sum of Total column for each grade) for each year

2014 through 2016. The data was aggregated as a proportion in order to avoid issues of scale due to the wide range of total student population in schools across Oklahoma.

For several school districts with extremely small number of students, data was redacted from public record. This is a rare occurrence and due to the small number of schools where this occurred, those records were dropped from the analytical file.

4. Methods

4.1 Neural Networks

Barker et al describe a Neural Network as “a data processing system consisting of a large number of simple, highly interconnected processing elements (artificial neurons) in an architecture inspired by the structure of the cerebral cortex of the brain” (cited from Tsoukalas and Uhrig (1997)). At its core, the neural network is a handy way of capturing interactive effects of input variables. A neural network can provide a more reasonable means to predicting categorical outcomes in high dimension data sets. Depending on the which type of school we are predicting the outcome for (letter grade) we can have anywhere between 85 to 115 variables. In this case, especially with a high dimension data set, we cannot take the ‘kitchen sink’ regression approach. The kitchen sink approach is simply throwing all 85 to 115 different variables into a regression predicting some outcome, either categorical or continuous. This is approach is undesirable is it leads to overfitting and bad predictive models but also is computational and time expensive. Furthermore, performing stepwise regression model selection is extremely time consuming. Therefore, neural networks provide a great advantage

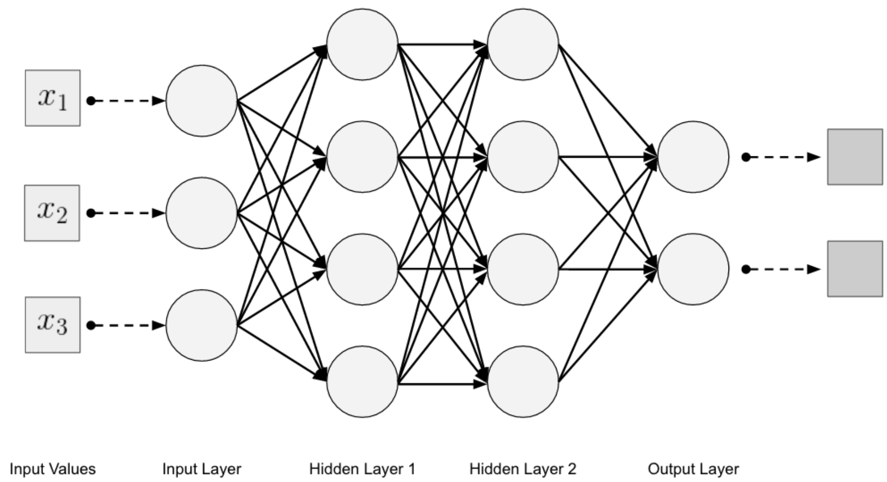


Figure 1: Example of Basic Neural Network Model [3]

that not only compute through the complexity of model selection but also capture non-linear relationships between variables that step wise regression model selection does not always yield. The diagram below provides an effective summary of what is happening in the neural network for our analysis. Using the middle school data as example, the first layer is simply the features in the data set. The output in our case would be the multiple letter grades possible. For simplification, we can call these 'A', 'B', 'C', 'D' and 'F', where each letter includes the plus and minus variations (except for those schools that are F that does not have the plus or minus variation).

The hidden layer, as Barker et al describe, "consists of several predetermined hidden nodes connected to the input layer with a set of weights". This hidden layer is again meant to capture and learn from possible non-linear relationships where each connection is weighted and then output a likelihood of being one of the five grades possible.

[3] https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.safaribooksonline.com%2Flibrary%2Fview%2Fdeep-learning%2F9781491924570%2Fassets%2Fdpln_0201.png&imgrefurl=https%3A%2F%2Fwww.safaribooksonline.com%2Flibrary%2Fview%2Fdeep-learning%2F9781491924570%2Fch04.html&docid=bhuS7mMOXDfCzM&tbnid=FTDyB4T1p8vFCM%3A&vet=10ahUKEwjgxdaK5L_bAhWk34MKHSlICIAQMwi0ASgPMA8..i&w=1030&h=573&bih=1204&biw=2160&q=neural%20network%20diagram&ved=0

Neural Networks were chosen as one of the potential frameworks for several reasons. First, the data that we set out to use is inheretly high dimensional. Neural networks are able to handle the high dimensionality and variation much more efficiently and effectively than other methods. Using techniques like LASSO regression with such a high dimensional data set would be hugely expensive both in terms of computing power and time. Since the focus of the framework is the outcome, we can forgo a quantitative measure of a single variable and its impact on the letter grade.

4.2 Support Vector Machines

A Support Vector Machine is a powerful classification algorithm that has become popular in the last several years because of its ability to handle data sets that are large volume and high dimensional. It is most commonly used to predict binary classifiers but can be used for multi-class classifiers. The goal of support vector machines is to find a hyperplane(s) that separates the data into distinct groups. Using the binary classification as an illustrative example, if we think of a 2-dimensional data (n = 2) set there will be n-1 hyperplane(s) that separate the data into two distinct groups. The key to having a high performing support vector machine is having a hyperplane that maximizes the distance between the two classes (in the case of the binary data set). We seek a maximized distance because it generalizes better to new, untrained data. In other words, a maximal distance is a preventative measure against overfitting training data.

The image below simplifies the idea of Support Vector Machines. As you can see the solid green line represents the optimal hyperplane because it clearly bisects the blue circles

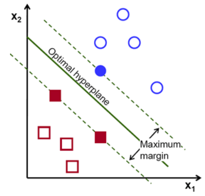


Figure 2: Example of a Support Vector Machine for Two Classes [4]

and red squares, allowing for clear, distinct boundaries and therefore, clear distinct classifiers. You may also noticed the solid filled blue circles and red squares. Those are support vectors. Those support vectors are the closest to the optimal hyperplane while maintaining distinct groups. The support vectors are critical to the optimal margin as changes in these points will change the hyperplane and therefore the maximum margin.

Support Vector Machines are distinct from Neural Networks, as Barket et al explain, because "[Support Vector Machines] solve a convex optimization problem which, theoretically, gives an optimal solution unlike the neural network algorithm that minimizes a non-convex error function". This example of a Support Vector Machine is a simplified scenario intended to illustrate the general concept. However, as is often the case with data sets, there are non-linear relationships that occur and consequently not captured by the linear construct as illustrated above. In such cases, we use kernel functions to attempt to capture these non-linear relationships. The four primary types of kernels used in this research, in addition to Linear, are Radial Basis Function, Polynomial and sigmoid.

Support Vector Machines are preferable to other frameworks again because of the inherent nature of the data. Techniques like Regression Trees are preferred for data sets where the outcome is mutli-class it does not scale well as the number of features increasues. That is to say Regression Trees do not work well with high dimension data sets and since the

[4]https://docs.opencv.org/2.4.13.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

data set we will use has over 60 variables Support Vector Machines are the obvious choice. Furthermore, while SVMs are typically used for binary classifiers, they can still support mutli-class classifiers using methodologies such as the "one-vs=all". Once again, we prefer Support Vector Machines to Tree Classification schemes because we are more concerned with the predicted outcome, not necessarily what is causing the outcome. [5]

5. Analysis:

Once the data was collected and extracted from OPSRC's Database, we scaled all of the data to have a mean zero. Scaling was critical to this research as different variables use very different units (e.g. average salary versus percent white). An initial attempt was made using unscaled data but performed very poorly on nearly every model that was trained and tested.

Once scaled, the data was partitioned into three different groups: Elementary, High and Middle Schools. The intent behind this grouping was to impose some control for the unobservable differences between middle, high and elementary schools. For example, in our data set, there is an unknown effect of having the same teacher throughout the day, as is done in elementary school, versus the effect of having several different teachers for different subject throughout the day. Therfore, in order to control for this and similar unobservable characteristics, it is critical to compare within levels of school and not compare outside of those designations. However, there was not always a clear designation between school types. For example, there are sevearl schools in the state that are a combined elementary and middle school, where as other districts (usually larger ones) have an elementary school and a

[5] https://github.com/UC-MACSS/persp-model_W18/blob/master/Notebooks/SVM/SVM.ipynb

separate middle school. Furthermore, each district has their own definition of which grade constitutes a middle school, but generally middle school starts in 6th grade and ends in 8th grade, but often times elementary schools defined middle school to be only seventh and eigth grade. Because there is no hard and fast rule, a decision was made to call any school housing between 6th and 8th grade students, was a middle school, anything kindergarten through fifth grade was elementary and ninth grade and beyond was a high school.

A key component to conducting this research was computational power. Because Neural NEtworks and Support Vector Machines are computationally expensive, it became increasingly necessary to use cloud computing services. We chose Google cloud compute services using Ubuntu 16.04 with GPU capability running python and keras (a Tensorflow wrapper) with GPU capabilities. Once the data was divided into the three types of schools, each data set was split into a training (75%) and test (25%) set. The training data sets were used to build the classification models for both the neural networks and the support vector machines and used a cross validation approach to minimize bias (K=10). For the neural networks, each training data set was run through an initial baseline model. The baseline model was a basic neural network structure with a single input layer and the "softmax" layer that placed the school into one of the five letter grade categories. The baseline model also used approximately 10% of the training data set as the validation data set. Once the baseline model had been constructed, the trianing data (with the same validation parameter) was run through several other models with differing regularizers and varying number of hidden layers in order to see which combination of attributes yielded the highest accuracy. Then following the training of the models, each model's accuracy was tested against the test data sets produced earlier in the workflow.

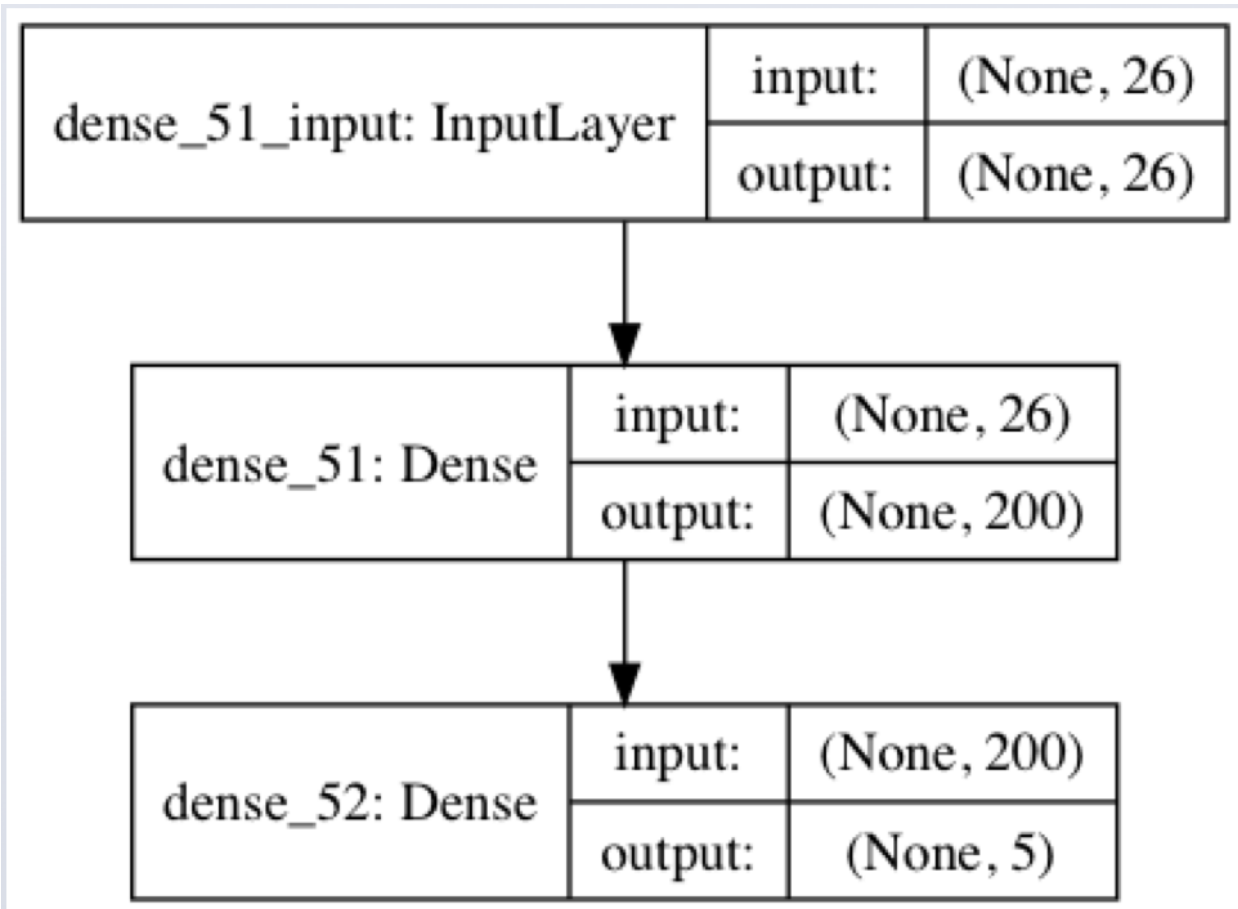


Figure 3: Baseline Neural Network Model

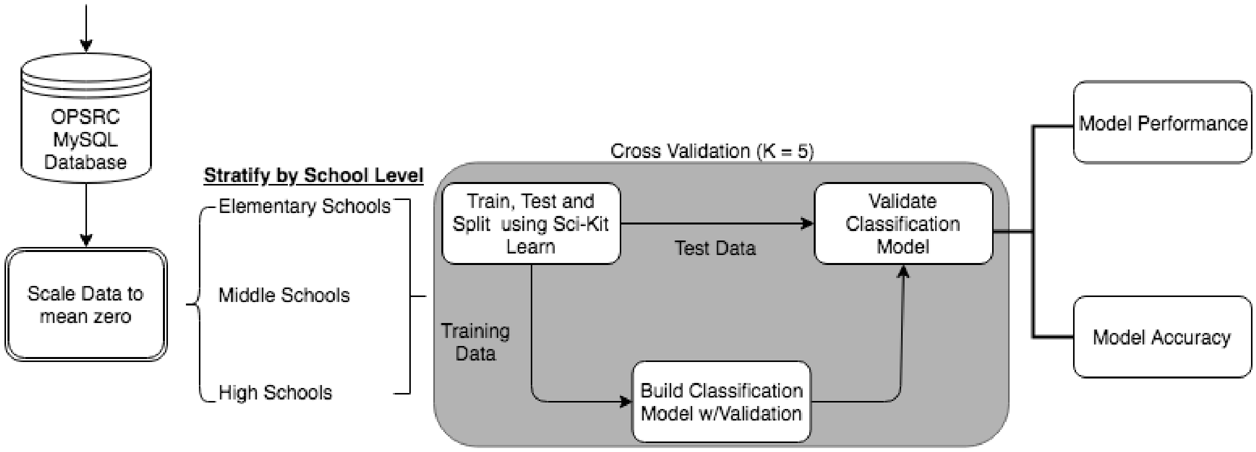


Figure 4: Workflow Diagram

This same cross validated approach was taken with constructing the Support Vector Machines. Using the training and testing data sets, each support vector machine was built with the training data, and it's performance and predicted power evaluated on the testing data set. Again similar to the neural entworks, different framework features (i.e. kernels) were used to see which performed the best. In addition to building the linear, polynomial, sigmoid and rbf support vector machines, the C penalty and gamma parameter were controlled for between uses of different kernels in order to compare performance between them.

6. Results

6.1 Neural Network Results:

The following tables summarize several of the neural network models constructed, trained and tested. The tables display the loss and accuracy using the training and validation data sets, and then the model's performance in terms of loss and accuracy (% predicted correctly) on the testing data sets.

Table 6: Neural Networks for Elementary Schools (Selected Models)

	Model Validation		Model Performance		Model Characteristics	
	Loss	Accuracy	Loss	Accuracy	Penalization	Number of Dense Units
Baseline	1.03	86.19%	3.72	45.89%	None	200
Model 2	.74	75.28%	1.86	48.84%	Dropout (.5)	512
Model 3	1.48	52.67%	1.73	46.62%	l1 regularization (0.001)	512
Model 4	1.12	83.69%	4.21	45.46%	l2 regularization (0.001)	512
Model 5	1.12	83.69%	3.88	44.09%	l2 regularization (0.001)	550

Table 7: Neural Networks for High Schools (Selected Models)

	Model Validation		Model Performance		Model Characteristics	
	Loss	Accuracy	Loss	Accuracy	Penalization	Number of Dense Units
Baseline	1.06	83.51%	3.72	38.22%	None	200
Model 2	.77	75.00%	1.86	40.98%	Dropout (.5)	512
Model 3	1.54	47.88%	1.73	41.61%	l1 regularization (0.001)	512
Model 4	1.01	83.85%	4.21	38.22%	l2 regularization (0.001)	512
Model 5	1.01	83.85%	3.88	36.94%	l2 regularization (0.001)	550

Table 8: Neural Networks for Middle Schools (Selected Models)

	Model Validation		Model Performance		Model Characteristics	
	Loss	Accuracy	Loss	Accuracy	Penalization	Number of Dense Units
Baseline	1.18	86.30%	3.88	50.00%	None	200
Model 2	.52	84.22%	3.25	53.53%	Dropout (.5)	512
Model 3	1.39	62.91%	2.08	40.71%	l1 regularization (0.001)	512
Model 4	.89	86.14%	3.05	49.03%	l2 regularization (0.001)	512
Model 5	.9536	81.63%	2.82	51.28%	l2 regularization (0.001)	550

6.2 Support Vector Machine Results

The tables below summarize the performance of the support vector machines for each school level, as well as for each kernel. The performance of the model is primarily judged on its testing performance.

Table 9: Support Vector Machhines Performance Metrics

	Middle School (% Correct)	
	Training Performance	Testing Performance
Linear	81.11%	70.83%
Poly	100.00%	60.89%
RBF	100.00%	35.90%
Sigmoid	32.24%	38.78%

Table 10: Support Vector Machhines Performance Metrics

	High School (% Correct)	
	Training Performance	Testing Performance
Linear	65.75%	54.99%
Poly	100.00%	49.89%
RBF	100.00%	33.12%
Sigmoid	27.84%	34.39%

Table 11: Support Vector Machhines Performance Metrics

	Elementary School (% Correct)	
	Training Performance	Testing Performance
Linear	66.19%	59.59%
Poly	100.00%	54.11%
RBF	100.00%	32.17%
Sigmoid	32.17%	27.72%

7. Conclusion

Based on the results summarized in the table above, it appears that in general neural networks perform poorly on all three types of schools and support vector machines generally only slightly out perform the neural network models. Support Vector Machines appear to work the best for predicting the A-F Letter Grade for Middle Schools using a Linear Kernel, while, High Schools (using a linear kernel) performed the worst.

Given these results and our goal of predicting school performance outcomes, I believe that we can indeed use socio-economic, workforce and demographic data to predict the A-F Letter Grade for schools. As such, the socio-economic, workforce and demograph data is best utilized with a Support Vector Machine framework that uses a linear kernel. However, though the combination of data and models has shown promise the models, particularly the linear support vector machine model, should be further optimized by changing it's parameters in order to imporove its accuracy.

7.1 Limitations:

Though the support vector machine approach does show promise, and to a lesser extent the neural networks model, there are significant draw backs to using these models. It is true that normal linear or logistic regression may not properaly encapsulate the interactive effects and be subject to omitted variable bias, they do have the distinct advantage of being eaiser to understand. Using a typical regression model found through out program evaluation literature allows for policu makers and administrators to identify statistically significant coefficients and their impact on school performance. Where as with support vector machines and neural networks, it is much harder (and in some cases nearly impossible) to identify the most significant influencer in the model. Drawing the causal relationships leads to more effective and targeted action and is more likely to induce proper intervention than simply being aware that a school is at risk of failing. Furthermore, the results from the Support Vector Machine should be approached with cautious optimism as well because the size of the middle school data set was relatively small compared to the high school and elementary schools.

7.2 Extensions of Research:

Based on the results of both the neural networks and support vector machines, further refinement of the models is needed. Specifically for the Neural Networkswe would like to find additional socio-economic and workforce data to add to the model. In addition, different combinations of layers and penalities could be added in order to increase accuracy. For the Support Vector Machine models, we will continue to use the linear kernal and refine the model by changing the values of the C penalty and gamma parameters.

References

K. Barker, T. Trafalis, and T. R. Rhoads. Learning from student data. In *Proceedings of the 2004 IEEE Systems and Information Engineering Design Symposium, 2004.*, 2004.

Jeremey D. Finn and Charles M Achilles. Tennessee's class size study: Findings, implications and misconceptions. *Educational Evaluation and Policy Anlysis*, 21(2):97–109, 1999.

Gene V. Glass and Mary Lee Smith. Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1(1):2–16, 1979.

David Grimmer. Class size effects: Assesing the evidence, its policy implications and future research agenda. *Educational Evaluation and Policy Analysis*, 21(2):231–248, 1999.

Caroline M. Hoxy. The effects of class size and composition on student achievement: New evidence from natural population variation. *Quarterly Journal of Economics*, 115:1239–1285, 2000.

Richard A. Huebner. A survey of educational data-mining research. *Research in Higher Education*, 19, 2013.

Christopher Kepsen and Steven Rivkin. Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. *The Journal of Human Resources*, 44(1):223–250, 2009.

Frederick Mosteller. The tennessee study of class size in the early school grades. *Critical Issues for Children and Youths*, 5(2):113–127, Summer/Fall 1995.

Office of Educational Quality and Accountability. 2016 district profiles - edmond. Report, Oklahoma State Department of Education, 2016a.

Office of Educational Quality and Accountability. 2016 school profiles - u.s. grant high school okcps. Report, Oklahoma State Department of Education, 2016b.

Office of Educational Quality and Accountability. 2016 district profiles - oklahoma city. Report, Oklahoma State Department of Education, 2016c.

Office of Educational Quality and Accountability. 2016 school profiles - star spencer high school okcps. Report, Oklahoma State Department of Education, 2016d.

Nurettin Yorek and Ilker Ugulu. A cfbpn artificial neural network model for educational qualitative data analyses: Example of students' attitudes based on kellerts' typologies. *Educational Research and Reviews*, 10:2606–2616, September 2015.