

Predicting English Language Learner Success in High School English Literature Courses

May 2008

The purpose of this paper is to assist ELL educators in making good decisions about English Literature and Composition course enrollment for their high school ELL students. The study uses extant data to explore the relationships between ACCESS proficiency levels and English Language Arts EOCT scale scores. ELL students' ACCESS Proficiency Levels and EOCT scale scores were matched. Results showed moderate to moderately strong positive correlations between ACCESS proficiency levels and American Literature and 9th Grade Literature EOCT scale scores. Regression equations for using ACCESS proficiency levels to predict pass/fail outcomes on the ELA EOCT are discussed and information about what level of English language proficiency a student should demonstrate in order to be 'potentially' successful in an English literature course is given.

Predicting English Language Learner Success in High School English Literature Courses

BACKGROUND

For students who are learning the English language, certain courses particularly those that require large amounts of reading English text can be difficult. Educators must use whatever knowledge they have about students' abilities to place them into courses that are a good fit with their skill level. Courses that are too easy provide no learning challenges for the student and may not count toward graduation course credit. Courses that are too rigorous can frustrate the student because she or he cannot learn the material and as a result, he or she may not pass the course. As with all student groups, factors such as motivation and determination play a role in English Language Learner (ELL) success in the classroom. However, for this population, English language proficiency plays a major role as well.

With the introduction of the Assessing Comprehension and Communication in English State to State for English Language Learners (ACCESS for ELLs) assessment, Georgia educators can track ELL students' progress in English language proficiency. Results from the ACCESS assessment can provide meaningful information about what a student is able to do in the classroom. Theoretically, these results should also provide meaningful information about what a student *should* be able to do in the next classroom, because they describe the student's proficiency level in English.

The purpose of this paper is to help inform decisions about course enrollment of ELL students in high school level English Language Arts (ELA) courses. The specific research question addressed is: What level of English language proficiency must an ELL student have in order to successfully learn the content standards that are taught in English Literature and Composition courses? For this study, level of English proficiency is defined as a student's score on the ACCESS for ELLs assessment. Level of success in an English Literature and Composition course is measured by the scale scores earned on the End of Course Test (EOCT) for the ELA course. The higher the scale score on the EOCT, the higher the achievement of the English Literature curriculum standards the student demonstrates.

ACCESS for ELLs®

The ACCESS for ELLs® is Georgia's state-mandated language assessment. It is a standards-based English Proficiency test crafted to be representative of the social and academic language demands within a school setting. The purpose of the assessment is to monitor student progress in

English language proficiency and to serve as a criterion when English Language Learners have attained full language proficiency. The ACCESS test is divided into three overlapping tiers: A for ELLs who are beginning to learn English, B for intermediate level ELLs, and C for advanced level ELLs. The tiers allow the test to best represent the entire range of English language proficiency for this diverse student population. Four language domains are assessed by the ACCESS test: Reading, Listening, Writing, and Speaking. Scores for the four individual domains and scores for four sub-domains, Literacy, Oral, Comprehension, and Composite, are available to the student. Three of the four additional domains are derived by using a combination of two language domains (e.g., Literacy combines the Reading and Writing domains). The Composite domain is derived by using a combination of all four language domains. Students receive raw scores, scale scores, and proficiency levels for each domain on the ACCESS assessment (Gottlieb, 2007).

ACCESS scale scores allow scores across grades (within each of the eight domains) to be compared on a single vertical scale from Kindergarten to Grade 12. ACCESS scale scores range from 100 to 600. ACCESS proficiency level scores (PL) provide an interpretation of the scale scores. They range from 1.0 to 6.0 and describe student performance in terms of six language proficiency levels defined by the World-Class Instructional Design and Assessment (WIDA) Consortium, the creators of the assessment. The six proficiency levels are a product of WIDA's research on English Language Learners; they account for both the maturational and the language proficiency growth of English language learners. The decimal in the PL (e.g., 4.2) indicates the proportion within the proficiency level range that the student's scale score represents, rounded to the nearest tenth. Proficiency level scores do not represent interval data because the interval between corresponding scale scores for PL 2.2 to 3.2 is not necessarily the same as for PL 3.2 to 4.2. However, proficiency levels do represent near-interval data because as the proficiency levels increase so does the amount of English language proficiency described (Gottlieb, 2007).

As mentioned previously, the calculation of the ACCESS Composite scale score and proficiency level is unique because it combines the earned scores of all four language domains, whereas the other scores describe performance in one or two language domains. The percentages used in calculating the Composite score is 35% from the student's Reading score, 35% from Writing, 15% from Listening, 15% from Speaking. These weights used to derive the Composite scores were established by WIDA's ELL education experts. According to WIDA, these follow best practice in testing ELL students in that they encompass what is believed to be necessary for academic success.

The Composite proficiency level is also important because the Georgia Department of Education (GaDOE) uses it as the criterion for exiting ELL students from ESOL services. The GaDOE has adopted a Composite Proficiency Level of 5.0 or greater on the ACCESS assessment, Tier C form as the ESOL exit criterion. When a student meets the exit criterion, the student is considered proficient in English and is no longer eligible for English language assistance services (provided that other criteria are also met). This exit criterion is the preferred way to exit students from ESOL services; however, the GaDOE allows some flexibility in exiting students from ESOL programs. A Language Assessment Committee (LAC) can exit a student from ESOL services if the committee feels that exiting the student is best for the student. To be considered for exit via a

LAC, a student must first earn a Composite Proficiency Level of least a 4.0 on the ACCESS Tier C form and must pass the regular state assessment in reading.

End-of-Course Tests

The Georgia End-of-Course Tests (EOCT) are designed to measure student mastery of specific course curriculum standards. Only students who enroll in courses for which there is an EOCT take the test and students may encounter the courses in different sequences and at different grades. Because the EOCT are course-specific and given directly after instruction, they are more expansive than the other high school assessments given in Georgia, the Georgia High School Graduation Tests (GHSGT). The EOCT require students to demonstrate intimate knowledge of the subject area. Conversely, the GHSGT require students to demonstrate mastery of core academic content and skills that are learned across different course paths that students may elect to take during their high school careers.

EOCT are offered for the following eight courses: 9th Grade Literature and Composition, American Literature and Composition, Algebra I, Geometry, Physical Science, Biology, U.S. History, and Economics. Students receive raw scores, scale scores, performance levels, and grade conversion scores for each EOCT subject. Scale scores allow scores across administrations (within a course) to be compared on the same scale. Performance levels are an interpretation of the scale score. They categorize students into three groups: those not meeting the standard, those who meet the standard, and those who exceed the standard. Like all criterion-referenced tests developed for Georgia, these performance levels are set and reviewed by Georgia educators. The 'passing' performance levels-meets the standard and exceeds the standard- reflect what Georgia educators believe students *should* know after completing a particular course in a Georgia public school. The student's scale score on the EOCT is converted to a grade conversion score. This grade conversion score is used in the calculation of the student's course grade. Accordingly, EOCT function as standardized final exams for the course to which they apply.

For the current analysis, only the English Language Arts EOCT were studied: 9th Grade Literature and Composition (9th Lit) and American Literature and Composition (Am Lit). For these assessments, the range of possible scale scores is 200-600. Scores between 400 and 449 denote 'meeting the standard' and scores at 450 and above denote 'exceeding the standard'. Scores below 400 denote 'not meeting the standard'.

Standard Error of Measure (SEM)

On any test, the score a student receives, also known as the observed score, represents the student's true score plus measurement error. The true score is the score a student would receive on the test if no measurement error was present. Some amount of measurement error always exists, and classification of students into different performance levels (e.g., Does Not Meet, Meets, and Exceeds) can be imperfect, especially for the students whose true scores fall on the border of performance level cut scores.

Good test design procedures and good testing policy for the use of the scores can mitigate the adverse effects of measurement error. Clustering items around the cut scores in order to provide

the most accurate description of student performance and using retest administration data in addition to main administration data to make decisions about student retention are a couple of examples of good test design and policy. Additionally, the calculation of a standard error of measure (SEM) can help document how accurate a student's test score describes his or her knowledge and skills in the subject area. The SEM is a mathematical definition of a test's reliability and it represents the range of possible *true* scores for a particular student. For example, if a student earns a scale score of 500 on a test and the SEM is 5, the student's true score is likely between a 495 and a 505. Moreover, the student would most likely earn a score between 495 and 505 on an equivalent form of the test if he or she were to take it. Because a student's true score is most likely to fall within a standard error band around his or her observed score, the SEM along with the scale score is a good metric for describing individual student performance especially for students on the cusp of meeting the standard.

The SEM can be calculated at the total test level and also for each score in the test score range. The SEM for a particular score is called the conditional standard error of measure (CSEM) because the amount of error is conditioned upon where the score is on the range. Scores falling at the tails of the distribution (very top or very bottom) have more error than scores falling toward the middle of the distribution. Consequently, the CSEM at the 'meets the standard' cut score is perhaps the best way to look for misclassification across the 'does not meet' and 'meets' performance levels. In spring 2007, the CSEM around the 9th Lit EOCT 'meets' cut score was 9 and the CSEM around the Am Lit 'meets' cut score was 8. This means that a student who earned a 400 on the 9th Lit test is likely to have true score between 391 and 409. A student who earned a 400 on the Am Lit test is likely to have a true score between 392 and 408.

Method

Students who took the ACCESS for ELLs and one of the English language arts EOCT (i.e., 9th Lit or Am Lit) in the spring of 2007 were potential candidates for inclusion in this study. Those students who had valid scores in both assessment files were included in the analysis. ELL status is defined by participation in the ACCESS assessment. Because ELL status on the EOCT is self-report, ELL status for this study was derived from the ACCESS assessment.

Simple and multiple ordinary least squares (OLS) regression analyses were computed to determine the relationships between a student's score on the ACCESS assessment and his or her score on the ELA EOCT. The regression equations for predicting success on the English Language Arts EOCT from a student's score on ACCESS were explored. Preliminary analyses showed that ACCESS proficiency levels had more predictive power than ACCESS scale scores. Because of this finding and because educators are already familiar with the proficiency levels, specifically the Composite proficiency level, the first analyses focused on the relationship between the ACCESS Composite proficiency levels and EOCT scale scores. Other analyses were conducted using the proficiency levels from all four language domains to predict EOCT scale scores. These secondary analyses, which augment the predictive capacity of the ACCESS test, are discussed in later sections.

Results

The 9th Lit and Am Lit analysis were run independently. The 9th Lit results are presented first followed by the Am Lit results.

ACCESS Assessment and 9th Grade Literature and Composition EOCT

There were 962 students who were matched across the spring 2007 ACCESS for ELLs and 9th grade Literature and Composition EOCT data files. For these students, ACCESS Composite Proficiency Levels ranged from 1.2 to 6.0. These observed Composite proficiency levels covered all but the very bottom of the ACCESS proficiency level range. The 9th grade Literature and Composition scale scores for these students ranged from 200 to 484. These scores did not cover the whole range of possible scale scores (200-600). Specifically, no student in the matched data set scored at the upper end of the 9th Lit scale and less than 25% of students earned a passing score.

Despite the low test scores for the ELL students in the sample, the score distributions for both assessments were mostly normal. Table 1 shows descriptive statistics for the students included in this analysis by assessment. Figures 1 and 2 depict the distribution of scores across the assessments for the students included.

Table 1. Descriptive Statistics for Students Taking Both the ACCESS for ELLs and 9th Grade Literature and Composition EOCT, Spring 2007.

Statistic	9 th Literature and Composition Scale Scores	ACCESS Composite Proficiency Level
N	962	962
Mean	376.8	3.4
Std. Deviation	26.1	1.0
Median	376.0	3.4
Skewness	0.1	0.3
Minimum	200	1.2
Maximum	484	6
Score at 25th Percentile	357	2.6
Score at 50th Percentile	376	3.4
Score at 75th Percentile	394	4.1

Figure 1. Distribution of Scale Scores for the 9th Grade Literature and Composition EOCT, Spring 2007.

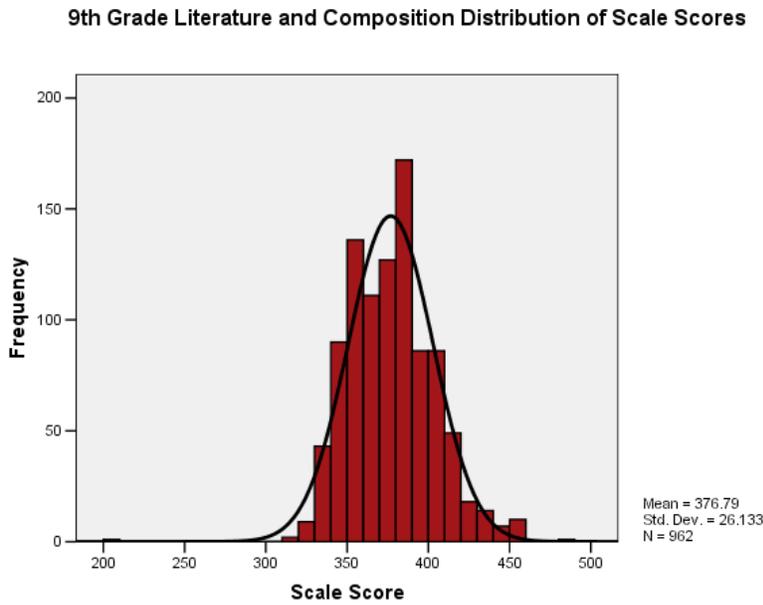
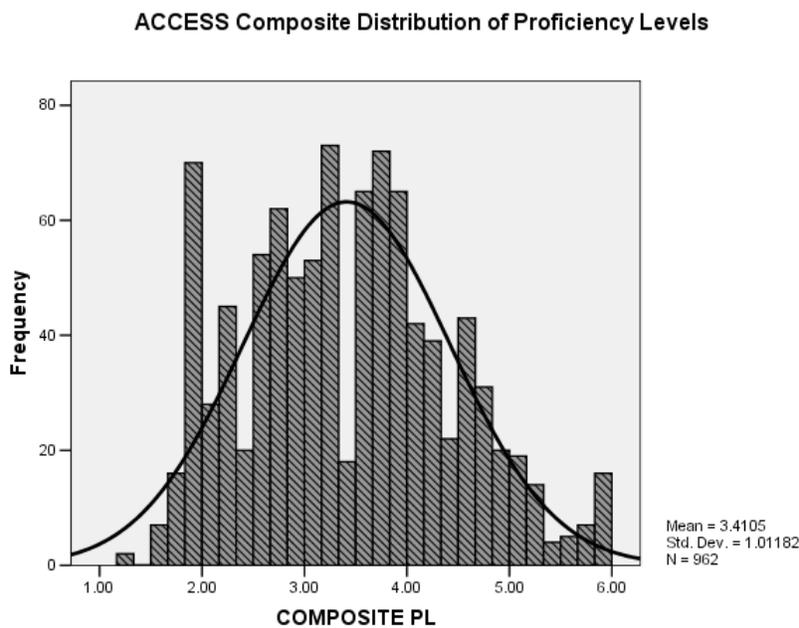


Figure 2. Distribution of Proficiency Levels for the ACCESS Composite, Spring 2007.



In Table 1, one can see that less than 25% of the ELLs who took the 9th Lit EOCT earned a scale score at the ‘meets the standard’ performance level. When comparing the 9th Lit score distribution

that was observed in this matched data set with what was reported for ELL students during the spring 2007 administration of the 9th Lit test, the score distributions were very similar. In the full spring 2007 9th Lit data file, 651 students who were classified as ELL took the 9th Lit EOCT. The range of scores observed for these ELL students was 200 to 458 with a mean score of 376.9.

The purpose of reporting how ELLs from the testing event performed along with how ELLs from the matched sample performed is to show that the matched sample provides an accurate representation of ELL test performance. While the performance from the two sources are similar, the difference in the number of ELLs from the matched set (962) and what was reported from the 9th Lit data file (652) was large. It is important to note that for the EOCT, ELL classification is derived from the answer document for which most cases students self-report via marking a bubble. ELL status for the matched set (n=962) comes from participation in the ACCESS administration; because only ELL students should take the assessment. It is believed that coding ELL status from participation in ACCESS is more accurate.

Even though the 9th Lit scores in the matched set are relatively low, the full range of ACCESS Composite PLs are observed. Because these proficiency levels are also mostly normally distributed, it is likely that the limited range of EOCT scores is not caused by a limited group of students who took both assessments. Instead, it seems that the range accurately covers an expected distribution of performance for ELL students.

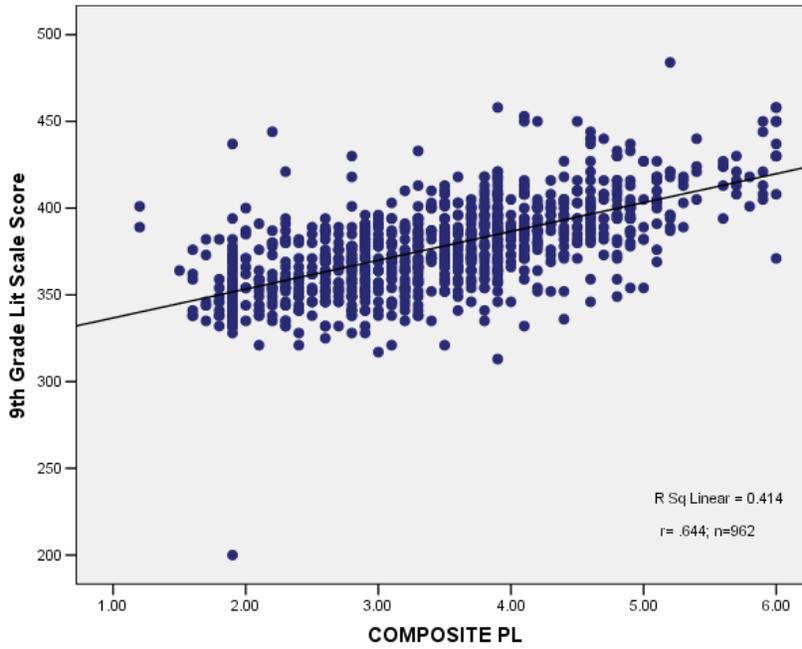
Normal distribution of scores from a non-truncated data set is only one of several assumptions that must be met when using a correlational study to produce valid conclusions about relationships. While it would be ideal to have a wider range of observed 9th Lit scale scores to more accurately describe the relationship between ACCESS and EOCT at the upper end of the scale score range, the matched data set satisfactorily meets the assumption of normality. Other assumptions that should be met when using correlation statistics are best examined by residual analysis. Residuals are the differences between a score that is predicted from the correlation/regression model and the score that was observed during the test event. These assumptions will be discussed following the correlation results.

The Pearson correlation coefficient (r) between ACCESS Composite proficiency levels and 9th grade Literature and Composition scale scores was .644. Correlation coefficients range between -1 and +1 depending on the strength and direction of the relationship between the two variables. This correlation clearly indicates that increases in the ACCESS Composite proficiency level are associated with increases in 9th grade Literature and Composition scale scores.

The R-squared statistic is another way of reporting the strength of a relationship between variables. It is the amount of variance in one variable that can be accounted for by the other variable. R-squared can be stated as a percent. For this relationship, 41% of the variation in the 9th Grade Literature and Composition scale scores can be accounted for by the variation in ACCESS Composite proficiency levels. Figure 3 shows a graphical representation of how the scale scores are related. On the scatterplot, each dot represents a student. The right upward slope indicates that the correlation is positive and the tight grouping of scores shows the strength of the relationship. The line drawn through the scatterplot is the regression line. It represents the

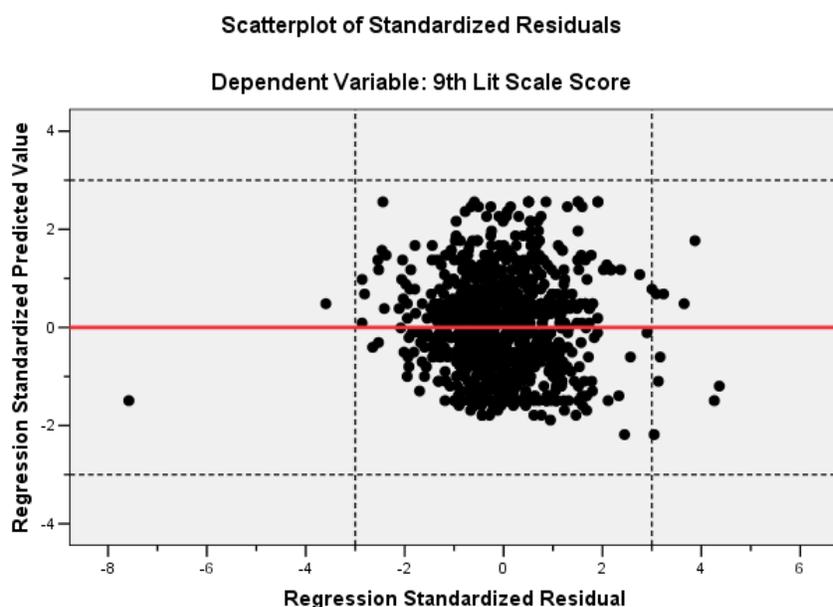
line of best fit, minimizing the squared distance of each point to the line, and helps show the direction of the correlation.

Figure 3. Relationship of ACCESS Composite Proficiency Levels and 9th Grade Literature and Composition Scale Scores.



Other assumptions that must be met when using a correlational study to produce valid conclusions about relationships can be examined using Figure 3. One assumption is that the relationship between the two variables must be linear, not curved. The pattern of data points in Figure 3 show that the relationship between the ACCESS Composite proficiency levels and 9th Lit scale scores is linear.

Another assumption of using a correlation to produce valid conclusions is the absence of outlying data points that may exert disproportionate influence on the correlation coefficient. Outliers are observations that fall outside the general linear pattern of the regression line. When looking at the correlation scatterplot, one can see at least one outlier in the bottom left corner. However, a more precise way to find outliers is by analyzing the residuals. From the residual analysis, 12 observations emerged as outliers. These were those datapoints that were more than 3.00 standard deviations away from the average difference between what was predicted and what was observed. Figure 4 plots the standardized residuals.

Figure 4. Scatterplot of Residuals: Predicted 9th Lit Scores and Observed 9th Lit Scale Scores

Like the correlation scatterplot, each dot represents a student on the residual plot. But, instead of representing scores, these dots represent how close the predicted scale score was to the observed scale score. The closer the dots are to zero, the smaller the difference between predicted and observed scale score. The dotted lines help illustrate where the outliers are. Predicted scores that are not at all close to the observed score are outliers; they fall outside the intersections of the dotted lines. Predicted scores that are relatively close to the observed scores fall inside the intersection of the dotted lines.

In the residual scatterplot, one can see that most of the 12 outliers indicate that the students' 9th Lit scale score was much higher than what was expected from their ACCESS Composite proficiency levels. Only two of the 12 had 9th Lit scale scores much lower than what was expected from their ACCESS Composite proficiency levels. When removing these 12 outliers, the r coefficient was .682. The correlation was made stronger by approximately 4 hundredths of a point from the original correlation coefficient of .644. However, out of 962 total data points, these 12 outliers were not enough to make the correlation invalid. Furthermore, there was no logical reason to take these outlying students out of the model, so they were left in.

Lastly, the assumption of homoscedasticity must be met before valid conclusions can be drawn from a correlation analysis. Homoscedasticity is the concept that the predicted model works equally for all levels of the independent variable. That is, students who have lower ACCESS Composite proficiency levels are just as likely to receive an accurate prediction about their EOCT

score as students who have middle and higher ACCESS Composite proficiency levels. Again, by looking at the residual plots, one can see if this assumption is met.

In Figure 4, one can see that the number and general pattern of dots above the line and below the line are similar and the overall plot has no 'shape'. This means that relatively equal amounts of prediction error are found across all levels of ACCESS Composite scores and the assumption is met. Because there were relatively few students performing at the upper end of the ACCESS Composite proficiency level scale, one may theorize that the 9th Lit scale score prediction may not be quite as accurate at the upper ends of the scale as it is at the lower and middle proficiency levels, where more of the students in the sample were clustered. This theory could be supported by the slight spreading of the residuals at the top middle. However, this spreading is not enough to nullify the assumption of homoscedasticity.

While the correlation between the Composite PL and 9th Lit scale scores was moderately strong, a second coefficient for describing the relationship between the two assessments was computed. The Composite proficiency levels are derived from summing a percentage of each of the four language domains measured by the ACCESS assessment. According to the Composite calculation, the Reading and Writing domain scores contribute more to the Composite score than the Listening and Speaking domains. Perhaps, using different weights of the four domain scores could explain more of the variation in 9th Lit scale scores.

A standard multiple regression was calculated using the proficiency levels for the four language domains, Reading, Listening, Writing, and Speaking as the independent variables and 9th Lit scale score as the dependent variable. When each of the four domains was entered separately, the regression coefficient (R) was .664; that is, 44% of the variation in the 9th Lit scale scores could be accounted for by the variation in the four ACCESS Language domain proficiency levels. Note that this model was slightly better at predicting 9th Lit scale scores ($R=.664$) than the model that used only the Composite proficiency level as the independent variable ($r=.644$).

A benefit of calculating the second regression analysis is that one can examine the relative predictive importance of each domain proficiency level by looking at the Beta weights and b-values in addition to examining the overall strength of the second predictive model. Beta weights are the standardized, individual contribution that each domain proficiency level makes to the model, when the other domain proficiency levels are held constant. Non-standardized individual contributions of each domain proficiency level are the b-values. Beta weights are best used to compare the influence of each domain to each other. The b-values are used in the actual regression equation for predicting an EOCT scale score from ACCESS proficiency levels.

All four language domain proficiency levels added predictive power to the model. However, the Reading domain proficiency level emerged as the most important predictor of 9th Lit scale score. Listening and Writing domain proficiency levels contributed a good amount of predictive power and the Speaking domain proficiency level contributed a small amount. Table 2 shows the Beta weights, b-values and the t statistic of significance for the four individual ACCESS language domains of the second predictive model for 9th Lit.

Table 2. Weights for the Proficiency Level Domains of Reading, Writing, Listening, and Speaking for the Four Domain Predictive Model.

Language Domain	Beta weights (standardized)	t	significance	b values (non-standardized)	Std. Error of b values
READING PL	0.36	11.03	p <.01	7.21	0.65
WRITING PL	0.15	5.06	p <.01	5.70	1.13
LISTENING PL	0.19	5.76	p <.01	4.57	0.79
SPEAKING PL	0.09	3.23	p <.01	1.29	0.40

Using the simple and multiple regression coefficients and constants obtained from the two analyses, equations for using ACCESS scores to predict 9th Lit scale scores were created. The equations are listed in Table 3 and follow the formulas below, where **y** is the predicted value, **x** is the independent variable or variables, **b** is the non-standardized weight of the individual contribution(s) of the independent variable(s), and **a** is the constant.

- Regression (Composite Proficiency Level Only)
 $y = b_1x_1 + a$
- Multiple regression (Four Proficiency Level Domains)
 $y = (b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4) + a$

Table 3. Regression Equations for the Two Predictive Models.

Information Used to Predict 9 th Lit from ACCESS	Equation
Composite PL to predict EOCT 9 th Lit Scale score	<i>Predicted 9th Lit Scale Score = (16.620 x ACCESS Composite PL) + 320</i>
Four Language Domain PL to predict EOCT 9 th Lit Scale score	<i>Predicted 9th Lit Scale Score = (7.213 x ACCESS Reading PL) + (5.700 x ACCESS Writing PL) + (4.574 x ACCESS Listening PL) + (1.292 x ACCESS Speaking PL) + 310</i>

Impact analyses were computed on the sample to gauge the prediction accuracy of the two regression models shown in Table 3 on 9th Lit pass/fail performance. First, the Composite only and the four domain regression models were used to predict two scale scores for each student. Then, the pass/fail impact criterion was applied to these predicted scores. Students with predicted scale scores of 400 and above were categorized as ‘passing’ and students with scale scores lower than 400 as ‘failing’. This 400 cut point is the actual passing cut score for the ELA EOCT.

It is important to note that the terms ‘pass’ and ‘fail’ used in these analyses refer to students predicted and actual outcome of the EOCT, not the course. For the purpose of these impact analyses, ‘pass’ means meeting or exceeding the standard set for the test. ‘Fail’ means not meeting the standard set for the test.

The number and percent of students who were predicted to pass the 9th Lit test using the Composite only are listed in Table 4. The number and percent of students who were predicted to pass the 9th Lit test using the four language domain Proficiency Levels regression model are listed in Table 5. Also listed in Tables 4 and 5 is the actual outcome observed from the 9th Lit test for the students in the sample. Note that the same sample that was used to derive the predictive models was used for the impact analysis.

Table 4. Impact of Predicted Model (Composite PL on 9th Lit Pass rate)

9th Lit Pass/Fail	PREDICTED OUTCOME		OBSERVED OUTCOME	
	N	Percent	N	Percent
Does Not Pass	877	91.2	777	80.8
Pass	85	8.8	185	19.2
Total	962	100.0	962	100.0

Table 5. Impact of Predicted Model (Four Language Domain PL on 9th Lit Pass rate)

9th Lit Pass/Fail	PREDICTED OUTCOME		OBSERVED OUTCOME	
	N	Percent	N	Percent
Does Not Pass	859	89.3	777	80.8
Pass	103	10.7	185	19.2
Total	962	100.0	962	100.0

When comparing the observed outcome to what was predicted from the two regression models, it appeared that both models under-predicted student performance on the 9th Lit test. Many fewer students were predicted to pass the 9th Lit test than actually did pass the test. In an attempt to mitigate the under-predicting quality of the models, the lower bound of the CSEM of the 9th Lit passing cut score was used as a hypothetical ‘passing’ criterion and the pass/fail impact of the regression models was calculated again. That is, students with predicted scale scores of 391 and above were categorized as ‘passing’ and those with predicted scale scores below 391 were categorized as ‘failing’. These hypothetical cut points represented the lower boundary of the conditional standard error of measure of the actual cut score for the 9th Lit test. Tables 6 and 7 show the impact of the predicted models when the lower CSEM bound of the true cut score was used as the ‘passing’ criterion.

Table 6. Impact of Predicted Model with 391 as the Hypothetical Passing Cut Score (Composite PL on 9th Lit Pass Rate)

9th Lit Pass/Fail	PREDICTED OUTCOME		OBSERVED OUTCOME	
	N	Percent	N	Percent
Does not Pass	767	79.7	777	80.8
Pass	195	20.3	185	19.2
Total	962	100.0	962	100.0

Table 7. Impact of Predicted Model with 391 as the Hypothetical Passing Cut Score (Four Language Domain PL on 9th Lit Pass Rate)

9th Lit Pass/Fail	PREDICTED OUTCOME		OBSERVED OUTCOME	
	N	Percent	N	Percent
Does Not Pass	741	77.0	777	80.8
Pass	221	23.0	185	19.2
Total	962	100.0	962	100.0

When using the lower bound of the CSEM range (391) as the passing cut score, the predictive models seemed more similar to the actual pass rate of the students in the sample.

Table 8 illustrates the 9th Lit scale score outcome from the Composite only predictive model. It lists a series of ACCESS Composite proficiency levels and the predicted 9th Lit EOCT scale scores associated with these PLs when the regression equation from this analysis is applied. In other words, if a student was to perform at the ACCESS level listed on the left, his or her *predicted* scale score for the 9th Lit EOCT is listed on the right. The higher the Composite PL, the higher the 9th Lit scale score is predicted to be. The lowest 9th Lit scale score listed in Table 8 is the scale score that falls at the lower bound of one SEM of the passing cut score (400). According to the regression model, students who perform at these levels on the ACCESS Composite Proficiency Level *could* be potentially successful on the 9th Lit EOCT. The Composite Proficiency Level regression equation for 9th Lit, which was used to derive these values, can be found in Table 3.

Table 8. 9th Lit Scale Scores Predicted from ACCESS Composite Proficiency Levels

ACCESS Composite Proficiency Level	Predicted 9 th Lit Scale Score from Regression Model
4.3	391
4.4	393
4.5	395
4.6	396
4.7	398
4.8	400
4.9	401
5.0	403
5.1	405
5.2	406
5.3	408
5.4	410
5.5	411
5.6	413
5.7	415
5.8	416
5.9	418
6.0	420

While the relationship between ACCESS Composite PL and 9th Lit scale score is strong, there is still some error associated with predicting a 9th Lit scale score from a Composite PL. This table can be most useful in confirming an educator's beliefs about a particular student's potential success in a 9th Lit course. It should be taken as a piece of information that can assist in the course enrollment decision-making process.

American Literature and Composition

The same design and analyses were conducted for the Am Lit EOCT. There were 1,005 students who were matched across the ACCESS for ELLs and American Literature and Composition EOCT data files. For these students, ACCESS Composite Proficiency Levels ranged from 1.3 to 6.0. These observed Composite proficiency levels covered all but the very bottom range of possible ACCESS scores. The observed Am Lit scale scores for these ELL students ranged from 346 to 474. Similar to 9th Lit, these scale scores did not cover the whole range of possible scale scores (200-600). No student in the matched data set scored at the lower or upper ends of the Am Lit scale. Unlike 9th Lit, however, the mean and median scale score for ELL students in the matched set was near the passing cut score of 400, $\mu=395$ and median=394.

The scale scores and proficiency levels for students across both assessments were normally distributed. Table 9 shows descriptive statistics for the students included in this analysis by assessment. Figures 5 and 6 display the distribution of scores across tests for students who took both assessments.

Table 9. Descriptive Statistics for Students Taking Both the ACCESS for ELLs and American Literature and Composition EOCT, Spring 2007.

Statistic	American Literature and Composition Scale Scores	ACCESS Composite Proficiency Level
N	1005	1005
Mean	395.0	3.8
Std. Deviation	20.4	1.0
Median	394.0	3.8
Skewness	0.3	0.2
Minimum	346	1.3
Maximum	474	6.0
Score at 25th Percentile	379	3.1
Score at 50th Percentile	394	3.8
Score at 75th Percentile	409	4.6

Figure 5. Distribution of Scale Scores for the American Literature and Composition EOCT, Spring 2007.

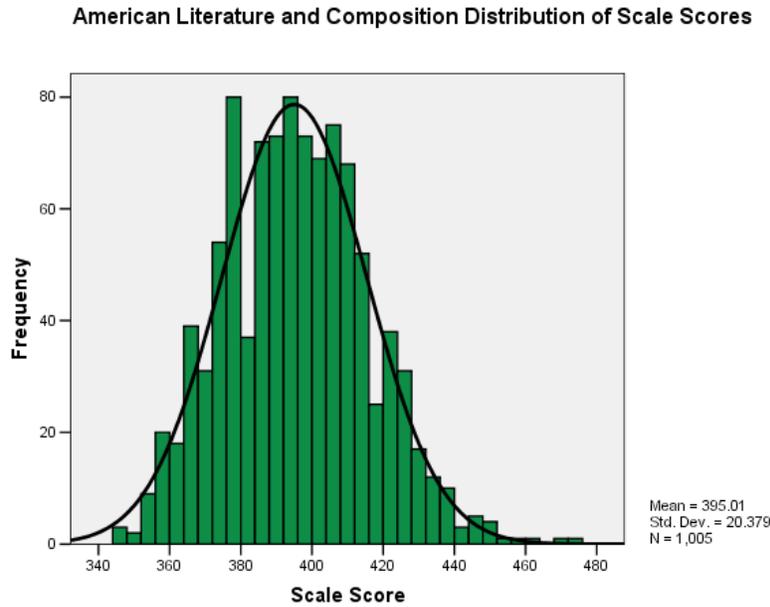
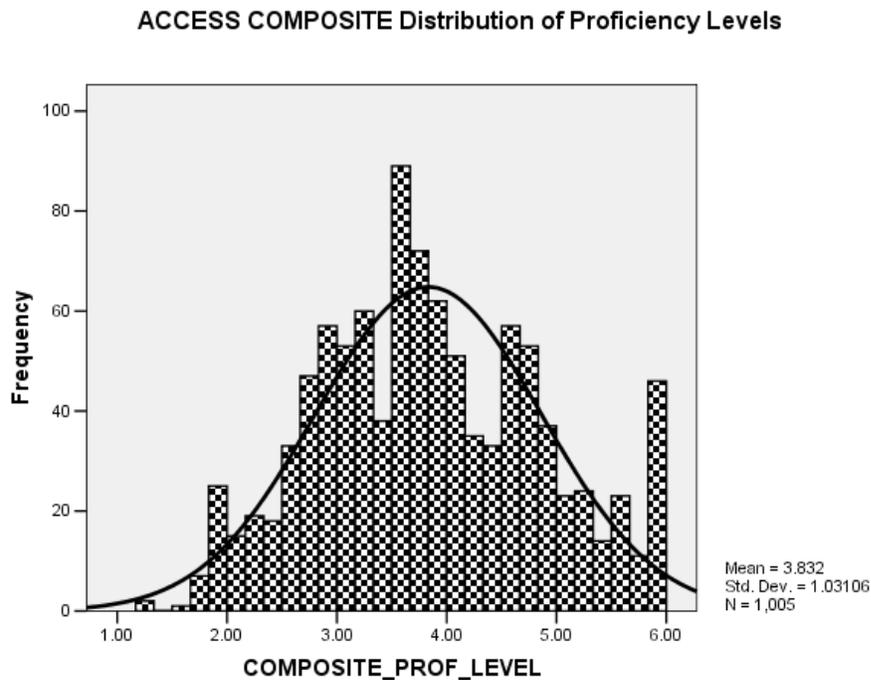


Figure 6. Distribution of Proficiency Levels for the ACCESS Composite, Spring 2007.

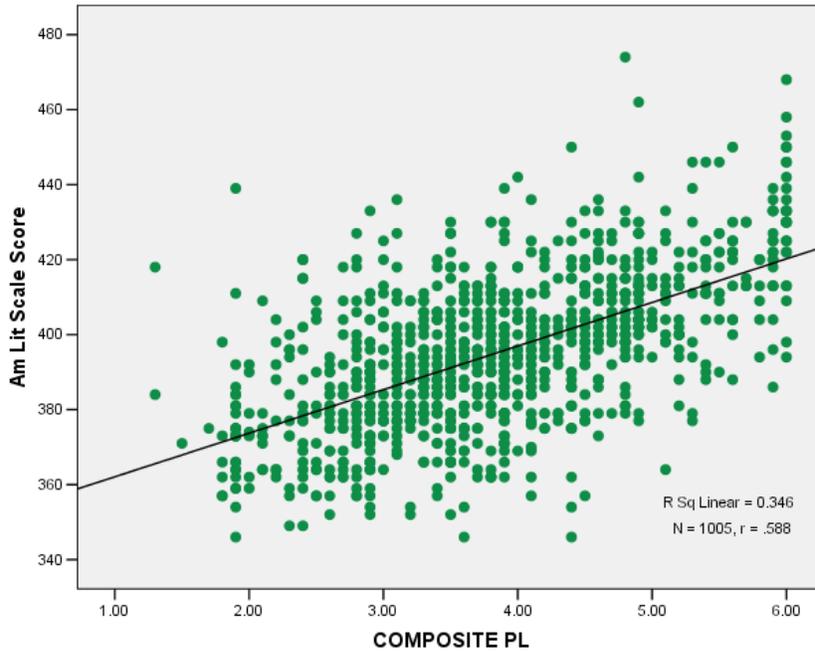


Because the full range of ACCESS Composite PLs are observed and are normally distributed, the limited range of EOCT scores is not likely due to a limited selection of students who took both assessments. Consistent with 9th Lit, it seems that few students with high ACCESS proficiency levels render few students with high scores on the Am Lit EOCT. As with the 9th Lit analysis, the distribution of scores from the matched data set was compared to what was reported for the spring 2007 Am Lit administration. The range of Am Lit scores observed in the sample is very similar to what was reported for Am Lit for the spring 2007 administration. In the spring 2007 American Literature data file, 659 students who were classified as ELL took the Am Lit EOCT. The scores for these ELL students ranged from 346 to 468. The mean scale score was 393.9. Again, the numbers of ELL students differed greatly between the matched set and the EOCT data, but it is believed that the matched data represent a more accurate tally of ELL students who took the Am Lit EOCT.

While it would be ideal to have a wider range of observed Am Lit scale scores to more accurately describe the relationship at the high and low ends of the scale score range, the matched data set satisfactorily meets the assumption of normality. Other checks on the assumptions of the correlation statistic for producing valid conclusions about relationships will be discussed in the residual analysis following the correlation results.

The Pearson correlation coefficient (r) between ACCESS Composite proficiency levels and American Literature and Composition scale scores was .588. This correlation indicates that increases in the ACCESS Composite proficiency levels are associated with increases in American Literature and Composition scale scores; however the relationship is not as strong as was the case with 9th Lit. For this relationship, 35% of the variation in the American Literature and Composition scale scores can be explained by the variation in ACCESS Composite proficiency levels (R -squared = .346). Figure 7 shows a graphical representation of how the scale scores are related. On the scatterplot, each dot represents a student. The right upward slope indicates that the correlation is positive and the grouping of student scores shows the strength of the relationship. The line drawn through the scatterplot is the regression line. It represents the closest straight line to all points of the plot. It helps show the direction of the correlation.

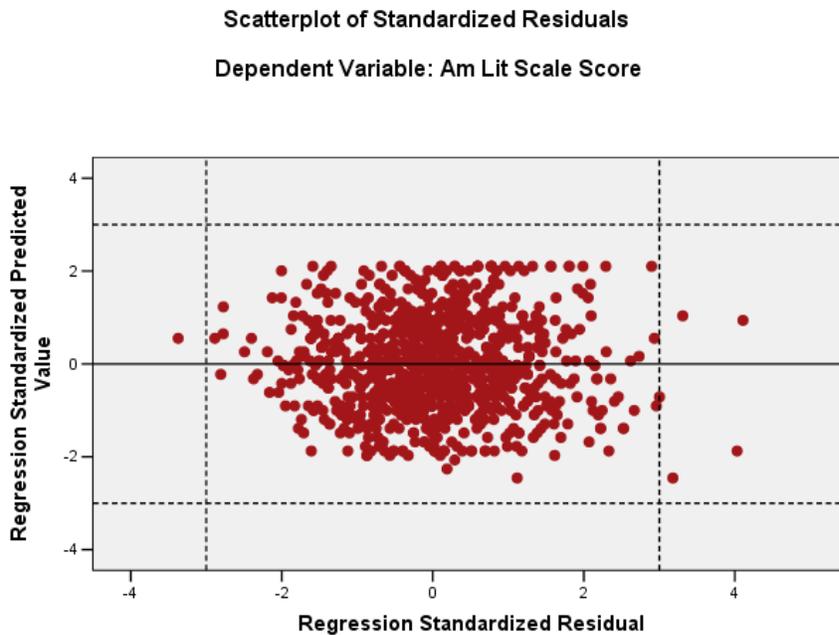
Figure 7. Relationship of ACCESS Composite Proficiency Levels and American Literature and Composition Scale Scores.



The assumption of linearity can be examined by looking at Figure 7. The pattern of data points in the Figure 7 shows that the relationship between the ACCESS Composite proficiency levels and Am Lit scale scores is linear.

Another assumption of using a correlation to produce valid conclusions is the absence of outlying data points that may exert disproportionate influence on the correlation coefficient. From the residual analysis, six observations emerged as outliers. These were those datapoints that were more than 3.00 standard deviations away from the average difference between what was predicted and what was observed. Figure 8 is the scatterplot of the standardized residuals. The dotted lines help illustrate where the outliers are. Predicted scores that are not at all close to the observed score are outliers; they fall outside the intersection of the dotted lines. Predicted scores that are near the observed scores fall inside the intersection of the dotted lines. In the residual plot, one can see that most of the six outliers indicate that the students' Am Lit scale score were much higher than what was expected from their ACCESS Composite proficiency level. Only one of the six had Am Lit scale scores much lower than what was expected from their ACCESS Composite proficiency level. When removing these six outliers, the correlation coefficient was .607. This coefficient is approximately two hundredths of a point stronger than the original correlation coefficient of .588. Out of 1,005 total data points these six outliers were not enough to make the correlation invalid; furthermore, since there was no logical reason to take these outlying students out of the model, they were left in.

Figure 8. Scatterplot of Residuals: Predicted Am Lit Scores and Observed Am Lit Scale Scores



The assumption of homoscedasticity can also be reviewed by looking at the residual plot. In Figure 8, one can see that relatively equal amounts of prediction error are found across all levels of ACCESS Composite scores. It appears this assumption is met.

Like in 9th Lit, a second coefficient for describing the relationship between the ACCESS and Am Lit assessments was computed. In this second regression analysis, each of the four ACCESS domain scores was entered into the model separately. The regression coefficient (R) was .593; that is, 35% of the variation in the Am Lit scale scores can be explained by the variation in the four ACCESS Language domain proficiency levels. Similar to findings in 9th Lit, this second regression model was slightly better at predicting Am Lit scale scores than the model that used only the Composite proficiency level as the independent variable ($r=.588$).

By examining the Beta weights one can see the relative predictive importance of each domain proficiency level. All four language domain proficiency levels added predictive power to the Am Lit model. However, the Reading domain proficiency level emerged as the most important predictor of Am Lit scale score; the Listening and Writing domain proficiency levels contributed a good amount of predictive power; the Speaking domain proficiency level contributed a small amount. This same pattern was noted in the 9th Lit analysis. Table 10 shows the Beta weights for the four ACCESS language domains for the second Am Lit model.

Table 10. Weights for the Proficiency Levels Domains of Reading, Writing, Listening, and Speaking for the Four Domain Predictive Model.

Language Domain	Beta weights (standardized)	t	significance	b values (non-standardized)	St. Error of b values
READING PL	0.34	10.59	p <.01	4.87	0.46
WRITING PL	0.19	6.18	p <.01	5.23	0.85
LISTENING PL	0.14	4.36	p <.01	2.70	0.62
SPEAKING PL	0.08	2.65	p <.01	0.89	0.34

Using the simple and multiple regression coefficients and constants obtained from the two analyses, equations for using ACCESS proficiency levels to predict Am Lit scale scores were created. The equations are listed in Table 11.

Table 11. Regression Equations for the Two Predictive Models; ACCESS Scores as Independent Variables, Am Lit scores as Dependent Variable.

Information Used to Predict Am Lit from ACCESS	Equation
Composite PL to predict EOCT Am Lit Scale score	<i>Predicted Am Lit Scale Score = (11.625 x ACCESS Composite PL) + 350</i>
Four Language Domain PL to predict EOCT Am Lit Scale score	<i>Predicted Am Lit Scale Score = (4.870 x ACCESS Reading PL) + (5.225 x ACCESS Writing PL) + (2.695 x ACCESS Listening PL) + (.894 x ACCESS Speaking PL) + 342</i>

Lastly, impact analyses were computed on the sample to gauge the prediction accuracy of the two regression models. The number and percent of students who were predicted to pass the Am Lit test using both regression equations are listed in Tables 12 and 13. The impact data from the Composite Proficiency Level regression model are listed in Table 12. The impact data from the four language domain Proficiency Levels regression model are listed in Table 13. Tables 12 and 13 also include the actual observed outcome from the Am Lit test for the students in the sample, for comparison purposes. As with 9th Lit, the same sample that was used to derive the predictive models was used for the impact analysis.

Table 12. Impact of Predicted Model (Composite PL on Am Lit Pass rate)

	Predicted Outcome		Observed Outcome	
Am Lit Pass/Fail	N	Percent	N	Percent
Does Not Pass	673	67.0	591	58.8
Pass	332	33.0	414	41.2
Total	1005	100.0	1005	100.0

Table 13. Impact of Predicted Model (Four Language PL Domains on Am Lit Pass rate)

	Predicted Outcome		Observed Outcome	
Am Lit Pass/Fail	N	Percent	N	Percent
Does Not Pass	620	61.7	591	58.8
Pass	385	38.3	414	41.2
Total	1005	100.0	1005	100.0

When comparing the observed outcome to what was predicted from each regression model, it appeared that the four domain model predicted pass/fail performance relatively well. The Composite-only model slightly under-predicted student performance on the Am Lit test (i.e., less students were predicted to pass the Am Lit test than actually did pass the test).

Tables 14 and 15 show the impact of the predicted models when the lower CSEM bound of the cut score (392) was used to as the ‘passing’ criterion for Am Lit.

Table 14. Impact of Predicted Model with 392 as the Hypothetical Passing Cut Score (Composite PL on Am Lit Pass rate)

	Predicted Outcome		Observed Outcome	
Am Lit Pass/Fail	N	Percent	N	Percent
Does Not Pass	429	42.7	591	58.8
Pass	576	57.3	414	41.2
Total	1005	100.0	1005	100.0

Table 15. Impact of Predicted Model with 392 as the Hypothetical Passing Cut Score (Four Language Domain PL on Am Lit Pass Rate)

	Predicted Outcome		Observed Outcome	
Am Lit Pass/Fail	N	Percent	N	Percent
Does Not Pass	417	41.5	591	58.8
Pass	588	58.5	414	41.2
Total	1005	100.0	1005	100.0

The second impact predictions where the lower bound of the CSEM for the Am Lit passing cut score (392) was used as a hypothetical ‘passing’ cut score did not appear to be as accurate in predicting pass/fail performance on the Am Lit EOCT. When the 392 criterion was used, the models substantially over-predicted student performance on the Am Lit test: many more students were *predicted* to pass the Am Lit test than actually did pass.

Table 16 illustrates the Am Lit scale score outcome from the Composite only predictive model. It lists a series of ACCESS Composite proficiency levels and the predicted Am Lit EOCT scale scores associated with these PLs when the regression equation from this analysis is applied. In other words, if a student was to perform at the ACCESS level listed on the left, his or her *predicted* scale score for the Am Lit EOCT is listed on the right. The higher the Composite PL, the higher the Am Lit scale score is predicted to be. The lowest Am Lit scale score listed in Table 16 is the scale score that falls at the lower bound of one SEM of the passing cut score (400). According to the regression model, students who perform at these levels on the ACCESS Composite Proficiency Level *could* be potentially successful on the Am Lit EOCT. The Composite Proficiency Level regression equation for Am Lit, which was used to derive these values, can be found in Table 11.

Table 16. Am Lit Scale Scores Predicted from ACCESS Composite Proficiency Levels

ACCESS Composite Proficiency Level	Predicted Am Lit Scale Score from Regression Model
3.6	392
3.7	393
3.8	394
3.9	395
4.0	397
4.1	398
4.2	399
4.3	400
4.4	401
4.5	402
4.6	403
4.7	405
4.8	406
4.9	407
5.0	408
5.1	409
5.2	410
5.3	412
5.4	413
5.5	414
5.6	415
5.7	416
5.8	417
5.9	419
6.0	420

While the relationship between ACCESS Composite PL and Am Lit scale score is moderately strong, there is still some error associated with predicting an Am Lit scale score from a Composite PL. This table can be most useful in confirming an educator’s beliefs about a particular student’s potential success in an Am Lit course. It should be taken as a piece of information that can assist in the course enrollment decision-making process.

Discussion

The results from this study show that moderate to moderately strong positive relationships exist between students’ scores on the ACCESS assessment and their scores on the English Language Arts

EOCT. The relationships found between the a) ACCESS Composite Proficiency Level and 9th Lit scale score and b) the ACCESS four language domain proficiency levels and the 9th Lit scale score were moderately strong. The relationships found between the c) ACCESS Composite Proficiency Level and Am Lit scale score and d) the ACCESS four language domain proficiency levels and the Am Lit scale score were moderate and moderately strong, respectively. In both sets of calculations, the regression that used all four domain proficiency levels was slightly better at predicting ELA scale scores than the Composite only regression. The coefficient was about two hundredths of a point stronger in both instances.

While the four domain models were stronger than the Composite only models, the Composite only models are still useful for describing a student's potential for success in an ELA course. Very few students performed extremely different from what the Composite only regression model predicted. Moreover, these few outlying students seemed to weaken the correlations between the ACCESS Composite proficiency levels and ELA scale scores, which further indicates that the relationships found between Composite proficiency levels and scale scores are valid.

Where the four domain regression model notably enhances the decision making process is in the information it provides about the individual contributions of each domain proficiency level to the ELA scale score. All of the four domains appeared to contribute positively to the ELA scale score, and interestingly, the relative amount of individual contributions by domain were similar across the ELA courses. The Reading domain contributed the most to the ELA scale score by far. The Writing and Listening domains were the two next highest contributors and the Speaking domain contributed the least to the two ELA scale score. This means, of course, that the student's ACCESS Reading domain proficiency level should get the most careful scrutiny before enrolling her or him into an ELA course; her or his Speaking domain proficiency level the least careful scrutiny.

These individual domain contributions most likely explain why the Composite only and four domain regression models were so similar in strength. In the calculation of the ACCESS Composite proficiency level, the Reading and Writing domains represent the greatest domain percentages used. Thirty-five percent of the Composite score comes from Reading and 35% comes from Writing. Fifteen percent of the Composite score comes from Listening and 15% comes from Speaking. In the four domain regression model derived from this analysis, Reading, Writing, and Listening emerged as important contributors to the ELA score. While not a perfect match, the variance of Composite PL associated with the Reading domain in the four domain model, and to a lesser extent the variance of Composite PL associated with the Writing domain were similar to the proportions used in the calculation of the Composite score. This finding is encouraging and provides support for using the Composite proficiency level as an 'overall' English proficiency score.

Nonetheless, correlations and regression coefficients alone are not sufficient to establish that a high score on the ACCESS assessment causes a high score on the ELA EOCT. However, the information gleaned from the computations show that ELA EOCT scale scores can be predicted from ACCESS proficiency levels. Both sets of correlations observed in this study fell within the moderate to moderately strong range. The scale score prediction will not be 100% accurate for each student, but one will likely be able to make a good approximation at the ELA performance

level outcome for each student (i.e., does not meet/meets and exceeds). By using the regression models obtained from this study, educators can identify ELL students who would be more or less likely to learn the content standards of an English Literature course from the overall population of ELLs at their school or district. How accurate the prediction will be for each individual student depends on how closely the student's test behavior on both assessments fits with the predictive models.

The impact analyses calculated in the study illustrate how accurate the predictive models are in the context of 'does not meet/meets and exceeds' (termed pass/fail) performance levels. For educators, this context is important. In making decisions about course enrollment for ELL students, it is more important to be able to predict does not meet/meets and exceeds status with a high degree of accuracy than to be able to predict a specific ELA scale score with lesser degree of accuracy. The impact data were calculated four ways for each ELA course. Each of the two regression models (the Composite only and the four domain models) were used to predict two sets of scale scores for each student. Then, the pass/fail impact criteria, using two different cut points, were applied to these predicted scores. The first set of impact analyses categorized students with predicted scale scores of 400 and above as 'passing' and those students with scale scores lower than 400 as 'failing'. This 400 cut point is the actual passing cut score for the ELA EOCT. The second set of impact analyses categorized students with predicted scale scores of 391 (for 9th Lit) and 392 (for Am Lit) and above as 'passing' and those with predicted scale scores below 391 and 392 as 'failing'. These lower cut points represented the lower boundary of the conditional standard error of measure of the actual cut scores for the two ELA tests.

All four pass/fail impact analyses were compared to the actual pass/fail rates for the ELL students in the sample. For 9th Lit, the percentage of ELL students who were predicted to pass the test from the ACCESS regression models was most similar to actual pass percentages when students who were predicted to earn a scale score of 391 or above were considered 'passing.' For Am Lit, the percentage of ELL students who were predicted to pass the test from the ACCESS regression models was most similar to actual pass percentages when students who were predicted to earn a scale score of 400 or above were considered 'passing.'

The predicted scale score tables (Tables 8 and 16) showed ACCESS Composite PLs that were associated with EOCT scale scores. These tables can provide educators with an additional piece of information to help them in deciding if an ELL student has demonstrated enough English proficiency to be potentially successfully if enrolled in an English Literature and Composition course.

Recommendations

The recommendations provided here refer to placing ELL students in ELA courses required for graduation, as opposed to placement into alternative ELA courses which may fulfill elective credit. In light of the results from the correlation and impact analyses computed in this study, a two-look approach to predicting student success in an English Literature course from the ACCESS performance is recommended. In brief, the first look applies the regression equation using the ACCESS Composite proficiency level to gauge if an ELL student is ready to take an ELA course. If yes, the second look focuses on the student's four ACCESS language domain proficiency levels to

ensure that he or she can demonstrate the specific skills needed for course success (e.g., reading English text).

Because the ACCESS test is vertically scaled, the Composite Proficiency Levels and ELA scale scores listed in Tables 8 and 16 apply to those students taking a Tier B or C form only. Students taking the Tier A (beginner) test will not demonstrate the English language proficiency needed for success in an ELA course. Since the Tier A form applies a 4.0 cap at two of the four language domain proficiency levels, it is very unlikely that a student taking a Tier A form would receive a Composite PL that is high enough to be associated with a passing ELA scale score on the EOCT anyway.

Once a student has been categorized as ‘potentially’ successful with the Composite regression model, it is recommended that educators take a second look at each student’s four ACCESS domain proficiency levels before making the final decision to enroll her or him into an ELA course. In this study, the four domain regression analysis showed that a student’s Reading proficiency level has the greatest predictive capability of his or her EOCT ELA scale score. A student’s Speaking proficiency level has the least amount of predictive capability. It seems, that a good rule of thumb would be if a student’s Reading proficiency level is above or at least the same as his or her other three domain proficiency levels, then he or she may be a good candidate for ELA course enrollment. Conversely, if a student’s Reading proficiency level is lower than his or her other three domain proficiency levels, then perhaps he or she would not be a good candidate for enrolling in an ELA course.

As an additional piece of information, it is recommended that teachers and administrators calculate a second predicted ELA score for these students using the four domain regression equation. These equations can also be found in Tables 3 and 11 for 9th Lit and Am Lit, respectively. If the four domain equation yields a much lower predicted ELA scale score for a student than the Composite only equation predicted, then perhaps the student should wait to be enrolled in the ELA course until she or he demonstrates greater English language proficiency.

With this study, the Georgia Department of Education hopes to assist educators in making good decisions about English Literature course enrollment for their high school ELL students. With more data available in future years, the relationship between the ACCESS and EOCT can be refined. It should be noted, though, that regardless of amount of research conducted, there can be no empirical approach that is 100% accurate. ESOL teachers’ evaluations of a student’s English language proficiency should be an integral component in course enrollment decisions. Educators should always use their best judgment for making any decision regarding their students and keep in mind that all students are most likely to be successful in courses when they are academically prepared for the courses and receive academic support while in the courses.

References

Gottlieb, M. (2007). *Assessing comprehension and communication in English state to state for English language learners: Interpretive guide for score reports* (Spring, 2007). Madison, WI: University of Wisconsin, Wisconsin Center for Educational Research.

Pearson Educational Measurement. (2007, November). *The Georgia end-of-course tests technical manual*. (Issue Spring 2007). Iowa City, IA: Author.