



2013-14 Assessment of General Education Outcomes
Written Communication

Office of Institutional Assessment and Evaluation

July, 2014

“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...”

2013-14 Assessment of General Education Outcomes *Written Communication*

The Assessment of Written Communication

In Spring Semester 2014, 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.

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 students' 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 departmental and institutional level goals and objectives.

For Spring Semester 2014, 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) articulate *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 to revise one of the essays. Students submitted a first version and then a revision of the same 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 number of criterion scores for all participating students who submitted the essay and the revision. The average total score on the essay increased from $M = 73.73$, $SD = 11.82$ to $M = 84.26$, $SD = 9.95$. The results also show that the development of students' writing skills improved for each criterion, the variability (e.g. the standard deviation) of students' scores decreased for each criterion, and that the minimum competency score of at least 14 or 70% was achieved. Paired correlations between before and after revision ranged from .44 for *Grammar/Mechanics* to .94 for *Style*.

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

Criterion	Valid N	Mean		Standard Deviation		Paired Samples Correlation
		Before Revision	After Revision	Before Revision	After Revision	
Organization	33	15.40	17.46	3.60	2.20	0.72**
Development A	33	14.80	17.60	3.12	2.44	0.76**
Development B	33	14.93	16.93	2.58	2.37	0.69*
Style	33	14.33	16.26	1.63	1.67	0.94**
Grammar/Mechanics	33	14.40	15.86	2.23	2.23	0.44**
Total Score	33	73.73	84.26	11.82	9.95	0.88**

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

Figure 1 shows that the mean score on each criterion increased between before and after revision. The largest increase in before and after revision scores was for *Development A*, Mean Difference = 2.80, and the smallest increase was for *Grammar/Mechanics*, Mean Difference = 1.46.

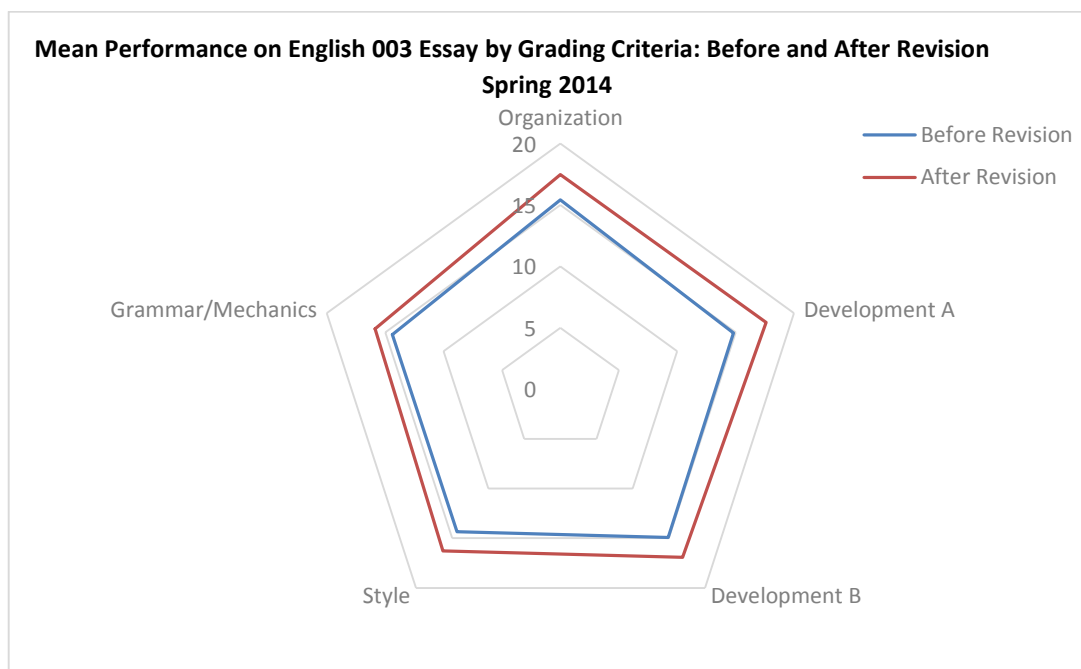


Figure 1

Table 2 presents the results of paired-sample *t*-tests that were conducted to compare mean criterion scores on essays before and after revisions. Specifically, there was a significant difference in the mean criterion scores for *Development A*, (Mean Difference=2.80; SE=.52); $t(14)=5.31$, $p<.001$). Similarly, there was a significant difference in the mean criterion scores for *Grammar/Mechanics*, (Mean Difference=1.46; SE=.60); $t(14)=2.41$; $p<.05$).

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	2.06	0.65	0.67	3.46	3.17	<0.001
Development A	2.80	0.52	1.66	3.93	5.31	<0.001
Development B	2.00	0.50	0.91	3.08	3.94	<0.01
Style	1.93	0.15	1.60	2.26	12.61	<0.001
Grammar/Mechanics	1.46	0.60	0.16	2.77	2.41	<0.05
Total Score	10.53	1.46	7.39	13.67	7.20	<0.001

Study 2

Table 3 presents the descriptive statistics by each criterion for Study 2 in which students wrote essays without submitting revisions. There were a total of 1705 valid observations. The scores of the original essay (e.g. “before revision”) used in Study 1 are also included in Study 2, but scores of the revised essays from Study 1 are excluded. A mean score for each criterion was calculated across all valid observations. *Organization* has the highest mean score, $M=16.97$, $SD=2.97$ and *Grammar/Mechanics* has the lowest mean score, $M=15.94$, $SD=2.80$.

Table 3. Descriptive Statistics by Each Criterion

Criterion	Valid N	Mean	Std. Dev.
Organization	1630	16.97	2.97
Development A	1630	16.66	4.00
Development B	1630	16.01	3.18
Style	1630	16.53	2.73
Grammar/Mechanics	1630	15.94	2.80
Total Score +	1705	80.60	17.38

The Valid N for Total Score exceeds the Valid N for each criterion score because some English professors reported the “Total Score” only and no criterion scores.

Figure 2 shows the mean score for each criterion on English 003 essays. The rankings for the mean are as follows: *Organization* > *Development A* > *Style* > *Development B* > *Grammar/Mechanics*.

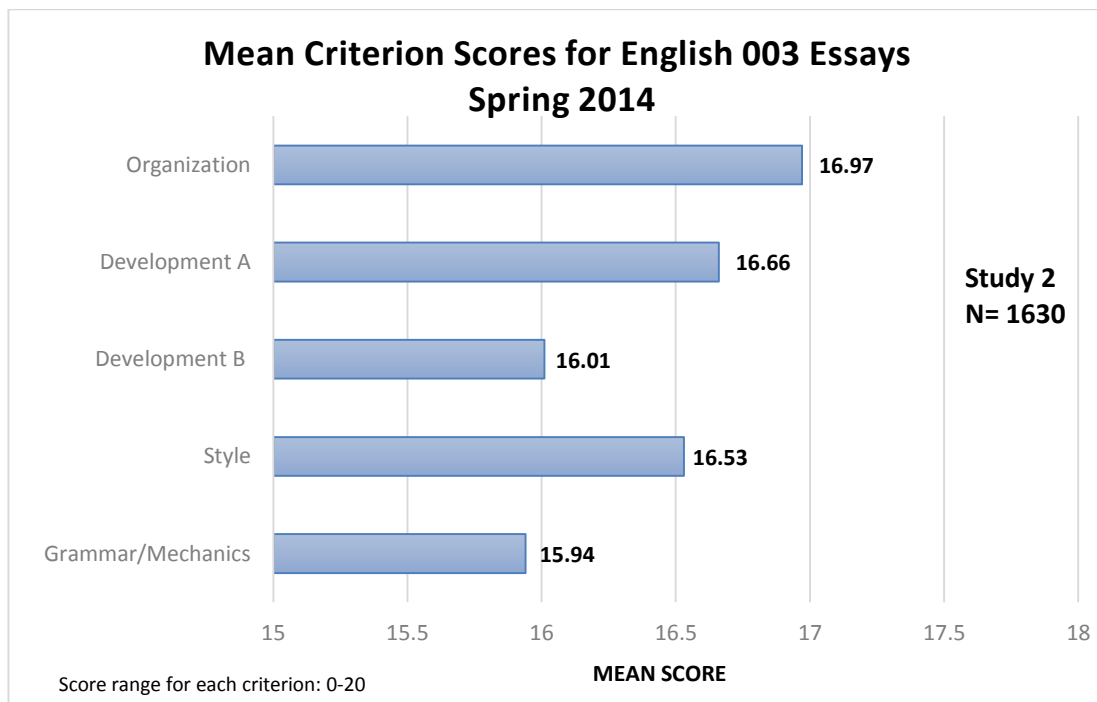


Figure 2

Figure 3 shows the comparison of mean criterion scores for English 003 essays from 2010 – 2014.

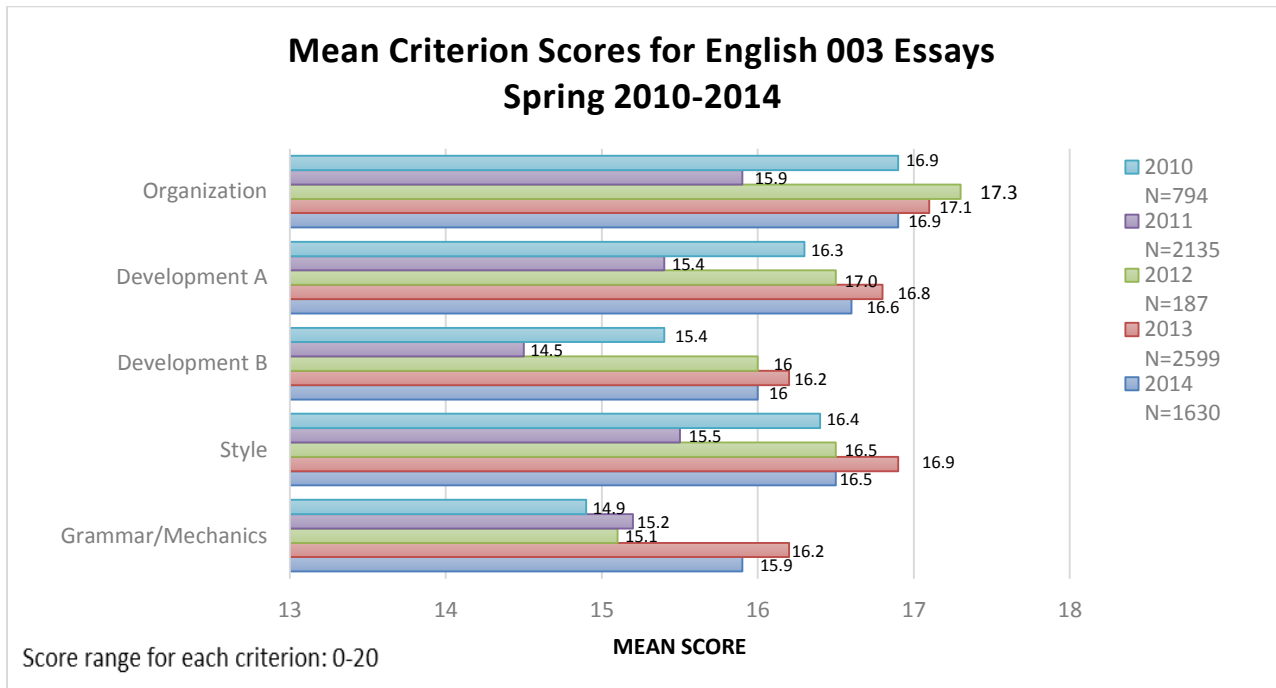


Figure 3

Table 4 presents the bivariate correlation coefficients between pairs of criterion scores and the Total Score. 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 .712. The correlation between *Development A* and *Total Score* is the largest with a value of 0.957.

Table 4. Pearson Correlations for Five Criteria and Total Score

	Organization	Development A	Development B	Style	Grammar/Mechanics	Total Score
Organization	1	.932**	.843**	.868**	.770**	.949**
Development A	.932**	1	.866**	.879**	.771**	.957**
Development B	.843**	.866**	1	.806**	.712**	.913**
Style	.868**	.879**	.806**	1	.881**	.949**
Grammar/Mechanics	.770**	.771**	.712**	.881**	1	.883**
Total Score	.949**	.957**	.913**	.949**	.883**	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 the five criteria are *not* equal and the overall difference is significant, $F(4,8145)=36.629$, $p<.001$. The results of the *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	1253.922	4	313.481	36.63	$p<.001$
Within Groups	69707.341	8145	8.558		
Total	70961.263	8149			

In order to compare the mean differences between criteria, a Post Hoc analysis was performed. The test of Homogeneity of Variances revealed a significant Levene Statistic, which indicates unequal variances. 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 differences between *Development A* and *Style* (MD = .131, $p=.871$) and between *Grammar/Mechanics* and *Development B* (MD= -.075, $p=.998$) are not significant. The largest mean difference is between *Organization* and *Grammar/Mechanics* (MD=1.031, $p<.001$)

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	.309*	.103	.027	.02	.60
	Development B	.956*	.108	<.001	.65	1.26
	Style	.440*	.100	<.001	.16	.72
	Grammar/Mechanics	1.031*	.101	<.001	.75	1.31
Development A	Organization	-.309*	.103	.027	-.60	-.02
	Development B	.647*	.107	<.001	.35	.95
	Style	.131	.099	.871	-.15	.41
	Grammar/Mechanics	.722*	.100	<.001	.44	1.00
Development B	Organization	-.956*	.108	<.001	-1.26	-.65
	Development A	-.647*	.107	<.001	-.95	-.35
	Style	-.516*	.104	<.001	-.81	-.23
	Grammar/Mechanics	.075	.105	.998	-.22	.37
Style	Organization	-.440*	.100	<.001	-.72	-.16
	Development A	-.131	.099	.871	-.31	.15
	Development B	.516*	.104	<.001	.23	.81
	Grammar/Mechanics	.591*	.097	<.001	.32	.86
Grammar/Mechanics	Organization	-1.031*	.101	<.001	-1.31	-.75
	Development A	-.722*	.100	<.001	-1.00	-.44
	Development B	-.075	.105	.998	-.37	.22
	Style	-.591*	.097	<.001	-.86	-.32

* 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 and Total Score

The passing competency requires a score of 70% or higher, (i.e., 14 out of total 20 points for criterion scores and 70 out of 100 for Total Scores). The results are shown in Table 7. The criterion with the highest passing rate is the *Style*, which has a passing rate of 96%. *Development B* has the lowest passing rate at 89.0%. The ranking for passing rates is *Style* > *Organization* > *Development A* > *Grammar/Mechanics* > *Development B*. Furthermore, 93% of all essays included in this analysis earned a Total Score that met the passing competency score of 70%.

Table 7. Passing Rate by Criterion

Criteria	Status	Frequency	Percent
Organization	Fail	83	5.1%
	Pass	1547	95%
	Total	1630	100%
Development A	Fail	93	5.7%
	Pass	1537	94%
	Total	1630	100%
Development B	Fail	179	11.0%
	Pass	1451	89%
	Total	1630	100%
Style	Fail	69	4.2%
	Pass	1561	96%
	Total	1630	100%
Grammar/Mechanics	Fail	144	8.8%
	Pass	1486	91%
	Total	1630	100%
Total Score	Fail	122	7.2%
	Pass	1583	93%
	Total	1705	100%

Note: Based on descriptive analytic rubric, the passing grade requires a score of 70% (i.e., 14 out of total 20 points for criterion score and 70 out of 100 for the Total Score)..

Figure 4 shows the passing rate by criterion for English 003 essays.

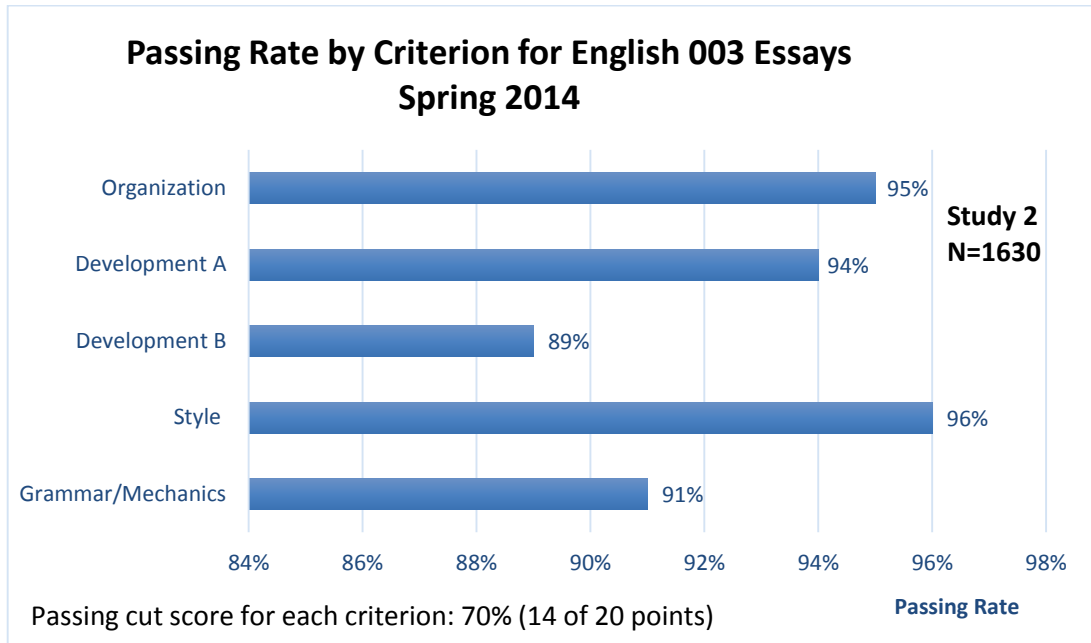


Figure 4

Figure 5 shows the comparison of passing rates for English 003 essays from 2010 – 2014.

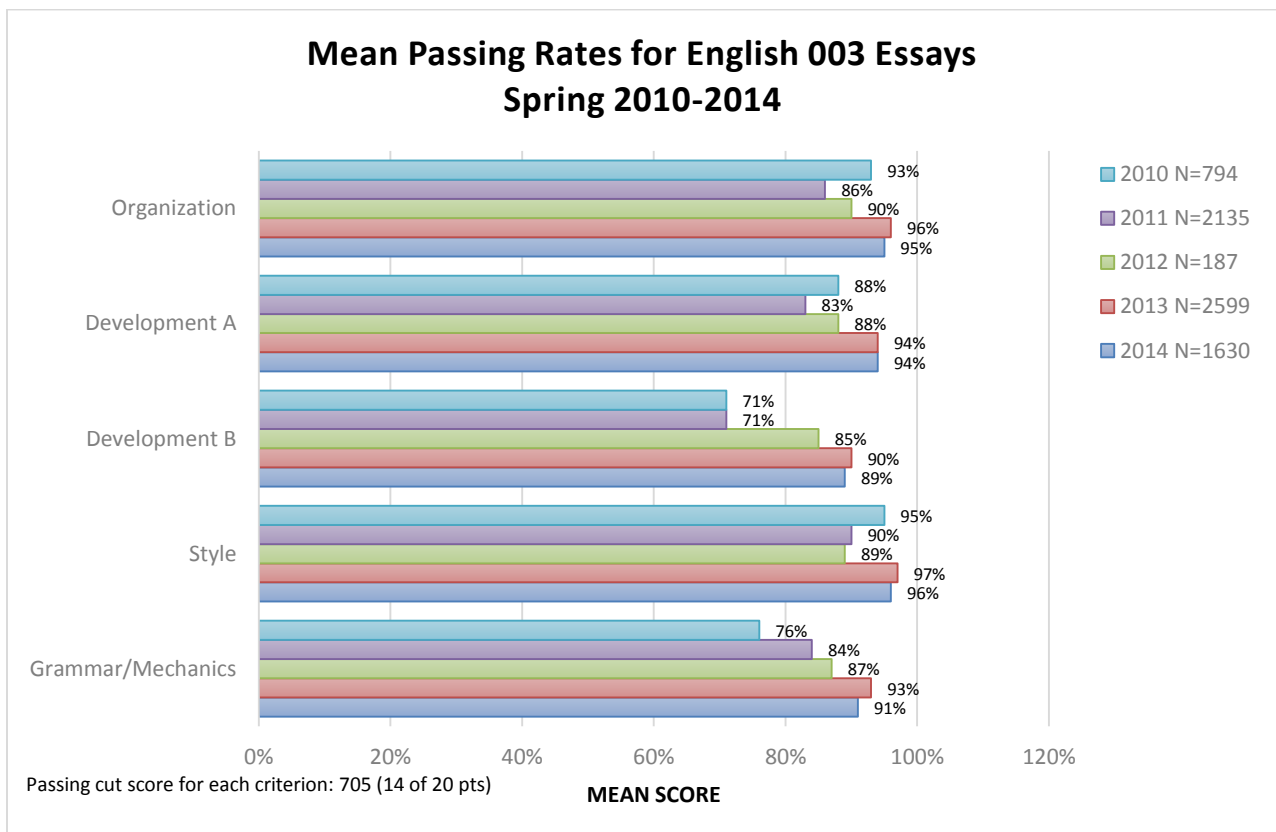


Figure 5

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 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 essays (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 2014, students enrolled in English 003 met the minimum performance criterion of 70% in *all* criterion areas. For four criterion areas, *Organization*, *Development A*, *Style* and *Grammar/Mechanics*, passing rates exceeds 90%; only *Development B* has a passing rate of 89%.

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.

Similar to the results from Spring Semester 2011, 2012, and 2013, *Development B* and *Grammar/Mechanics* have the lowest mean scores (See Table 3) and passing rates (Table 7). Enhanced training in these two areas can effectively improve students' writing competencies in English 003. Furthermore, the results for Spring Semester 2014 were less than or equal to the results for Spring Semester 2013 for each criterion. The Office of Institutional Assessment and Evaluation (OIAE) will continue to monitor student learning outcomes in written communications to identify discernible 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.

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 & WRITTEN 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: _____