Predicting Student Success on the Texas Chemistry STAAR Test: A Logistic Regression Analysis

By

William L. Johnson, Ed.D., LMFT

William.Johnson@tylerist.org

Robert E. Lee High School

411 ESE Loop 323

Tyler, TX 75701

(903) 262-2625

&

Annabel M. Johnson, Ph.D.

235 Texas Dr.

Hideaway, TX 75771

&

Jared Johnson, B.S.

235 Texas Dr.

Hideaway, TX 75771

Paper Submitted August 25, 2012

Predicting Student Success on the Texas Chemistry STAAR Test: A Logistic Regression Analysis

By: William L. Johnson, Ed.D., Annabel M. Johnson, Ph.D. & Jared Johnson, B.S.

Parents, students, teachers, superintendents, and university professors assembled in force for the June 19, 2012, Texas House Public Education Committee meeting. Nearly all agreed that there was something seriously wrong with the state's testing system. Parents offered critiques positing the distortion of students' education by the over-emphasis on test preparation in Texas' public schools and decried the current five-year near $500 million test-development contract with the contracted testing company. Others noted the nationwide groundswell of opposition to the excesses and misuse of standardized testing. Several House members on the committee made comments suggesting they were willing to try again to scale back the testing excesses; however, there was no group consensus on what to do.

Meanwhile, school boards in more than 425 school districts in Texas, and increasing numbers in other states, have signed onto a resolution against the current excesses of high-stakes testing. See http://fairtest.org/national-resolution-highstakes-testing for the full text of the national resolution. In June 2012, Florida's State School Board Association adopted a resolution condemning the overuse of high-stakes tests and objected to their use as the primary basis for evaluating teachers, administrators, schools and school districts. Without doubt, there are critics in abundance decrying the excesses and misuse of standardized testing.

No doubt, there are limitations with testing. Testing can be insensitive to students; that is, test questions are systematically skewed to reproduce the score distribution achieved by student groups in the past. A very high percentage of test variance in standardized test scores from year-to-year may largely reflect students' test-taking skills and be insensitive to what is being taught in the classroom. Furthermore, tests sample only limited amounts of the domain being tested, and a distortion may occur when testing a limited amount of a domain.

However, according to the critics, this is not the real issue. The real issue is that we have tested students for years, and the test scores have not gone up much. And second, the amount of money spent to develop these tests has done little to improve the graduation rates from the public schools. The levels of proficiency represented in many of these tests are so low that the testing has really not changed anything significantly.

Many want to get rid of student standards and testing because they don't like what happens with the results; that is, they see testing used as a social and political tool. However, it is argued that tests have served the purpose of identifying and recognizing effective teachers and administrators in schools and highlighting successful educational practices. Test data have also provided educators with the ability to simplify complex issues. Perhaps critics do have a valid point about testing. It really is about using tests in a productive and ethical way in order to improve education.

The Uses of Data and Testing

Both sides of the testing issue are likely over reactive and missing a middle ground. The purpose of this article was to bridge-the-gap somewhat by giving examples of how test data and testing could be used in a productive and ethical way to help students. The authors have addressed testing in a way to help our students. We have found in our teaching that students are not very test sophisticated. Thus, before our science TAKS testing we reviewed the five major components (objectives) tested on the State of Texas science TAKS test. Our data analysis of three years of previous science test scores showed the following: if students answered at least 10 questions correctly on objective one (out of 17 questions on that objective), 95% of our students passed science TAKS. If they did not answer 10 questions correctly, most could still pass the science TAKS test if they answered at least seven questions correctly on objective five (15 physics formula problem types). Objectives two and three (biology) did not fail our students since they had taken biology the previous year. Low scores the year before on objective four (chemistry) was not a problem since the students were currently taking chemistry. We told our students our estimates of the number of questions needed to pass the yearly science TAKS test and gave them strategies for answering questions. We also gave students a summary sheet of their past TAKS scores, noting areas of strength and weakness. We then gave our students remedial work based on their areas of need on their previous TAKS test (Johnson & Johnson, 2010).

Understanding Test Construction

Furthermore, our analysis of the structure of the science TAKS test showed that 40% of the test was process skills, and this percentage has held from year-to-year. Process skills would include the following: being able to understand and use formulas; draw inferences; ability to communicate conclusions and evidence; collect and organize data; analyze, evaluate and critique data; plan and implement experiments; evaluate changes based on data; and plan, implement and ask questions based in scientific settings. Generally these skills and abilities encompass observation, communication, classification, measurement, inference and prediction (Johnson & Johnson, 2012).

This information is invaluable in helping the students prepare for the process-skill portions of the science TAKS test. It has also been invaluable in helping students "know science." In lab work, the students have used process skills to work in a sophisticated way. Using our Vernier LabQuest units, the students have collected data from experiments and modeled the data as linear and nonlinear. However, we have primarily used the general linear model (GLM) in our data analysis. Our students have applied the evidence/findings from our science experiments to really begin to know and understand science. Students have not just memorized facts. These skills and the knowledge of test structure and previous test scores have resulted in exemplary science TAKS scores for our students for the past several years. This illustrates using strategies and test data in a productive manner to foster student success and increased graduation rates. Most of our special education students have passed science TAKS. Many of them had never passed a science TAKS test before.

The State of Texas began moving incrementally from TAKS testing to the new State of Texas Assessment of Academic Readiness test in 2011-2012. Only juniors and senior re-testers will take the TAKS test in the 2012-2013 academic school year. This will be the last administration of the TAKS test except for retesting students who failed the test; thus, one sees the need to be preparing for the STAAR test. Furthermore, this knowledge would be most valuable considering that only 28% of Texas school districts and 44% of Texas campuses met the “adequate yearly progress” (AYP) requirement during the 2011-2012 school year under the No Child Left Behind Act. See http://www.tea.state.tx.us/ayp/ for additional AYP information.

Under the federal school accountability system, a school or district in the 2011-2012 school year met AYP requirement if 87% or more of their students passed the state reading/English language arts test; 83% passed the state mathematics test; 95% participated in the state testing program; and, depending on the grade level, had either a 75% graduation rate or a 90% attendance rate. Under the current structure of the NCLB Act, the passing standards will rise to 100% in the math and reading tests by 2014. This means steep increases in the passing requirements through 2014.

Furthermore, the Texas Education document titled: ”2012 AYP Requirements Rise,” provides additional information pertaining to sanctions for missing AYP. See the following news release: http://www.tea.state.tx.us/news_release.aspx?id=2147508185, dated August 8, 2012. The news release notes that Non-Title 1 schools that miss AYP must revise their already existing campus improvement plans to address the reasons that the campus missed AYP. Schools or districts in Stages 2-5 face stronger sanctions at each additional stage. A school that has reached Stage 2 sanctions, for example, must offer tutoring to its students.

The TEA news release notes that at Stage 5 (the most advanced intervention level), a school must adopt an alternative form of governance. Along with tutoring options and offering school transfer, a school at Stage 5 could do the following: reopen as a charter school; replace all or most of the school staff; contract with a private management company to operate the school;

turn the school's operation over to TEA; or adopt any other major restructuring of school governance. At Stage 5 for two or more years, TEA staff with meet with the campus and district staff to discuss way to revise the restructuring plan to make it more successful. See the following: http://www.tea.state.tx.us/index4.aspx?id=44598&menu_id=798.

Furthermore, the 2009 Texas legislature directed that educator-preparation programs be held accountable for the impact of their graduates on student achievement. As the metrics of the education-preparation effectiveness program are replicated for two-to-three years to determine reliability and validity, the suggestions given in this document will be invaluable to beginning teachers. Also, the following discussion of our logistic regression pilot study will be of great value to new teachers.

Logistic Regression Analysis

In addition to what has been noted (Johnson & Johnson, 2010), the authors piloted an SPSS logistic regression program to calculate the actual probability that each student would pass or fail the STAAR test. Since logistic regression does not make any assumptions about normality, linearity and homogeneity of variance for the independent variables in the analysis, logistic regression is being used more-and-more frequently because it can be interpreted similarly to other general linear model (GLM) solutions. As the statistic of choice, logistic regression has catapulted data analysis to a much higher level by assigning a probability that each student will pass or fail the STARR end-of-course (EOC) test. This has added significantly to the knowledge a teacher would have about each student in his or her classes. Knowing the probability of each student passing the EOC test would alert the teacher to his or her students needing special preparation and help for the EOC test.

The authors had a convenience sample of n = 32 sophomores and n = 68 juniors at Robert E. Lee High School (5A) Tyler, Texas. The data were analyzed in the spring of 2012. In our initial logistic regression analysis, the authors entered the students' state pilot chemistry STAAR test (coded pass = 1 or fail = 0), their last science TAKS raw score, their last science TAKS Z-score, their last science TAKS test score (coded pass = 1 or fail = 0) and their grade level (coded sophomore = 0 or junior = 1).

In our analysis, using the last science TAKS Z- score instead of the raw science TAKS score, there was a numerical problem with a standard error in the modeled event. That is, one should not interpret the numerical logistic regression solution if any of the independent variables had standard errors (SEs) greater than 2.0. This was the problem in our first analysis. Note, however, that the SE rule does not apply to the constant in the final solution. When we reentered the dependent variable and only the last science TAKS raw score and the grade level, the logistic regression solution converged with no numerical problems. Since the SPSS program does not compute tolerance values, the authors affirmed the solution by examining the standard errors in the final logistic solution. Both independent variables had SEs less than 2.0

The authors also examined the data for outliers. When the outliers were removed and the analysis rerun without the outliers, the solution differed by only 0.32% from the original solution. Since the solution did not differ by at least 2% from the original solution, we kept the original solution. This is the standard approach for examining outliers and selecting which solution to use.

We also analyzed the statistical output for the usefulness of the model based on classification accuracy. To be classified as a useful model, the accuracy rate should be at least 25% higher than chance accuracy. Our analysis showed that the criteria for classification accuracy was satisfied. The accuracy rate was 30% higher than by chance. We did this calculation “by hand” since SPSS does not compute a cross-validated accuracy rate for logistic regression.

Following is the SPSS version 20.0 printout of the binary logistic regression analysis, as well as the student data. The authors will also briefly explain the statistical meaning of each step in the SPSS analysis. The statistical output will give the actual probability that the student will pass or fail the STAAR test and the accuracy of the prediction model. Our analysis showed a very high classification accuracy. The model showed that 52 out of 56 students (92.9%) were correctly predicted as passing the STAAR EOC test, and 41 out of 44 students (93.2%) were correctly predicted as failing the EOC test. The overall accuracy of the model was 93%. Following is the SPSS 20.0 binary logistic regression analysis printout with explanatory comments noted for each table or chart. See Brace, Kemp, and Snelgar (2003) and Meyers, Gamst, and Guarino (2006) for an explanation of SPSS logistic regression analysis.

References

Brace, N., Kemp, R. & Snelgar, R. (2003). SPSS for psychologists (2nd Ed.). Mahwah, NJ: Erlbaum.

Johnson, W.L., & Johnson, A.B. (2010). An innovative program that results in science school improvement. Texas Study of Secondary Education, 19(2), 21-23.

Johnson, W.L., & Johnson A.B. (2012). Planning for the STAAR test: The third generation correlates of effective schools. Texas Study of Secondary Education, 21(2), 5-9.

Meyers, L.S., Gamst, G. & Guarino, A.J. (2006). Applied Multivariate Research Design and Interpretation. Thousand Oaks, CA: Sage Publications.

National Center for Fair and Open Testing (2012). Retrieved July 16, 2012, from http://fairtest.org/national-resolution-highstakes-testing.

Texas Education Agency (2012, August 8). 2012 AYP requirements rise. Austin, TX. Retrieved August 16, 2012, from http://www.tea.state.tx.us/ayp/.

Texas Education Agency (2012, August 8). 2012 AYP requirements rise. Austin, TX. Retrieved August 17, 2012, from http://www.tea.state.tx.us/index4.aspx?id=4459&menu_id=798.

Texas Education Agency (2012, August 8). 2012 AYP requirements rise. Austin, TX. Retrieved August 18, 2012, from http://www.tea.state.tx.us/news_release.aspx?id=2147508195.

```
LOGISTIC REGRESSION VARIABLES pofeoc
  /METHOD=ENTER taksraw
  /CLASSPLOT
  /PRINT=GOODFIT ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

LOGISTIC REGRESSION

[DataSet1] C:\Users\Johnson\Documents\Logregression.sav]

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	100	100.0
	Missing Cases	0	.0
	Total	100	100.0
Unselected Cases		0	.0
Total		100	100.0

a. If weight is in effect, see classification table for the total number of cases.

(This table reports that 100% of the cases have been processed.)

Dependent Variable Encoding

Original Value	Internal Value
.00	0
1.00	1

(This table tells you how the two outcomes have been coded.)

Categorical Variables Codings

		Frequency	Parameter coding
			(1)
sophjr	.00	32	1.000
	1.00	68	.000

(This table gives the categorical variables coding.)

BLOCK 0: BEGINNING BLOCK

(This section reports the results of the most basic attempt to predict the outcome; that is, this section describes a "null model": a model with no predictors and just the constant which is analogous to the y-intercept in OLS regression. This is why one will see all the variables put into the model in the table titled, "Variables not in the Equation.")

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	137.186	.240
	2	137.186	.241
	3	137.186	.241

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 137.186

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

(The first Iteration History table shows the estimation was terminated after three iterations. The -2 Log likelihood (-2LL) is a likelihood ratio and represents the unexplained variance in the outcome variable. The smaller the value, the better the fit. If a model fits perfectly, the likelihood = 1 and -2LL = 0. Recall that likelihood is the probability of the observed results given the parameter estimates.).

Classification Table[a,b]

	Observed		Predicted		
			pofeoc		Percentage Correct
			.00	1.00	
Step 0	pofeoc	.00	0	44	.0
		1.00	0	56	100.0
	Overall Percentage				56.0

a. Constant is included in the model.

b. The cut value is .500

(The Classification Table reports the results of this simple prediction. The table shows how well the null model correctly classifies cases. The key information is the percentage in the lower right corner which shows the null model is only 56% accurate. This is only slightly greater than the accuracy of random guessing. The full model should be much more accurate.)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.241	.201	1.433	1	.231	1.273

(The Variables in the Equation table show the logistic coefficient (B) associated with the intercept included in the model. The table is similar to a standard logistic regression table. SE is the standard error associated with the coefficient. The Wald statistic is a chi-square 'type' of statistic and is used to test the significance of the variable in the model. The hypothesis is accepted, and one concludes the constant is zero. The df equals one for the Wald test since there is only one predictor in the model, namely the constant. Exp(B) refers to the change in the odds ratio attributed to the variable; that is, 56/44 = 1.273.)

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	taksraw	60.780	1	.000
		sophjr(1)	.688	1	.407
	Overall Statistics		60.932	2	.000

The Variables not in the Equation table list the Wald test score, df, and p-value for each variable not included in the beginning block model. This is a Score Test used to predict whether or not an independent variable would be significant in the model. This table shows that the TAKS raw predictor would be significant, and the sophjr(1) predictor would not be significant. The df is degrees of freedom for each variable. The overall statistics coefficient shows the result of including all the predictors in the model. Notice that this is not a total but an estimate of the overall Wald statistic associated with the model had all the variables been included.)

BLOCK 1: Method = Enter

(This block reports the results of the logistic regression analysis. It is more accurate than Block 0. It is the most interesting part of the output: the overall test of the model in the "Omnibus Tests of Model Coefficients" table and the coefficients and odds ratios in the "Variables in the Equation" table.)

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients		
			Constant	taksraw	sophjr(1)
Step 1	1	65.928	-7.539	.195	.168
	2	45.755	-14.247	.361	.287
	3	37.506	-21.506	.541	.382
	4	34.995	-28.022	.702	.508
	5	34.629	-31.676	.792	.607
	6	34.618	-32.460	.811	.633
	7	34.618	-32.489	.812	.634
	8	34.618	-32.489	.812	.634

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 137.186

d. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

(The iteration history shows that the estimation was terminated at iteration #8 because the parameter estimates did not change by more than 0.001. The -2LL is a likelihood ratio and represents the unexplained variance. Thus, the smaller the value the better. Notice here the -2LL value (34.618) is much lower than the value in the null model (137.186). Though the iteration history is not usually of interest, the table does show how the model chi-square value was derived. Thus, 137.186 – 34.618 = 102.568. This is the value for the model chi-square in the Omnibus Tests of Model Coefficients table that follows next. The -2LL is a likelihood ratio and represents the unexplained variance in the outcome variable.)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	102.568	2	.000
	Block	102.568	2	.000
	Model	102.568	2	.000

(Omnibus tests are general tests of how well the model performs. The table reports the chi-square associated with each step in the Enter model. We can use either this Omnibus test or the Hosmer-Lemeshow test that will follow later. There is only one step from the constant model to the block containing predictors; thus, all three values are the same. The null hypothesis that there was no difference between the model with only a constant and our model (Block 1, with predictors) with independent variables was rejected. The existence of a relationship between the independent variables and the dependent variable was supported; that is, our model was statistically significant from the constant only model because of the significance (p-value) noted in the Omnibus table. There is a significant effect for the combined predictors on the outcome variable.)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	34.618[a]	.641	.859

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

(This is the first absolute measure of the validity of the model and evaluates whether or not the set of independent variables improves prediction of the dependent variable better than chance. It tests if at least one of the independent variables (covariates) is statistically different from zero. The -2 log likelihood (-2LL) is used because it is a chi-squared distribution, while –LL is not. Therefore, the -2LL measure can be used for assessing the significance of the logistic regression model. This table also provides “pseudo” equivalent R-squares somewhat similar to the R-squared value that is found in OLS regression; that is, the proportion of variance explained by the predictors. It is not possible to compute an exact R-squared value in log regression; thus, one should interpret these R-squared values with some caution. However, most researchers prefer the Nagelkerke pseudo R-squared because it can achieve a value between 0 and 1; thus, it can be evaluated as indicating a model fit. The Cox and Snell pseudo R-squared value can have a value greater than one, and the larger the value the better the estimate.)

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.318	7	.621

(This table gives the results of the Hosmer and Lemeshow Test. This is the preferred test of goodness-of-fit. This table also provides a formal test of the agreement between the observed and predicted outcomes. A large p-value indicates a good match. If the p-value is less than 0.05, the model does not adequately fit the data, and one needs to look for alternate variables. In this table, the p-value is large (0.621) and greater than our established cutoff (generally 0.05) indicating a good fit. In other words, a non-significant chi-square means that the predicted probabilities match the observed probabilities, and our model predicts values not significantly different from what we observed.)

Contingency Table for Hosmer and Lemeshow Test

		pofeoc = .00		pofeoc = 1.00		Total
		Observed	Expected	Observed	Expected	
Step 1	1	9	9.000	0	.000	9
	2	10	9.990	0	.010	10
	3	10	10.817	1	.183	11
	4	10	9.597	2	2.403	12
	5	5	3.719	8	9.281	13
	6	0	.662	11	10.338	11
	7	0	.166	10	9.834	10
	8	0	.042	10	9.958	10
	9	0	.007	14	13.993	14

(This Contingency Table was used in the calculation of the previous table and shows more detail. The table showed the observed and expected values for each category of the outcome variable. The cases are ranked by the estimated probability on the criterion variable; that is, the test divides the data into approximately 10 groups. These groups are defined by increasing order of estimated risk. The first group has nine students, the second group 10 students, the third group 11 students, etc. Notice the close match between the observed and expected values for each group.)

Classification Table[a]

	Observed		Predicted		
			pofeoc		Percentage Correct
			.00	1.00	
Step 1	pofeoc	.00	41	3	93.2
		1.00	4	52	92.9
	Overall Percentage				93.0

a. The cut value is .500

(The Classification Table summarizes the results of the full logistic regression prediction model. It shows how well our full model correctly classified cases. The model correctly predicted 93.0% of the students would pass or fail the end of course STAAR test. This is the overall rate for the model. One can see that this percentage has increased from 56.0 in the earlier Classification Table to 93.0 in the Full Prediction Model.)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	taksraw	.812	.192	17.819	1	.000	2.252
	sophjr(1)	.634	1.024	.383	1	.536	1.885
	Constant	-32.489	7.807	17.318	1	.000	.000

a. Variable(s) entered on step 1: taksraw, sophjr.

(The first column in the table gives the logistic coefficients(B) of each predictor variable. The Wald statistic provides Wald chi-square values used in testing the null hypothesis. The statistic indicates how useful each predictor was. It tells us whether or not the logistic coefficient (B) is different than zero. Recall that the Wald statistic value is (B/SE), but in most in most software it is calculated as $(B/SE)^2$. The degrees of freedom for each of the tests of the coefficients are listed in the df column. The Exp(B)is the odds ratio associated with each predictor. The column gives the indication of the change in the predicted odds of “pass or fail” the End of Course (STAAR) test for each unit change in the predictor variable. We expect predictors which increase the logit to display Exp (B) values greater than 1.0, and Exp (B) values less than 1.0 to decrease the logit. The logit is what is being predicted. It is the odds of membership in the category of the outcome variable with the numerically greater value (here a one value rather than a zero value).

Recall, if p = probability; then odds = p/1-p; and log odds or logit (p)= log(p/1-p). Also, logit (p) = a + bx. For two independent variables, logit (p) = a + B_1X_1 + B_2X_2 where a is the constant, the Xs are the independent variables, the Bs are the logistic coefficients, and logit (p) is the odds of membership in the outcome variable. Then the following equation is correct: p = $e^{logit(p)}$ / 1 + $e^{logit(p)}$. We then have our predicted probabilities for pass or fail the STAAR test, and the group membership follows. Note also in the Variables in the Equation that e^B or($e^{.812}$) = 2.252 and that ln(2.252) = .812. Entering the raw TAKS score (taksraw) in the model, note that both the constant and raw TAKS score are significant.

One can now use the results of the binary logistic regression analysis to calculate the STAAR passing odds for each new student who will take the STAAR test. To do this, add the data available for each new student. Repeat the analysis as before, but in the logistic regression dialogue box, click on the “save” button. This will bring up the “logistic regression save variable” box. Then select the “Probabilities” and “Group membership” boxes. Next, Click the “ok” button. Two new variables will be added to the data file: the “probability of passing” the STAAR test and the “predicted group” noted as pass or not. One can use these values to make real predictions.)

Step number: 1

Observed Groups and Predicted Probabilities

```
      32 +
+
         I
I
         I
I
F        I
1I
R     24 +
1+
E        I0
1I
Q        I0
1I
U        I0
1I
E     16 +0
1+
N        I0
1I
C        I0
1I
Y        I0
1I
      8 +0
1+
         I00                            1
1              11I
         I00     0                      0
1              1       1    1111I
         I00 0  00      0 0         0   0              0   0
0           1  0      11   11111I
Predicted ---------+---------+---------+---------+---------+---------+-------
--+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6
.7        .8        .9         1
  Group:
000000000000000000000000000000000000000000000000001111111111111111111111111111
11111111111111111111111
```

Predicted Probability is of Membership for 1.00
The Cut Value is .50
Symbols: 0 - .00
1 - 1.00
Each Symbol Represents 2 Cases.

(The graph above shows how the full model predicts membership. The better the model, the less zeros and ones are in the middle of the graph. When the model is less accurate, there are more symbols (zeros or ones) in the middle of the graph, displaying their probability (x-axis).)

Student	Taksraw	Sophjr	Porfeoc	Porftaks	ztaksraw
1	36.00	.00	.00	1.00	-.45129
2	24.00	.00	.00	.00	-1.94315
3	43.00	.00	1.00	1.00	.41896
4	44.00	.00	1.00	1.00	.54329
5	39.00	.00	.00	1.00	-.07832
6	24.00	.00	.00	.00	-1.94315
7	49.00	1.00	1.00	1.00	1.16490
8	44.00	.00	1.00	1.00	.54329
9	42.00	.00	1.00	1.00	.29464
10	42.00	1.00	1.00	1.00	.29464
11	45.00	.00	1.00	1.00	.66761
12	44.00	.00	1.00	1.00	.54329
13	43.00	1.00	1.00	1.00	.41896
14	38.00	.00	.00	1.00	-.20264
15	43.00	.00	1.00	1.00	.41896
16	12.00	.00	.00	.00	-3.43501
17	38.00	1.00	.00	1.00	-.20264
18	44.00	1.00	1.00	1.00	.54329
19	48.00	.00	1.00	1.00	1.04057
20	48.00	.00	1.00	1.00	1.04057

Data for the Logistic Regression Analysis

21	50.00	.00	1.00	1.00	1.28922
22	41.00	1.00	.00	1.00	.17032
23	54.00	.00	1.00	1.00	1.78651
24	39.00	1.00	.00	1.00	-.07832
25	37.00	1.00	.00	1.00	-.32697
26	39.00	1.00	.00	1.00	-.07832
27	17.00	1.00	.00	.00	-2.81340
28	35.00	1.00	.00	1.00	-.57561
29	34.00	1.00	.00	1.00	-.69993
30	27.00	1.00	.00	.00	-1.57019
31	41.00	.00	1.00	1.00	.17032
32	30.00	.00	.00	.00	-1.19722
33	49.00	1.00	1.00	1.00	1.16490
34	34.00	1.00	.00	1.00	-.69993
35	44.00	1.00	1.00	1.00	.54329
36	41.00	1.00	.00	1.00	.17032
37	29.00	1.00	.00	1.00	-1.32154
38	37.00	1.00	.00	1.00	-.32697
39	28.00	.00	.00	.00	-1.44586
40	35.00	1.00	.00	1.00	-.57561
41	42.00	1.00	.00	1.00	.29464
42	45.00	1.00	1.00	1.00	.66761
43	48.00	1.00	1.00	1.00	1.04057
44	29.00	1.00	.00	1.00	-1.32154
45	43.00	1.00	1.00	1.00	.41896
46	29.00	1.00	.00	1.00	-1.32154
47	45.00	1.00	1.00	1.00	.66761
48	39.00	1.00	.00	1.00	-.07832
49	20.00	1.00	.00	.00	-2.44044
50	51.00	1.00	1.00	1.00	1.41354
51	37.00	1.00	.00	1.00	-.32697
52	47.00	.00	1.00	1.00	.91625
53	49.00	.00	1.00	1.00	1.16490
54	42.00	1.00	1.00	1.00	.29464
55	49.00	1.00	1.00	1.00	1.16490
56	47.00	1.00	1.00	1.00	.91625
57	40.00	1.00	.00	1.00	.04600
58	49.00	1.00	1.00	1.00	1.16490
59	32.00	1.00	.00	1.00	-.94858
60	31.00	1.00	.00	1.00	-1.07290
61	34.00	1.00	.00	1.00	-.69993
62	35.00	.00	.00	1.00	-.57561
63	33.00	1.00	.00	1.00	-.82425
64	39.00	.00	1.00	1.00	-.07832
65	43.00	1.00	1.00	1.00	.41896
66	46.00	1.00	1.00	1.00	.79193

67	47.00	1.00	1.00	1.00	.91625
68	47.00	1.00	1.00	1.00	.91625
69	50.00	1.00	1.00	1.00	1.28922
70	31.00	1.00	.00	1.00	-1.07290
71	49.00	1.00	1.00	1.00	1.16490
72	35.00	1.00	.00	1.00	-.57561
73	45.00	.00	1.00	1.00	.66761
74	39.00	1.00	1.00	1.00	-.07832
75	49.00	1.00	1.00	1.00	1.16490
76	35.00	1.00	1.00	1.00	-.57561
77	35.00	1.00	.00	1.00	-.57561
78	43.00	1.00	1.00	1.00	.41896
79	24.00	.00	.00	.00	-1.94315
80	49.00	1.00	1.00	1.00	1.16490
81	42.00	1.00	1.00	1.00	.29464
82	39.00	1.00	1.00	1.00	-.07832
83	27.00	.00	.00	.00	-1.57019
84	29.00	.00	.00	.00	-1.32154
85	45.00	1.00	1.00	1.00	.66761
86	42.00	1.00	1.00	1.00	.29464
87	45.00	1.00	1.00	1.00	.66761
88	37.00	.00	.00	1.00	-.32697
89	42.00	1.00	1.00	1.00	.29464
90	32.00	.00	.00	.00	-.94858
91	46.00	1.00	1.00	1.00	.79193
92	47.00	1.00	1.00	1.00	.91625
93	35.00	.00	.00	1.00	-.57561
94	44.00	1.00	1.00	1.00	.54329
95	46.00	1.00	1.00	1.00	.79193
96	48.00	1.00	1.00	1.00	1.04057
97	34.00	.00	.00	1.00	-.69993
98	44.00	1.00	1.00	1.00	.54329
99	45.00	1.00	1.00	1.00	.66761
100	41.00	1.00	1.00	1.00	.17032

(Column 1 Taksraw is the science TAKS raw score; column 2 Sophjr is the student class level of sophomore = 0 or junior = 1; column 3 Porfeoc is pass = 1 or fail = 0 for the STAAR (EOC) test; column 4 Porftaks is pass = 1 or fail = 0 for the science TAKS test; and column 5 ztaksraw is the Z score for the science TAKS raw score listed in column 1).

Predicted Probabilities for Pass/Fail STAAR (PRE_1) and Group (PGR_1)

taksraw	sophjr	pofeoc	porftaks	Ztaksraw	PRE_1	PGR_1
36	0	0	1	-0.45129	0.0675	0
24	0	0	0	-1.94315	0	0
43	0	1	1	0.41896	0.95511	1
44	0	1	1	0.54329	0.97956	1
39	0	0	1	-0.07832	0.45264	0
24	0	0	0	-1.94315	0	0
49	1	1	1	1.1649	0.99932	1
44	0	1	1	0.54329	0.97956	1
42	0	1	1	0.29464	0.90428	1
42	1	1	1	0.29464	0.83364	1
45	0	1	1	0.66761	0.99082	1
44	0	1	1	0.54329	0.97956	1
43	1	1	1	0.41896	0.91861	1
38	0	0	1	-0.20264	0.26857	0
43	0	1	1	0.41896	0.95511	1
12	0	0	0	-3.43501	0	0
38	1	0	1	-0.20264	0.16301	0
44	1	1	1	0.54329	0.96215	1
48	0	1	1	1.04057	0.99919	1
48	0	1	1	1.04057	0.99919	1
50	0	1	1	1.28922	0.99984	1
41	1	0	1	0.17032	0.68993	1
54	0	1	1	1.78651	0.99999	1
39	1	0	1	-0.07832	0.3049	0
37	1	0	1	-0.32697	0.07959	0
39	1	0	1	-0.07832	0.3049	0
17	1	0	0	-2.8134	0	0
35	1	0	1	-0.57561	0.01676	0
34	1	0	1	-0.69993	0.00751	0
27	1	0	0	-1.57019	0.00003	0
41	0	1	1	0.17032	0.8075	1
30	0	0	0	-1.19722	0.00055	0
49	1	1	1	1.1649	0.99932	1

34	1	0	1	-0.69993	0.00751	0
44	1	1	1	0.54329	0.96215	1
41	1	0	1	0.17032	0.68993	1
29	1	0	1	-1.32154	0.00013	0
37	1	0	1	-0.32697	0.07959	0
28	0	0	0	-1.44586	0.00011	0
35	1	0	1	-0.57561	0.01676	0
42	1	0	1	0.29464	0.83364	1
45	1	1	1	0.66761	0.98283	1
48	1	1	1	1.04057	0.99847	1
29	1	0	1	-1.32154	0.00013	0
43	1	1	1	0.41896	0.91861	1
29	1	0	1	-1.32154	0.00013	0
45	1	1	1	0.66761	0.98283	1
39	1	0	1	-0.07832	0.3049	0
20	1	0	0	-2.44044	0	0
51	1	1	1	1.41354	0.99987	1
37	1	0	1	-0.32697	0.07959	0
47	0	1	1	0.91625	0.99818	1
49	0	1	1	1.1649	0.99964	1
42	1	1	1	0.29464	0.83364	1
49	1	1	1	1.1649	0.99932	1
47	1	1	1	0.91625	0.99657	1
40	1	0	1	0.046	0.49696	0
49	1	1	1	1.1649	0.99932	1
32	1	0	1	-0.94858	0.00149	0
31	1	0	1	-1.0729	0.00066	0
34	1	0	1	-0.69993	0.00751	0
35	0	0	1	-0.57561	0.03114	0
33	1	0	1	-0.82425	0.00335	0
39	0	1	1	-0.07832	0.45264	0
43	1	1	1	0.41896	0.91861	1
46	1	1	1	0.79193	0.9923	1
47	1	1	1	0.91625	0.99657	1
47	1	1	1	0.91625	0.99657	1
50	1	1	1	1.28922	0.9997	1
31	1	0	1	-1.0729	0.00066	0
49	1	1	1	1.1649	0.99932	1
35	1	0	1	-0.57561	0.01676	0
45	0	1	1	0.66761	0.99082	1
39	1	1	1	-0.07832	0.3049	0

49	1	1	1	1.1649	0.99932	1
35	1	1	1	-0.57561	0.01676	0
35	1	0	1	-0.57561	0.01676	0
43	1	1	1	0.41896	0.91861	1
24	0	0	0	-1.94315	0	0
49	1	1	1	1.1649	0.99932	1
42	1	1	1	0.29464	0.83364	1
39	1	1	1	-0.07832	0.3049	0
27	0	0	0	-1.57019	0.00005	0
29	0	0	0	-1.32154	0.00025	0
45	1	1	1	0.66761	0.98283	1
42	1	1	1	0.29464	0.83364	1
45	1	1	1	0.66761	0.98283	1
37	0	0	1	-0.32697	0.14018	0
42	1	1	1	0.29464	0.83364	1
32	0	0	0	-0.94858	0.00281	0
46	1	1	1	0.79193	0.9923	1
47	1	1	1	0.91625	0.99657	1
35	0	0	1	-0.57561	0.03114	0
44	1	1	1	0.54329	0.96215	1
46	1	1	1	0.79193	0.9923	1
48	1	1	1	1.04057	0.99847	1
34	0	0	1	-0.69993	0.01407	0
44	1	1	1	0.54329	0.96215	1
45	1	1	1	0.66761	0.98283	1
41	1	1	1	0.17032	0.68993	1

END OF DOCUMENT