



2010-11 Assessment of General Education Outcomes
Written Communication

Office of Institutional Assessment and Evaluation

July, 2011

“Working together to create a culture of evidence-based decision making...”



Office of Institutional Assessment and Evaluation
“Working together to create a culture of evidence-based decision making...”

2010-11 Assessment of General Education Outcomes
Written Communication

The Assessment of Written Communication

In Spring Semester 2011, the Office of Institutional Assessment and Evaluation collected data on student performance in English 003 from the English Department for the purpose of determining the extent to which students demonstrated competence in the general education outcome, *written communication*.

During the semester, students were required to compose essays in the areas of Style Analysis, Definition, Investigative, and Solution. A descriptive analytic rubric was used for grading each essay based on five criteria: (1) Organization, (2) Development A, (3) Development B, (4) Style, and (5) Grammar/Mechanics. A copy of the analytic rubric is in the Appendix.

For each criterion, a score is given in one of six levels, with maximum of 20 points. The top level is *Strong*, which has a score range of 18-20 points. The second level is *Satisfactory*, which has a score range of 16-17 points. The third level is *Passing/Acceptable*, which has a score range of 14-15 points. A score that is lower than 14 is considered a failing grade. Failing scores fall into three levels: *Failing, Inadequate* (12-13 pts), *Failing, Seriously Flawed* (1-11 pts), and *Total Failure* (0 pts). A student’s total score on an essay is the sum of all criterion scores. The total score on an essay can range from 0 to 100.

Description of Criteria on the Analytic Rubric

The descriptions of each of the five criteria are as follows:

- **Organization:** Organizes ideas and connects them with clear transitions.
- **Development A:** Identifies and analyzes important features of the argument and supports the main points of the argument.
- **Development B:** Explains opposition and refutes it; uses documentation.
- **Style:** Control of language, including diction and syntactic variety.
- **Grammar/Mechanics:** Facility with conventions (grammar, usage, and mechanics).

For each criterion, a score is earned on one of six levels. The maximum number of points for each criterion is 20. The first/highest level is *Strong*, which has a score range of 18-20 points. The second level is *Satisfactory*, which has a score range of 16-17 points. The third level is *Passing/Acceptable*, which has a score range of 14-15 points. A score that is lower than 14 is considered a failing score. Failing scores fall into three levels: *Failing, Inadequate* (12-13 pts), *Failing, Seriously Flawed* (1-11 pts), and *Total Failure* (0 pts). A student’s total score on an essay is the sum of all criterion scores. The total score on an essay can range from 0 to 100.

Studies on the Development of Students' Writing Skills in English 003

Scoring rubrics have become widely used in educational settings as a tool for assessing student writing skills (Andrade, 2000). When descriptive analytic rubrics are used, students receive a score on their performance on each writing criterion as well as a description of what that score means. The results can be used to make improvement in areas of weakness. Rubrics can also be used for purposes of evaluation and accountability. The analyses of students' performances across multiple sections of a course can inform the attainment of institutional level goals and objectives.

For Spring Semester 2011, two studies were conducted on the English 003 essay data. Study 1 focused on the effects of writing revisions of essays on the improvement of students' writing skills. Study 2 focused on the effects of writing multiple essays, without revisions, on the improvement of students' writing skills.

Study 1

The literature on formative assessment provides guidance on the characteristics of effective assessment and feedback (Sadler, 1989; Hattie and Timperley, 2007; Hughes, 2010). Two important components of formative assessment and feedback include: (1) articulated *clear criteria* for assessing writing, and (2) provide *opportunities for improvement through revision* (Andrade, 2000).

Procedure and Results: During the course of the semester, students in one section of English 003 were assigned only two essays. Students submitted a first version and then a revision of each essay after the first version was scored using a descriptive analytic rubric. Table 1 presents the means and standard deviations of criterion scores for the two essays before and after the revision. The valid "N" consists of the number of observations; that is, the scores for all participating students who submitted two essays and two revisions. The average total score on the two essays increased from $M = 68.95$, $SD = 8.18$ to $M = 77.82$ to $SD = 7.278$. The results also show that the development of students' writing skills improved for each criterion, the minimum competency score of at least 14 or 70% was achieved, and variability of students' scores decreased on every criterion. Paired correlations between before and after sample means ranged from .31 for Development A to .57 for Style.

Table 1. Descriptive Statistics Before and After Revision by Grading Criteria

Criterion	Valid N	Mean		Std. Dev.		Paired Samples Correlation
		Before Revision	After Revision	Before Revision	After Revision	
Organization	55	14.24	16.36	1.795	1.747	0.36*
Development A	55	14.20	15.91	1.726	1.724	0.31*
Development B	55	12.36	14.22	2.857	2.192	0.38*
Style	55	14.03	15.84	2.322	1.782	0.57**
Grammar/Mechanics	55	14.12	15.49	2.323	1.783	0.54**
Total Score	55	68.95	77.82	8.182	7.278	0.52**

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 2 presents the results of paired t-tests that were performed on mean scores for each criterion. All criterion scores after the revisions were significantly larger than the criterion scores before the revisions at $p < .001$. The largest gain in mean scores was for the criterion Organization, $Mean\ Difference = 2.13, SE = .27, p < .001$. The smallest gain in mean scores was for Grammar/Mechanics, $Mean\ Difference = 1.37, SE = .27, p < .001$.

Table 2. Paired *t*-test Before and After Revision by Grading Criteria

Criterion	Mean Difference	Std. Error	95% Confidence Interval		<i>t</i>	Sig. (2-tailed)
	(After - Before Revision)		Lower	Upper		
Organization	2.13	0.27	1.59	2.67	7.89	<0.001
Development A	1.71	0.27	1.16	2.26	6.23	<0.001
Development B	1.86	0.39	1.08	2.63	4.79	<0.001
Style	1.81	0.27	1.28	2.34	6.82	<0.001
Grammar/Mechanics	1.37	0.27	0.82	1.92	5.01	<0.001
Total Score	8.87	1.02	6.82	10.92	8.68	<0.001

Study 2

Table 3 presents the descriptive statistics by each criterion for Study 2 in which students wrote 4-5 essays without submitting revisions. There were a total of 2135 valid observations. Minimum and maximum scores for each criterion are 0 and 20, respectively. A mean score for each criterion was calculated across all valid observations. *Development B* has the lowest mean score, $M = 14.49, SD = 4.33$ and *Organization* has the highest mean score, $M = 15.9, SD = 3.79$.

Table 3. Descriptive Statistics for English 003 by Grading Criteria - Spring Semester 2011

Criterion	Valid N	Minimum	Maximum	Mean	Std. Dev.
Organization	2135	0	20	15.90	3.79
Development A	2134	0	20	15.39	3.71
Development B	2135	0	20	14.49	4.33
Style	2135	0	20	15.53	3.50
Grammar/Mechanics	2135	0	20	15.22	3.48
Total Score	2138	0	100	76.52	17.21

The bivariate correlation coefficients are shown in Table 4. All correlation coefficients are positive and significant at the 0.01 level. The correlation between *Grammar/Mechanics* and *Development B* is the smallest with a value of .625. The correlation between *Development A* and *Total Score* is the largest with a value of 0.938.

Table 4. Pearson Correlations for Five Criteria and Total Score

	Organization	Development A	Development B	Style	Grammar/Mechanics	Total Score
Organization	1	0.889**	0.774**	0.872**	0.772**	0.937**
Development A	0.889**	1	0.822**	0.846**	0.758**	0.938**
Development B	0.774**	0.822**	1	0.708**	0.627**	0.862**
Style	0.872**	0.846**	0.708**	1	0.899**	0.934**
Grammar/Mechanics	0.772**	0.758**	0.627**	0.899**	1	0.866**
Total Score	0.937**	0.938**	0.862**	0.934**	0.866**	1

** Correlation is significant at the 0.01 level (2-tailed).

F-test and Post Hoc Analysis

An Analysis of Variance (ANOVA) was performed on the five criterion means to determine if students performed equally across all criteria. The *F*-test shows that the means for five criteria are *not* equal and the overall difference is significant, $F(4,10669)=40.51, p<.001$ The results of *F*-test are shown in Table 5.

Table 5. ANOVA for Five Criteria Used in the Grading Rubric

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2308.158	4	577.039	40.51	<.001
Within Groups	151972.706	10669	14.244		
Total	154280.863	10673			

In order to compare the mean difference between criteria, a Post Hoc analysis was performed. From Table 3, we know that the variances are different. Therefore, equal variances are not assumed. Tamhane's T2 Post Hoc test was used. The results are shown in Table 6.

In Table 6, the mean differences are shown in the third column. The mean for *Organization* is significantly larger than any other criterion. The mean for *Development B* is significantly smaller than any other criterion. The ranking for the mean is *Organization > Style > Development A > Grammar/Mechanics > Development B*. The mean of *Style* is slightly larger than the mean of *Development A*, but the difference is not significant. The largest mean difference is between *Organization* and *Development B*, and the difference is 1.408.

Table 6. Post Hoc Multiple Comparisons

(I) Criterion	(J) Criterion	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Organization	Development A	0.513*	0.115	<0.001	0.19	0.83
	Development B	1.408*	0.125	<0.001	1.06	1.76
	Style	0.369*	0.112	0.010	0.06	0.68
	Grammar/Mechanics	0.686*	0.111	<0.001	0.37	1.00
Development A	Organization	-0.513*	0.115	<0.001	-0.83	-0.19
	Development B	0.895*	0.123	<0.001	0.55	1.24
	Style	-0.145	0.110	0.879	-0.45	0.16
	Grammar/Mechanics	0.173	0.110	0.712	-0.14	0.48
Development B	Organization	-1.408*	0.125	<0.001	-1.76	-1.06
	Development A	-0.895*	0.123	<0.001	-1.24	-0.55
	Style	-1.039*	0.120	<0.001	-1.38	-0.70
	Grammar/Mechanics	-0.722*	0.120	<0.001	-1.06	-0.39
Style	Organization	-0.369*	0.112	0.010	-0.68	-0.06
	Development A	0.145	0.110	0.879	-0.16	0.45
	Development B	1.039*	0.120	<0.001	0.70	1.38
	Grammar/Mechanics	0.317*	0.107	0.030	0.02	0.62
Grammar/Mechanics	Organization	-0.686*	0.111	<0.001	-1.00	-0.37
	Development A	-0.173	0.110	0.712	-0.48	0.14
	Development B	0.722*	0.120	<0.001	0.39	1.06
	Style	-0.317*	0.107	0.030	-0.62	-0.02

*. The mean difference is significant at the 0.05 level.

Note: "Tamhane's T2" Post Hoc test is used, and equal variances are not assumed.

Passing Rate by Criterion

The passing competency requires a score of 70% or higher, i.e., 14 out of total 20 points. The results are shown in Table 7. The criterion with the highest passing rate is the *Style*, which has a passing rate of 89.8%. *Development B* has the lowest passing rate at 70.5%. The ranking for passing rate is *Style* > *Organization* > *Development A* > *Grammar/Mechanics* > *Development B*.

Table 7. Passing Rate by Criterion

Criteria	Status	Frequency	Percent
Organization	Fail	291	13.6%
	Pass	1844	86.4%
	Total	2135	100.0%
Development A	Fail	357	16.7%
	Pass	1777	83.3%
	Total	2134	100.0%
Development B	Fail	629	29.5%
	Pass	1506	70.5%
	Total	2135	100.0%
Style	Fail	218	10.2%
	Pass	1917	89.8%
	Total	2135	100.0%
Grammar/Mechanics	Fail	342	16.0%
	Pass	1793	84.0%
	Total	2135	100.0%
Total Score	Fail	366	17.1%
	Pass	1772	82.9%
	Total	2138	100.0%

Note: Based on descriptive analytic rubric, the passing grade requires a score of 70% of the total score, i.e., 14 out of total 20 points.

Summary

The ultimate goal for outcomes assessment is to use the results of assessment to improve student learning and performance. This report embraces this goal by emphasizing student performance at the criterion level. The disaggregation of scores at the criterion level vis-à-vis total scores at the essay level provides an advantage in implementing strategies for improvement. First, it informs the instructors in the Department of English about the strengths and weaknesses of students enrolled in English 003 in the aggregate. Second, student performance data at the criterion level provide instructors with specific targets for improvement in the teaching and learning process. Third, analysis at the criterion level provides information about students' relative strengths and weaknesses among different types of essay composition (e.g., Style Analysis,

Definition, Investigative and Solution). This information can help instructors balance time and effort devoted to helping students develop competency in each criteria and for each essay type.

For Spring Semester 2011, students enrolled in English 003 met the minimum performance criterion of 70% in *all* criterion areas. For four of the five criterion areas, *Organization, Development A, Style, and Grammar/Mechanics*, students' performance exceeded the minimum performance criterion by more than 13% to 19%. Student performance on Development B exceeded the minimum performance criterion by only .5% (rounded to 1%).

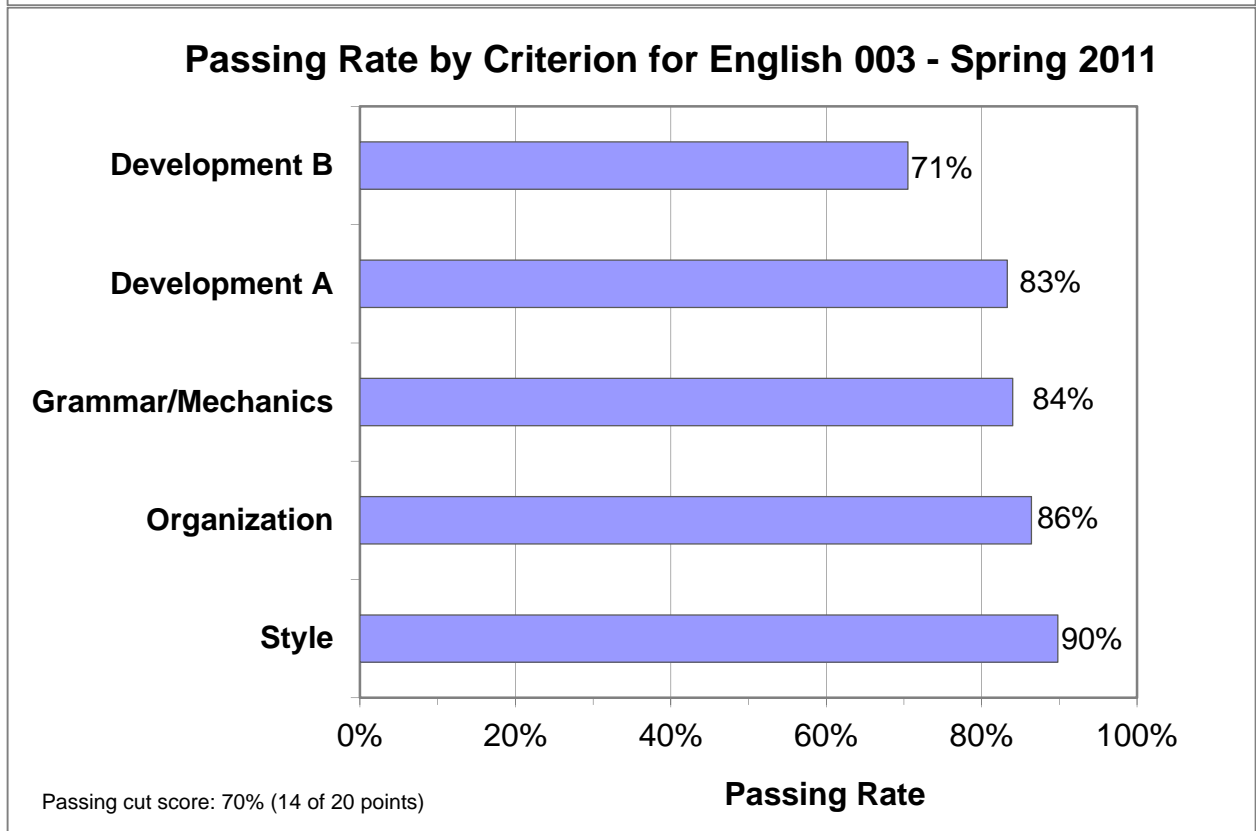
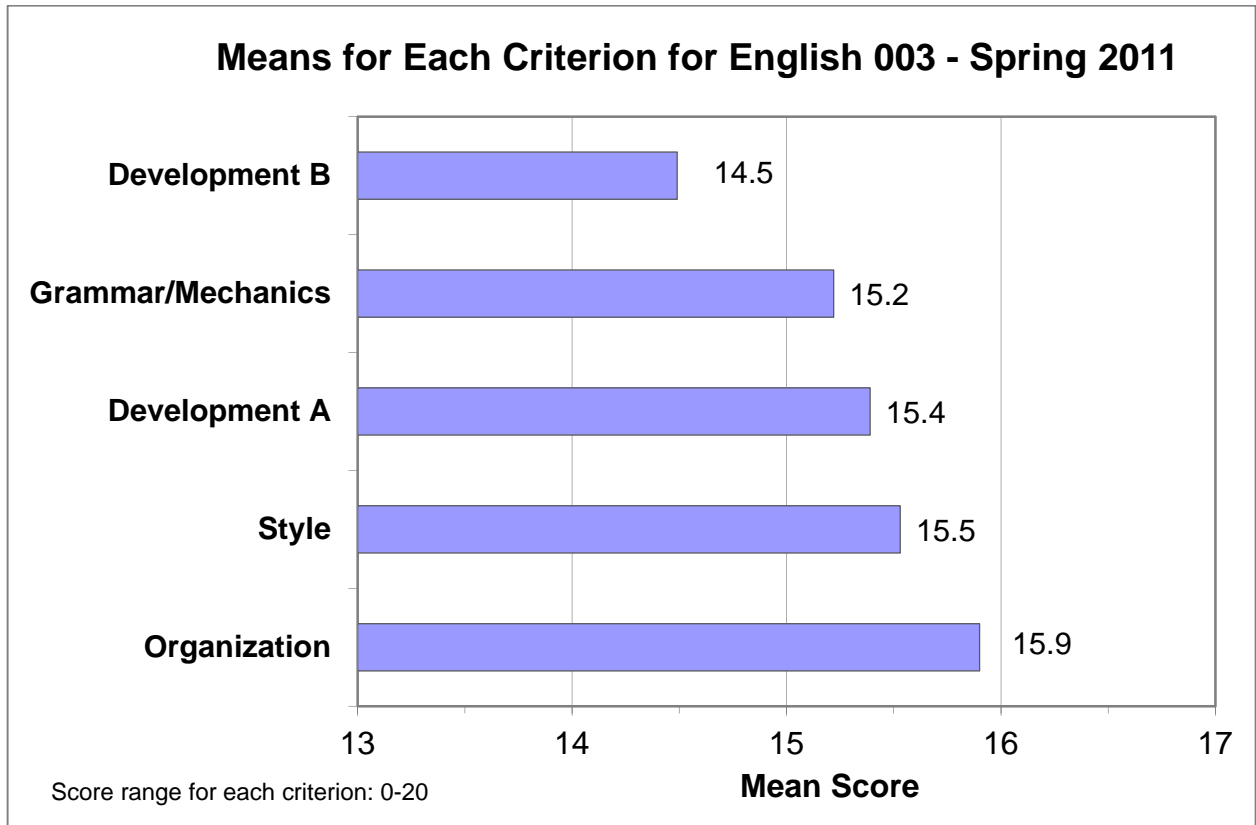
It is also instructive to note that when students were allowed to revise their essays and were given feedback on their performance, their scores increased significantly and the variability in their performance decreased. Additionally, the mean criterion scores for the students in the “revision” class exceeded the mean criterion scores of students who wrote 4-5 essays in every case, except *Development B*.

Similar to the results from Spring Semester 2010, *Development B* and *Grammar/Mechanics* have the lowest mean scores (See Tables 1 and 3) and passing rates (Table 7). Enhanced training in these two areas can effectively improve students' writing competencies in English 003. The results for Spring Semester 2010 exceeded the results for Spring Semester 2011 for every criterion except *Grammar/Mechanics*. The Office of Institutional Assessment and Evaluation (OIAE) will continue to monitor student learning outcomes in written communications to identify discernable trends.

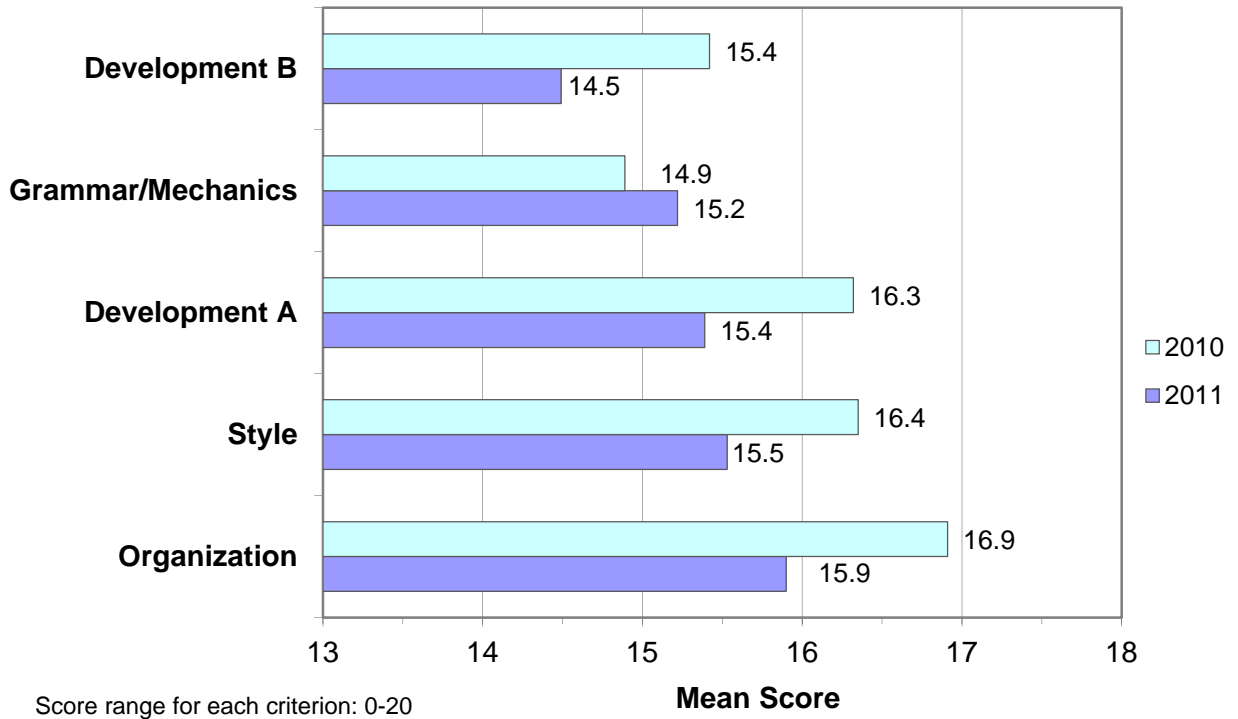
References

- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Hattie, J. A., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hughes, G. (2010). Formative assessment practices that maximize learning for students at risk. In. H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 212-232). New York: Routledge.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.

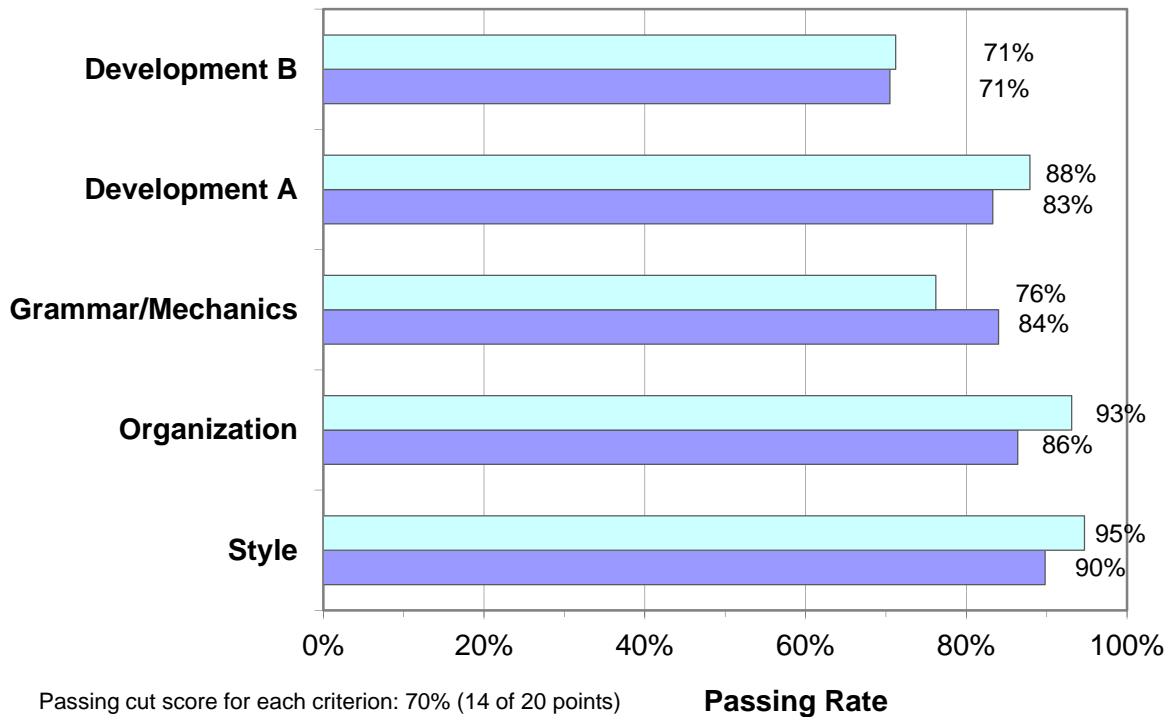
Graph Presentations



**Comparison of Criterion Means for English 003
Between Spring 2010 and Spring 2011**



**Comparison of Passing Rate for English 003
Between Spring 2010 and Spring 2011**



Appendix

ANALYTIC RUBRIC FOR THE ARGUMENTATIVE ESSAY (ENGLISH 003)
LETTER GRADES COORDINATE WITH THIS SCHEME AS FOLLOWS: 100-90=A; 89-80 = B; 79-70 = C; 69-60 = D; 59-BELOW F

VALUE → CRITERIA ↓	Strong 18-20 pts. Total 100-90 = A	Satisfactory 16-17 pts. 89-80 = B	Passing / Acceptable 14-15 pts. 79-70 = C	Failing, Inadequate 12-13 pts. 69-60 = D	Failing, Seriously Flawed 1-11 pts. 59 and below	Total Failure 0 pts.	SCORE
OVERALL & WITHIN PARAGRAPHS: Weight = 20 pts	Consistently organizes ideas logically and effectively and connects them w/ clear transitions: 18-20 pts.	Usually organizes ideas logically and connects them with appropriate transitions: 16-17 pts.	Organizes ideas satisfactorily but may not always connect them with transitions: 14-15 pts.	Has limited organization of ideas: 12-13 pts.	Is very disorganized; weak to no evidence of organization: 1-11 pts.	Responds off topic; evidences plagiarism: 0 pts.	
DEVELOPMENT: A) Analysis of Argument and Supporting Evidence; Incl. Research Documentation as Relevant Weight = 20 pts	Clearly identifies and insightfully analyzes important features of the argument and effectively supports the main points of the argument: 18-20 pts.	Clearly identifies important features of the argument and analyzes them in a generally thoughtful way and sensibly supports the main points of the argument: 16-17 pts.	Identifies and capably analyzes important features of the argument and supports the main points of the argument: 14-15 pts.	Does not identify or analyze most of the important features of the argument and offers support of little relevance or value for main points of the argument: 12-13 pts.	Does not understand, identify or analyze main features of the argument and/or provides little, erroneous, or illogical support for the main points of the argument: 1-11 pts.	Provides no evidence of ability to identify or analyze features of argument and/or provides no support; off topic; evidences plagiarism: 0 pts.	
DEVELOPMENT: B) Opposition and Refutation; Incl. Research Documentation as Relevant (as needed) Weight = 20 pts.	Clearly and fully explains opposition and completely and logically refutes it; consistently uses documentation effectively: 18-20 pts.	Fully explains opposition and logically and persuasively refutes it; usually uses documentation effectively: 16-17 pts.	Adequately explains opposition and gives logical refutation; adequately uses documentation: 14-15 pts.	Does not adequately explain opposition, only superficially explains. Provides some refutation; uses documentation somewhat ineffectively: 12-13 pts.	Barely explains or does not explain opposition and/or provides weak or no refutation; uses documentation ineffectively: 1-11 pts.	Provides no opposition or refutation; responds off topic; does not use required documentation; evidences plagiarism: 0 pts.	
STYLE: Language Control (Syntax & Diction) Weight = 20 pts.	Demonstrates superior control of language, including diction and syntactic variety: 18-20 pts.	Demonstrates clear control of language, including diction and syntactic variety: 16-17 pts.	Demonstrates adequate control of language, including diction and syntax but may lack syntactic variety: 14-15 pts.	Uses language imprecisely and/or lacks sentence variety: 12-13 pts.*	Has serious or severe and frequent problems in the use of language and sentence structure: 1-11 pts.	Has no sentence structure; incorrect diction a problem; evidences plagiarism: 0 pts.	
GRAMMAR AND MECHANICS (Punctuation): Weight = 20 pts.	Demonstrates superior facility with conventions (grammar, usage, and mechanics): 18-20 pts.	Demonstrates facility with the conventions of standard English but may have minor flaws: 16-17 pts.	Displays control of the conventions of standard written English but may have some flaws: 14-15 pts.	Has occasional major errors or frequent minor errors in grammar, usage and mechanics: 12-13 pts.	Has numerous errors in grammar, usage and mechanics that interfere with meaning or result in incoherence: 1-11 pts.	Has illegible and gross errors in grammar, usage and mechanics; evidences plagiarism: 0 pts.	

Student Name/ID: _____

Total Score: _____

Rater: _____