Using ITS Generated Data to Predict Standardized Test Scores

Kim Kelly
Worcester Polytechnic Institute
kkelly@wpi.edu

Ivon Arroyo
Worcester Polytechnic Institute
iarroyo@wpi.edu

Neil Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT

This study suggests that the data generated by intelligent tutoring systems can be used to accurately predict end-of-year standardized state test scores. A traditional model including only past performance on the test yielded an R^2 of 0.38 and an enhanced traditional model that added current class average improved predictions (R^2=0.50). These models served as baseline measures for comparing an ITS model. Logistic regression models that include features such as hint percentage, average number of attempts and percent correct overall improved the R^2 to 0.57. The predictive power of the data is as effective with only a few months of use. This lends support for the increased use of the systems in the classroom and for nightly homework.

Keywords

Intelligent tutoring system, homework, standardized test, prediction, regression, classification, decision tree

1. INTRODUCTION

With the introduction of the No Child Left Behind Act in 2001, assessing student performance became a significant focus of schools. With the high-stake nature of these tests, it is imperative to identify at-risk students accurately and as early in the year as possible to provide time for interventions. Intelligent tutoring systems (ITS) allow teachers to evaluate student performance while students are learning. Furthermore, the ITS provides data to teachers which can be used to predict standardized-state-test scores (Feng, et al. 2006, Feng, et al. 2008). Specifically, help request behavior is effective at predicting student proficiency (Beck et al. 2003).

While the above studies are promising, the content used to generate the data was very narrow, consisting of previously released state test questions. Therefore the material mapped directly to the test that was being predicted. The present study uses ASSISTments (www.assistments.org), a web-based intelligent tutoring system, which allows teachers to enter their own content in addition to using certified problem sets. This content can include in-class warm ups, challenge problems, and questions from the textbook. Some of the problem sets may include tutoring in the form of hints or scaffolding while others include correctness only feedback with varying numbers of attempts allowed. What impact does this diverse data have on the previously established usefulness of ITS in predicting end-of-year test scores? The present research attempts to determine if the data collected from student use of an ITS over an entire school year accurately predicts student performance on a standardized-state-test.

2. APPROACH

For the 2010-1011 school year, 129 students in a suburban middle school used ASSISTments as part of their 7th grade math class. The different types of assignments completed during the course of the year include classwork, homework and assessments. Student data from August through May was used to predict MCAS (Massachusetts Comprehensive Assessment System) scaled score and to classify performance. A smaller date range (August through October) was also considered to determine if the model is equally effective with less data, earlier in the school year.

2.1 Modeling

The **traditional model** most schools use to predict 7th grade MCAS scaled scores is a student's 6th grade MCAS score. An **enhanced traditional model** added student's average. These models serve as a baseline to compare the **ITS models** for the different date ranges. Based on the previous literature that successfully predicted state test scores from an ITS, many variables were constructed to be included in the model (number of questions answered, percent correct on first attempt, percentage of hints used, and average number of attempts per question).

For predicting a student's 7th grade MCAS scaled score, linear regression was used for both the traditional and enhanced traditional models. Whereas step-wise linear regression models were generated for both date ranges. The models were compared using R^2 and accuracy. To measure the accuracy of each model, the predicted score was used to classify each student (advanced, proficient, needs improvement, warning) and this classification was compared to the actual classification on the 7th grade MCAS.

For classifying purposes, decision trees were generated to predict specific performance level for each time frame as well. Cross validation was used to assess the accuracy of the models.

3. RESULTS

Students who were not enrolled in the course for the entire time period considered in this study were not included in the analysis (n=8). Finally, students whose 6th or 7th grade MCAS scores were not available were not included (n=4).

3.1 Prediction

Traditionally, prior performance on a standardized test is used to predict future performance on the same test. A linear regression using only 6th grade MCAS score to predict 7th grade MCAS scaled score serves as a comparison model for the more complex models. These scores are highly correlated ($r(115)=0.617$, $p<0.001$) and was 75% accurate in categorizing students.

The **enhanced traditional model** included 6th grade MCAS score ($\beta = 0.341$, $t(115) = 4.03$, $p < .001$) as well as Percent_Correct_First_Attempt ($\beta = 0.448$, $t(115) = 5.30$, $p <$

.001). This model was more successful (F(2,114)=57.10, p<0.001) and accounted for 51% of the variance. This model also yielded 75% accuracy in categorizing students.

A step-wise linear regression was used to generate a model that incorporates data collected from student use of ASSISTments. The regression model using the data from August through May yielded a model that included percent correct overall ($\beta = 0.479$, $t(115) = 6.80$, $p < .001$) and class ($\beta = 0.399$, $t(115) = 5.67$, $p < .001$). This model was 82% accurate. A step-wise regression generated an identical model using only the data from August through October and resulted in 81% accuracy. This confirms that the data collected during the first quarter is equally sufficient when predicting end-of-year test results. Furthermore, the inclusion of a measure only available through the use of the intelligent tutoring system supports its use in the classroom. The results of these models can be found in Table 1.

3.2 Classification

A J48 Decision tree with cross validation predicted MCAS classification with 68.4% accuracy for the full year. The attributes included in the tree were prior MCAS performance, total number of questions answered, and percentage of hints used. While the tree does well with predicting the classification of Advanced and Proficient, it was unable to identify the students who fell in the Needs Improvement category. This is a significant limitation of this model. However, it is important to note that with only 2 students falling in the Needs Improvement category, it will be very challenging to identify them.

A separate decision tree was constructed based on the data from August through October. This model predicted MCAS classification better with 76% accuracy. See Figure 1 for the tree. The attributes that were included were average number of attempts and percentage of hints used.

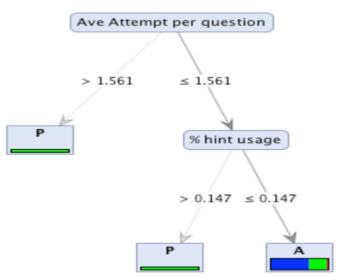

Figure 1. J48 Decision Tree for predicting MCAS classification using data from August through October.

4. Contribution and Discussion

Being able to predict how students will perform on end-of-year standardized tests allows teachers to identify at risk students and offer interventions. Traditionally, teachers had only the previous year's performance. While this is highly correlated with current performance, the current study shows that ITS provide additional data that allow teachers to better predict performance, and earlier in the year. Using this data to offer remediation and interventions should be considered by educators who use ITS regularly.

This study is unique in that the ITS was used throughout the year for nightly homework, often with correctness only feedback and diverse content that is not closely mapped to the final measurement of performance. The ability to still accurately predict performance provides evidence of yet another use of ITS within schools.

Table 1. Predictive and classification model performance.

Model	R^2	Accuracy	Kappa
Traditional	0.381	75%	0.44
Enhanced Traditional	0.505	75%	0.46
ITS (First Quarter)	0.566	81%	0.61
ITS (Full Year)	0.566	82%	0.63
Classification (First Quarter)	N/A	76%	0.39
Classification (Full Year)	N/A	68%	0.36

Both prediction and classification required data that could only be generated by the use of ITS, and not through traditional classroom measures. Specifically, percent correct overall was a useful predictor. This can only be generated by allowing students to learn while doing their homework and measuring their success beyond just their first response. Similarly, average number of attempts on each problem and hint percentage add to the model when trying to classify student performance. This lends more support for the use of ITS for nightly homework.

5. ACKNOWLEDGMENTS

Bill and Melinda Foundation via EDUCAUSE, IES grants R305C100024 and R305A120125, and NSF grants ITR 0325428, HCC 0834847, DRL 1235958.

6. REFERENCES

[1] Beck, J., Jia, P., Sison, J., Mostow, J. (2003). Predicting student help-request behavior in an intelligent tutor for reading. Proc. of the 9 th Int. Conf. on User Modeling, LNAI 2702

[2] Feng, M., Heffernan, N., Koedinger, K., (2006). Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required. Proceedings of the 8th International Conference on Intelligent Tutoring Systems. Springer-Verlag: Berlin.

[3] Feng, M., Beck, J., Heffernan, N., Koedinger, K. (2008). Can an intelligent tutoring system predict math proficiency as well as a standardized test. In Baker & Beck (Eds.). Proceedings of the First International Conference on Educational Data Mining.