

COMPARISON OF EFFECTS OF COGNITIVE LEVEL AND QUALITY WRITING ASSESSMENT (CLAQWA) RUBRIC ON FRESHMAN COLLEGE STUDENT WRITING

I. SUZANNE PENNER

Youngstown State University

The study investigated the effects of the Cognitive Level and Quality Writing Assessment (CLAQWA) rubric on the cognitive and writing skill growth in freshmen composition classes. The participants were enrolled at a Midwestern state university. The nonequivalent control group design used quantitative analysis with selected criteria from the CLAQWA rubric as measurements. Two independent raters graded the essays, and results confirmed a statistically significant correlation of grades on both sets of essays. Results from both raters confirmed no statistically significant differences on either type of skill score between the experimental or control group for the final essay. These results suggest that although a specific rubric enhances the learning environment, a specific rubric does not define the learning environment. Results indicated that the measurement of student outcomes, mandated by recent legislative efforts, may be accomplished through the use of a rubric, but at the same time, a specific rubric may not be a universal answer.

Keywords: cognitive, cognitive level quality writing assessment, writing skills, freshmen composition

The study investigated the effects of the Cognitive Level and Quality Writing Assessment (CLAQWA) rubric (*CLAQWA online*, 2007) on the cognitive and writing skill growth in freshmen composition classes. The participants were enrolled at a Midwestern state university. The nonequivalent control group design used quantitative analysis with selected criteria from the CLAQWA rubric as measurements. Two independent raters graded the essays, and results confirmed a statistically significant correlation of grades on both sets of essays. Results from both raters confirmed no statistically significant differences on either type of skill score between the experimental or control group for the final essay. These results suggest that although a specific rubric enhances the learning environment, a specific rubric does not define the learning environment. Results indicated that the measurement of student outcomes,

mandated by recent legislative efforts, may be accomplished through the use of a rubric, but at the same time, a specific rubric may not be a universal answer.

Accrediting bodies and campuses have mandated assessments of student learning (*Action Plan*, 2006; *Boards*, 2006; *Executive*, 2007; *Four Pillars*, 2004; *U.S. Department*, 2007). As a result, colleges are concerned about measuring academic quality and outcomes of students. A general assumption has existed that there is a close link between writing skills and cognitive skills, so colleges have seen a need to accurately measure both skills at the same time in undergraduate student work. Because colleges have frequently required two semesters of writing as core requirements, freshmen writing courses have constituted a logical place to evaluate such change in student writing.

To begin, ideas previously investigated

should be highlighted. Specific, measurable data has been analyzed (Hillocks, 1986). The Hillocks text has marked an important beginning to the enormous task of developing a bibliography and synthesizing information learned from composition studies. It has also set in motion the establishment of experimental and quasi-experimental designs in the area of academic English, a practice rarely seen in previous decades. More generalized practices have been collated (Roen et al., 2002). This book has represented encouragement, interpretation, and new directions for many pedagogical endeavors within the freshman college composition classroom.

Additionally, thought processes of college students and practical suggestions for undergraduates have been studied (Chickering & Gamson, 1987; Perry, 1999). Perry has used psychological theories to explain the behavior, motivation, and thinking of college students before he has indicated that professors should apply these aspects of intellectual development to individual teaching situations. Chickering and Gamson have succinctly summarized seven principles that can be applied in student-faculty settings, including student/faculty contact, student reciprocity, active learning techniques, prompt feedback, focused time on task, high student expectations, and mutual respect (1987).

With the perspectives of accreditation, teaching, and learning in mind, Ramey et al. (2007) have articulated the idea that a systematic way of improving college student writing will best serve campus, faculty, and student needs. Accordingly, when professors introduce a rubric to students at the beginning of the term, when professors give frequent feedback of writing to students, when professors explain to students how a rubric functions during the writing course, and when professors achieve consensus on essay standards (p. 70), the overall experience of the writing course could help students improve their writing (p. 71).

History and Terms. The history of composition studies might be a logical place to begin discussing college writing. Recently, Duncan (2007) outlined the history of paragraph teaching/composition studies, summarizing the last two centuries. He asserted that although labels have changed, three predominant methods of teaching composition have continually prevailed, labeling them prescriptive, descriptive, and cognitive.

Prescriptive methods of teaching composition have depended on a concrete conception of the structure of writing. Topic sentences, paragraphs, and writing structure have played a major role in the prescriptive method. Descriptive methods of teaching composition have deemphasized the structure of writing. As long as the writing has fulfilled the function of the writer's intention, the writing has been considered adequate. Cognitive methods and process models of teaching composition have emphasized critical thinking skills, a psychological construct that is now beginning to be investigated more fully.

Cognition is now beginning to be considered as the more important aspect of writing. According to Duncan (2007), prescriptive and descriptive methods have tended to cycle up and down in popularity for the last two hundred years while cognitive methods have been a relative newcomer to the realm of composition theory. Even so, these three labels have remained helpful in sifting through material about composition studies.

Duncan would like to see all three methods coalesce and collaborate with other disciplines in a way that both unites old ways of thinking and enhances newer classroom practices (p. 487). Yet, he has correctly perceived the difficulty of teaching the more nebulous concepts that belong to the descriptive and cognitive aspects of writing; the popularity of prescriptive methodology has dominated due to its more tangible aspects. It may be that the educational paradigm shifts brought about by

the NCLB Act (*Four Pillars*, 2004) and the Commission on the Future of Higher Education (*Executive*, 2007) will force educators to accomplish this challenging and lofty goal in composition studies.

Much research in composition studies in the last several decades has forged a general perspective about the content of composition theory (Bloom, 1984; Chickering & Gamson, 1987; Duncan, 2007; Hillocks, 1986; Hillocks, 1995; Perry, 1999). The history of composition studies, research about composition studies, theories about college student learning, and applications of those theories within individual courses have covered the main areas of concern in composition studies. Most of them have indirectly inferred either the prescriptive, descriptive, or cognitive methods, and all of them have assumed that word skills measure cognitive levels. Clearly, a need still exists for a way to link many parameters in a more concrete and identifiable way

Cognition. The measurements of cognitive growth have stemmed from the six levels of the cognitive taxonomy, defined by Bloom (1984, pp. 201-7). He gave the most emphasis to the aspects of knowledge, the first level of cognition. The knowledge level included both the more concrete recital of facts as well as the more abstract recollection of universals, interrelations, or patterns (p. 62). Also, student comprehension of cognitive levels may help to pinpoint areas where change is most needed. Bloom's identification of cognitive levels has aided both students and professors.

To further explain the nature of cognition, Perry (1999) extended and developed the cognitive model for the undergraduate level of learning. He understood that cognitive development from concrete experience to abstract functioning repeated itself at older levels of development (p. 32). In other words, every time an adult begins to learn a new concept, the person (who may be mature in other ways) necessarily also falls back to a more concrete

level of understanding before the person can move forward to more abstract functioning at the new concept.

In addition to the changeable, cyclical way that humans learn new concepts, other factors (not necessarily developmental factors) have influenced the way campuses think about cognition. For example, Perry observed that the movement toward diversity on campuses accompanied a movement toward relativity in knowledge, yet he viewed cognition as a growth process, not just a change process (pp. 2-5). This neutralizing tendency toward knowledge at the collegiate level has tended to diminish or overlook the importance of cognitive growth patterns. In other words, philosophical constructs can overtake psychological constructs, but returning to basic psychological concepts of growth may be a first step in reordering such omissions.

As a result, Perry's efforts to pinpoint aspects of the cognitive process have helped college level educators rediscover the importance of the developmental steps in order to evaluate cognitive growth. His cognitive scheme has opened up "the possibility of assessing, in developmental terms, abstract structural aspects of knowing and valuing in intelligent late-adolescents" (p. 16). To enable this evaluative process, he outlined four stages of cognitive maturation in post-adolescents: dualism, multiplicity, relativism, and commitment.

Dualism meant that students came to college thinking in a black/white, frame of mind. Multiplicity meant that students came to realize that many answers exist to solve a problem. Relativism meant that students abandoned their faith in the perspective of dualism. Commitment meant that students accepted a particular position and learned how it affected their lives.

Due to the introductory nature of freshmen courses, one would expect to see some cognitive growth in students, but probably not all four levels of it. The application of Perry's

thinking about college student cognition has been an appropriate beginning point for many composition studies. These principles have now been disseminated for more than two decades, and many colleges still consider the advice current, perhaps because the authors “address the teacher’s *how*, not the subject-matter *what*, of good practice in undergraduate education” (p. 4). In other words, it is generally recognized that beginning this task of more effective learning has meant starting with generalizations about many specifics.

Whereas Perry (1999) described the steps of cognition, Chickering and Gamson (1987) gave summative recommendations (student/faculty contact, student reciprocity, active learning techniques, prompt feedback, focused time on task, high student expectations, and mutual respect) that support the efforts of campuses to move in the direction of teaching more cognitively. Colleges have still considered the advice of Chickering and Gamson to be sound today, even though some of the terminology has changed. To restate their advice for a college writing course, one might say that their advice has leaned toward a balance of prescriptive and descriptive methods, with a slight emphasis on cognition.

The thinking of both Hillocks and Duncan—one who has studied writing treatments and one who has assessed historical trends—have seemed to say that when professors gave a framework to the writing assignment, the quality of student writing may have shown improvement. Professor-led involvement—neither domination nor absence—in the learning process is critical. Their thinking has reflected age-related cognitive theory about college students explained by Perry—that appropriate classroom management can foster cognitive growth revealed through student writing.

Without doubt, cognitive theory has been clearly related to the teaching of freshman writing because writing is the primary means

of “assessing, encouraging, and grading student thought” (Flateby & Metzger, n.d.). The fluid, flexible natures of cognition and maturity have enhanced the potential of initiating intellectual growth in the college situation. Also, because writing courses have frequently been required at the beginning of the undergraduate experience, it has seemed logical to assume that the most accurate measurements of cognitive growth could be obtained at that juncture in student writing. (Flateby, 2007; Flateby & Eubanks, 2008) This research project recognizes that the Cognitive Level and Quality Writing Assessment (CLAQWA) (Flateby & Metzger, 2001) is a rubric that was designed for use in the above-mentioned way, and this research represents an experimental design consistent with the research suggestions.

When taking into account several factors: 1) legislative and accreditation directives (*Executive*, 2007; *Four Pillars*, 2004); 2) the history of composition studies (Duncan, 2007); 3) theories of cognition and their application the college level (Bloom, 1984; Perry, 1999); 4) summaries of writing practices (Hillocks, 1986); and 5) summations of advice for undergraduates (Chickering & Gamson, 1987), one begins to see a need for a method, program, or tool that may incorporate all aspects of a complex situation into a cohesive unit and allow valid measurements of its individual components.

With new demands and requirements in mind, Flateby & Metzger (2001) from the University of South Florida developed the Cognitive Level and Quality Writing Assessment (CLAQWA) System, an assessment tool that measured both “writing skills and cognitive level” (p. 4). The CLAQWA rubric was developed at the University of South Florida as an instrument whose intention was “to help instructors standardize their evaluation of writing and assess the cognitive level attained in student writing” (Flateby & Metzger, n.d., p. 2).

The rubric has been in development since 1999 (Flateby & Metzger, 2001, p. 4), and it helps both students and professors clarify writing objectives. The instrument was developed by an interdisciplinary team and pilot testing (*Peer Review*, 2006). The CLAQWA has undergone refinement and revision as a result of surveying faculty and students (Flateby, 2007). Because this rubric defines and standardizes the skills involved in the writing processes and outcomes, it allows a valid comparison between two sets of final papers.

It is an instrument that has fulfilled a genuine need. The CLAQWA has defined writing skills and cognitive levels, giving professors a rubric for evaluation (*Primary*, 2007). To use the CLAQWA correctly, professors have presented an explanation of cognitive levels, assignment expectations, and requisite skills to students before an assignment is completed. When students have gained an understanding of the level of work required for the assignment, it may be assumed that their work has reflected this increased understanding, i.e. showed improvement.

All in all, the perspectives of legislation, psychology, history, and pedagogy seem to have drawn the same conclusion—that careful planning may produce the greatest skill improvement in the writing of college freshmen. If student writing holds the potential to indicate cognitive growth, then an assessment tool that measures both cognitive level and writing skills may prove invaluable to undergraduates by enabling educators to begin moving in the direction of accountability.

As the previous information indicates, college teaching is not as simple as it may initially appear. Before professors enter the classroom to teach, much preparation and thought has already occurred. While professors may be more individually concerned with presenting content to students, campuses may be more broadly concerned with the external dynamics that affect measurable teaching outcomes. In future, it may be that links between

the external dynamics and individual concerns will need to be made clearer. This study investigated whether the implementation of a specifically designed rubric could be a useful part of assessment, by bridging the broader requirements of the college and the narrower concerns of the individual classroom. The benefit of using this specific rubric is that it includes both cognitive level and writing skill measurements, a different enough approach that it may help campuses move toward more useful and accurate measurement of student outcomes.

There remains a need for an instrument that is both flexible and comprehensive enough to be useful in many disciplines. Because the CLAQWA instrument evaluates both cognitive level and word skill in writing and also focuses on outcomes, it meets current needs. The 16-point rubric categorizes and standardizes the foci of instruction, another valid concern arising from mandates. As a widespread measurement, the CLAQWA instrument may prove to be an invaluable asset to instructors and campuses as they seek to define and clarify assessment goals.

Method

This quasi-experimental, evaluative research study was a nonequivalent control group design. The subjects were assigned to predetermined groups (Writing I class sections). All of the students wrote frequently and were expected to show improvement in the final essay as a result of completing the course.

In the research study, college freshmen from ten different sections of Writing I courses participated. Five courses, taught by the experimental professors, taught and evaluated students using the CLAQWA rubric. Five courses, taught by the control professors, neither taught nor evaluated students using the CLAQWA rubric.

The writing quality and cognitive level quality of the final papers of the two groups of

students were compared by using the CLAQWA rubric. Two independent raters scored the final papers according to the CLAQWA rubric. The SPSS data analysis program assessed the scores. The study investigated any statistically significant differences in the final paper scores. The purpose of this comparison was to determine whether teaching writing according to a specific rubric yielded greater improvement (in either word skill or reasoning skill) in student essays than not teaching writing according to the specific rubric.

Participants

The subjects of this research study were 107 freshmen enrolled at an urban Midwestern state university. The university required all students to take two freshman writing courses, and the first course of this sequence was under consideration. The first required writing course focused on critical thinking skills and writing skills, both of which can be measured by the CLAQWA. Among ten freshmen writing sections of 25 students in each section (students who are primarily 18-year-olds), students were assigned to predetermined course sections via the college enrollment process.

Materials

The diagnostic/sample essay and the final/third essay, served as the data for this research study. Two raters evaluated the student essays according to the CLAQWA instrument. Both the diagnostic and the final essays from both the experimental group and the control group were coded, so that raters did not know which papers belonged to which group. College students submitted assigned writing according to professor instructions. Although both the experimental group and the control group submitted other essays throughout the semester, only the aforementioned essays were evaluated. In addition, only the essays of students who turned in all previous writing assignments on time were used for this research.

This typical case sampling is necessary for an accurate comparison.

The purpose of comparing the diagnostic essays was to determine whether the two groups showed any significant differences between them, in other words, whether the experimental group and the control group could be considered homogeneous at the beginning of the semester. The purpose of comparing the final essays was to determine whether the two groups showed any significant differences in either word skill outcomes or reasoning skill outcomes, in other words, whether the experimental group showed greater skill gains than the control group at the end of the semester.

Procedures

The experimental professors implemented the CLAQWA in the following way. Before the assignments were given, the experimental professors explained the use of the cognitive scale to the experimental group of 55 students, so that students were made aware of the aspects of thinking and learning that take place when a professor gives each assignment. Before each assignment was given, the experimental professors explained the parameters of the assignment, the writing skills, and other specific skills required in the assignment. Instructions from the professor coincided with the guidelines of the CLAQWA rubric. The experimental students received feedback from the CLAQWA rubric.

The control professors did not implement the CLAQWA in the classroom. Before the assignments were given, the control professors did not explain the use of the cognitive scale. To the control group of 52 students, the control professors did not give the same instructions. The control students did not receive feedback from the CLAQWA rubric.

By analyzing 107 students' diagnostic essays and final essays, the researcher hypothesized that the students in the experimental group would show more significant outcomes

in overall writing skills and cognitive level than the students in the control group did.

During the semester, the students in the experimental group submitted three essays to the experimental professors; these essays were evaluated according to the CLAQWA rubric. The diagnostic/sample essay was administered to students before instruction began; it was a timed, 50-minute essay. The third/final essay of the semester was the culminating effort of the semester; it incorporated writing skills, structural development, and analytical skills. The experimental professors evaluated student writing from all essays written during the semester according to criteria from the CLAQWA instrument.

During the semester, the students in the control group submitted three essays to the control professors; these essays were not evaluated according to the CLAQWA rubric. The diagnostic/sample essay was administered to students before instruction began; it was a timed, 50-minute essay. The third/final essay of the semester was the culminating effort of the semester; it incorporated writing skills, structural development, and analytical skills. The control professors did not evaluate student writing from any essays written during the semester according to criteria from the CLAQWA instrument.

The control professors neither taught nor evaluated student essays according to the CLAQWA rubric. The equivalence of both groups was measured by the diagnostic/sample essay. The performance of both groups was measured by the final/third essay. Performance was separated into word skill scores and reasoning skill scores.

All final/third essays papers were assessed by the CLAQWA, a 16-point rubric that can evaluate both writing skills and cognitive skills. Professors can use the scale values (writing, cognitive) separately or in combination, to meet specific needs of specific writing assignments. Cognitive values were derived

from Bloom (1984). Writing quality values were derived from commonly accepted goals, like unity, support/development, coherence, and sentence skills (Langan, 2006; Lunsford, 2010; Rosa & Eschholz, 2007), goals frequently named in college writing textbooks. Both skills were evaluated on a 5-point continuum.

All 107 students submitted two essays: a diagnostic essay before instruction began; and a final essay after completing instruction. Two raters scored essays according to the CLAQWA rubric.

Results

Statistical analysis of this data does not support the initial assumption of the research study—that the experimental group of students who were taught according to a specific rubric would show greater improvement in writing skills and reasoning skills than the control group. The homogeneous groups showed similar enough improvement in writing skills and cognitive skills that it could not be definitely stated that the improvement could be attributed to the use of the specific rubric. In contrast to what was expected, statistical analyses confirmed a commonly held assumption about freshmen college students, that “most students should show improvement” (Hillocks, 1995, p. 207). In other words, professors can feel confident that their manner of explaining a given assignment to students will enable them to succeed at writing.

Discussion

This conclusion may seem more acceptable to some because it supports the various approaches of professors to teach in the way that seems best to them. Statistical analyses may also mean that the control professors were already teaching a mixture of methods that may have included elements of prescriptive, descriptive, and cognitive methods

(Duncan, 2007), i.e. a combination that includes the more balanced or environmental ways of teaching writing. Statistical analyses may also mean that other, carefully presented rubrics may accomplish similar ends. This type of measurement gives one answer to a simple question that has been generated as a result of recent legislative mandates: quantitative measurements of writing change may be beneficial to campuses as they strive to meet educational mandates.

The study did not inquire into students' composing habits but focused instead on professors' outcomes. This suggests that studying other factors, which may also affect writing outcomes, may enhance the understanding of student outcomes, i.e., how to bring about maximum improvement.

The research should remind educators that rating essays according to a rubric does not perform the same function as grading an essay. The purpose of grading an essay is specialized and geared to individual students. Writing change is expected from students as a result of grading. The purpose of rating according to a rubric is generic and geared to established standards. Writing change has already occurred at the time of rating. Rubrics intended for public purposes can indicate improvement, but they cannot be compared to the copious grading that professors give to student essays, nor can they be compared to the very exacting standards of standardized testing measurements that students and professionals may be familiar with.

The research indicates that professors can continue to teach to realistic goals commonly understood to be the aims of most college level writing courses. Professors can continue to expect to see student improvement without worrying about teaching according to a specific method. In addition, colleges need not move in the direction of standardized teaching methods, for educational freedom remains of paramount importance to instructors. At the

same time, colleges are advised to learn how to conduct educational research appropriately, with the understanding that such research is limited in what it can measure.

Clearly, rubrics will continue to hold central importance to those who evaluate student writing for personal, institutional, or accountability purposes. Using an appropriate rubric such as the CLAQWA will allow professors the freedom in teaching that is so central to any academic discussion. Perhaps departments may employ this rubric as a teaching tool for instructors. The quantification of results and consistent record keeping may allow for more long-range comparisons, beneficial to departments when conducting self-assessment.

This research confirms that guidance from professors—neither a complete domination of instruction nor a total absence of instruction—helps students learn the skills of the course. This coincides with the findings of several researchers (Chickering & Gamson, 1987; Duncan, 2007; Hillocks, 1986, 1995) who have inquired, in a systematic way, about writing practices, habits, and methods since the 1980s up to the current time.

While this research study confirms that students do improve their skills during the duration of the semester-long course with or without the use of a specific rubric, it in no way undermines the efficacy of the CLAQWA rubric, an instrument that measures both cognitive level and quality of writing in college student essays. Nevertheless, this research study does affirm that the use of a specific rubric may not be the universal answer to measurement concerns. As a result, this research supports the diverse initiatives of professors to teach according to their preference.

Finally, this research refutes a prevailing idea that there is one specific way to teach writing or to measure change in writing. Instead, it supports previous research about composition, research that comes from pedagogical,

historical, psychological, and quantitative perspectives, research that began in the 1980s and continues to the present day. At the same time, the analyses raised points that may be of interest to future researchers, such as what professors may reasonably expect to accomplish when teaching composition.

References

- Action plan for higher education: Improving accessibility, affordability and accountability.* (2006). Retrieved from United States Department of Education, No Child Left Behind page: <http://www.ed.gov/nclb>
- Bloom, B.S., ed. (1984). *Taxonomy of educational objectives: The classification of educational goals, Handbook 1, Cognitive domain.* New York: Longman.
- Boards and Commissions: Higher education update.* (2006). Retrieved from United States Department of Education, No Child Left Behind page: <http://www.ed.gov/nclb>
- Chickering, A.W., & Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin*, 39(7), 3-7.
- CLAQWA online: A new tool for improving student writing and thinking.* (2007). Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Duncan, M. (2007). Whatever happened to the paragraph? *College English*, 69(5), 470-495.
- Executive summary of changes to the principles of accreditation.* (2007). Retrieved from Southern Association of Colleges and Schools, Commission on Colleges: www.sacscoc.org
- Flateby, T.L. (2007). *Closing the loop: Assessing and improving writing and thinking throughout the curriculum.* Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Flateby, T.L., & Eubanks, D. (2008). *CLAQWA online: Next steps in writing assessment.* Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Flateby, T.L., & Metzger, E. (2001). Instructional implications of the cognitive level and quality of writing assessment (CLAQWA). *Assessment Update*, 13(1), 4-5, 11.
- Flateby, T.L., & Metzger, E.A. (n.d.). *CLAQWA: Cognitive level and quality writing assessment system.* Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Four pillars of NCLB.* (2004). Retrieved from United States Department of Education, No Child Left Behind page: <http://www.ed.gov/nclb>
- Hillocks, G. (1986). *Research in written composition: New directions for teaching.* Urbana, IL: National Conference of Research in English.
- Hillocks, G. (1995). *Teaching writing as reflective practice.* New York: Teachers College Press.
- Langan, J. (2006). *English skills with readings* (6th ed.). Boston: McGraw Hill.
- Lunsford, A.A. (2010). *Easy writer: A pocket reference* (4th ed.). Boston: Bedford/St. Martin's.
- Peer review of student writing examined.* (2006). Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Perry, W.G. (1999). *Forms of intellectual and ethical development in the college years: A scheme.* San Francisco: Jossey-Bass.
- Primary trait analysis used by the CLAQWA system.* (2007). Retrieved from University of South Florida, Assessment Center: <http://usfweb2.usf.edu/assessment/index.shtml>
- Ramey, S., VandeVusse, L., & Gosline, M. (2007). Using a written communication rubric to improve students' writing. *International Journal of Learning*, 13(10), 67-74.
- Roen, D.H., Pantoja, V., Yena, L., Miller, S.K., & Waggoner, E. (Eds.). (2002). Strategies for teaching first-year composition. ERIC Document Reproduction Service No. ED459476.
- Rosa, A. & Eschholz, P. (2007). *Models for writers: Short essays for composition* (9th ed.). Boston: Bedford/St. Martin's.
- U.S. Department of Education awards \$2.45 million for improved measures of achievement for post-secondary students.* (2007). Retrieved from United States Department of Education, No Child Left Behind page: <http://www.ed.gov/nclb>

APPENDIX E: SUMMARY OF RESULTS

Table E1: Inter rater reliability Diagnostic Essay

Descriptive Statistics			
	Mean	Std. Deviation	N
diag1	16.9252	4.90610	107
diag2	14.4953	3.39637	107

Correlations			
	diag1	diag2	
diag1	Pearson Correlation	1	.636**
	Sig. (2-tailed)		.000
	N	107	107
diag2	Pearson Correlation	.636**	1
	Sig. (2-tailed)	.000	
	N	107	107

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

Both graders showed a statistically significant correlation on the grading of the diagnostic essay. The Pearson correlation was .636 (sig. = 0.00) at $\alpha = 0.01$ in a two-tailed test.

Table E1: Inter rater reliability Final Essay

Descriptive Statistics			
	Mean	Std. Deviation	N
final1	25.5421	6.33541	107
final2	24.7850	3.09604	107

Correlations			
	final1	final2	
final1	Pearson Correlation	1	.252**
	Sig. (2-tailed)		.009
	N	107	107
final2	Pearson Correlation	.252**	1
	Sig. (2-tailed)	.009	
	N	107	107

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

Both graders showed a statistically significant correlation on the grading of the final essay. The Pearson correlation was .252 (sig. = 0.09) at $\alpha = 0.01$ in a two-tailed test.

Table E2: One Way ANOVA for Diagnostic Essay (Experimental and Control Groups)

Descriptives									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
diag1	experimental	55	17.0182	4.60471	.62090	15.7734	18.2630	6.00	25.00
	control	52	16.8269	5.24954	.72798	15.3654	18.2884	5.00	25.00
	Total	107	16.9252	4.90610	.47429	15.9849	17.8656	5.00	25.00
diag2	experimental	55	14.2364	3.10891	.41921	13.3959	15.0768	8.00	20.00
	control	52	14.7692	3.68684	.51127	13.7428	15.7957	8.00	25.00
	Total	107	14.4953	3.39637	.32834	13.8444	15.1463	8.00	25.00

Comparison of Effects of CLAQWA Rubric on Freshman College Student Writing / 457

Test of Homogeneity of Variances				
	Levene Statistic	df1	df2	Sig.
diag1	.900	1	105	.345
diag2	1.077	1	105	.302

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
diag1	Between Groups	.978	1	.978	.040	.841
	Within Groups	2550.424	105	24.290		
	Total	2551.402	106			
diag2	Between Groups	7.590	1	7.590	.656	.420
	Within Groups	1215.158	105	11.573		
	Total	1222.748	106			

ANOVA is robust to heterogeneity of variance if sample sizes are equal or near equal. However, a Levene analysis was run to confirm homogeneity of variance in the samples. The Levene statistic tests the null hypothesis of homogeneity of variance with $\alpha < .10$ being needed to reject the null hypothesis. Both samples failed to reject the null hypothesis of homogeneity of variance.

There were no statistically significant differences between the experimental and control groups for the diagnostic essays graded by rater one or rater two. With an alpha level of $\alpha = 0.05$, diagnostic essay one measured $\alpha = 0.841$ and diagnostic essay two measured $\alpha = 0.420$. Thus, the null hypothesis of no significant difference between the experimental and control groups in performance on the diagnostic essay was not rejected.

Table E2: One Way ANOVA for Final Essay (Experimental and Control Groups)

		Descriptives							
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
final1	experimental	55	26.0727	6.55138	.88339	24.3016	27.8438	14.00	35.00
	control	52	24.9808	6.11167	.84754	23.2793	26.6823	7.00	35.00
	Total	107	25.5421	6.33541	.61247	24.3278	26.7563	7.00	35.00
final2	experimental	55	24.6364	3.63809	.49056	23.6529	25.6199	15.00	33.00
	control	52	24.9423	2.42061	.33568	24.2684	25.6162	19.00	28.00
	Total	107	24.7850	3.09604	.29931	24.1916	25.3784	15.00	33.00

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
final1	.351	1	105	.555
final2	5.407	1	105	.022

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
final1	Between Groups	31.871	1	31.871	.792	.375
	Within Groups	4222.690	105	40.216		
	Total	4254.561	106			
final2	Between Groups	2.502	1	2.502	.259	.612
	Within Groups	1013.554	105	9.653		
	Total	1016.056	106			

ANOVA is robust to heterogeneity of variance if sample sizes are equal or near equal. However, a Levene analysis was run to confirm homogeneity of variance in the samples. The Levene statistic tests the null hypothesis of homogeneity of variance with $\alpha < .10$ being needed to reject the null hypothesis. Both samples failed to reject the null hypothesis of homogeneity of variance.

There were no statistically significant differences between the experimental and control groups for the final essays graded by rater one or rater two. With an alpha level of $\alpha = 0.05$, final essay one measured $\alpha = 0.375$ and final essay two measured $\alpha = 0.612$. Thus, the null hypothesis of no significant difference between the experimental and control groups in performance on the final essay was not rejected.

Table E2: Word Skills Analysis for Final Essay (Experimental and Control Groups)

The study in question sought to test for statistically significant differences between the experimental group ($N = 55$) and the control group ($N = 52$) on the word skills portion of the Cognitive Level and Quality Writing Assessment system (CLAQWA) instrument. The analysis was further discriminated by the use of two independent graders.

**Word Skills Analysis
Descriptives**

		N	Mean	Std. Deviation	Std. Error
wordsk1	experimental	55	10.4545	2.03505	.27441
	control	52	10.6923	1.52802	.21190
	Total	107	10.5701	1.80205	.17421
wordsk2	experimental	55	14.1455	2.02227	.27268
	control	52	14.2308	1.42272	.19730
	Total	107	14.1869	1.74902	.16908

Comparison of Effects of CLAQWA Rubric on Freshman College Student Writing / 459

The descriptive statistics show little significant difference between the mean scores for the experimental and control groups as reflected between the two groups and the overall mean score. This pattern holds for both graders.

The Levene statistic was used to test for homogeneity of variance. The null hypothesis of the test of homogeneity of variance is that there is no significant difference between the variances of the two groups. The hypothesis is tested at $\alpha = .10$.

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
wordsk1	3.855	1	105	.052
wordsk2	4.244	1	105	.042

The null hypothesis of no difference between the variances of the experimental and control groups was rejected for both graders in the word skills portion of the test. However, this does not prove to be problematic for an ANOVA analysis since the sample sizes are almost equal. ANOVA is robust to heterogeneity of variance if the groups sizes are equal or near to equal.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
wordsk1	Between Groups	1.511	1	1.511	.463	.498
	Within Groups	342.713	105	3.264		
	Total	344.224	106			
wordsk2	Between Groups	.195	1	.195	.063	.802
	Within Groups	324.067	105	3.086		
	Total	324.262	106			

The ANOVA shows no statistically significant difference between the two groups measured as wordsk1 (Grader 1). The alpha level for the analysis was $\alpha = .05$. The between groups significance was $\alpha = .498$, thus failing to reject the null hypothesis of no significant difference between the two groups. This also held true for the groups designated wordsk2 (Grader 2). The between groups significance was $\alpha = .802$, thus failing to reject the null hypothesis of no significant difference between the groups.

Table E2: Reasoning Skills Analysis for Final Essay (Experimental and Control Groups)

**Reasoning Skills Analysis
Descriptives**

		N	Mean	Std. Deviation	Std. Error
reason1	experimental	55	11.1455	2.84422	.38351
	control	52	10.7115	2.61471	.36260
	Total	107	10.9346	2.73092	.26401
reason2	experimental	55	14.9273	3.74094	.50443
	control	52	14.2692	3.54310	.49134
	Total	107	14.6075	3.64396	.35227

The descriptive statistics show little significant difference between the mean scores for the experimental and control groups as reflected between the two groups and the overall mean score. This pattern holds for both graders.

The Levene statistic was used to test for homogeneity of variance. The null hypothesis of the test of homogeneity of variance is that there is no significant difference between the variances of the two groups. The hypothesis is tested at $\alpha = .10$.

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
reason1	.826	1	105	.366
reason2	.171	1	105	.680

The null hypothesis of homogeneity of variance was not rejected. It can be safely assumed that the two groups, for each grader, possessed equal or near equal variances.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
reason1	Between Groups	5.033	1	5.033	.673	.414
	Within Groups	785.509	105	7.481		
	Total	790.542	106			
reason2	Between Groups	11.574	1	11.574	.871	.353
	Within Groups	1395.940	105	13.295		
	Total	1407.514	106			

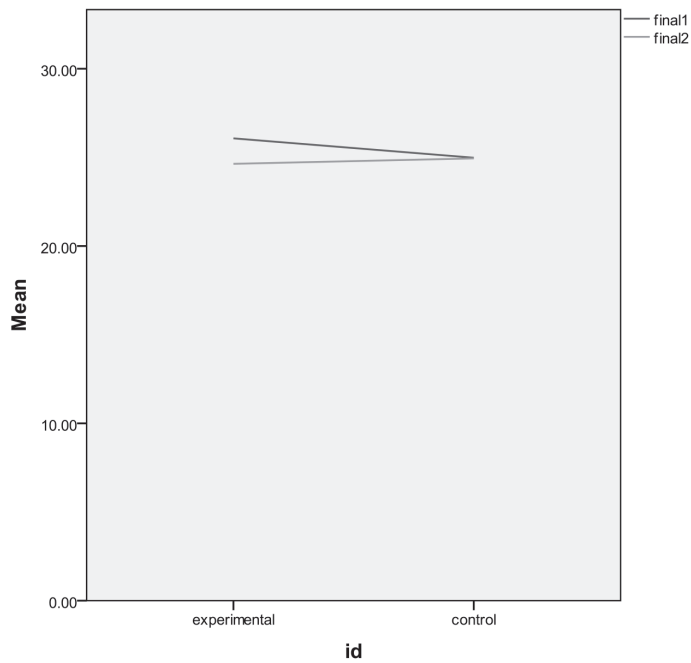
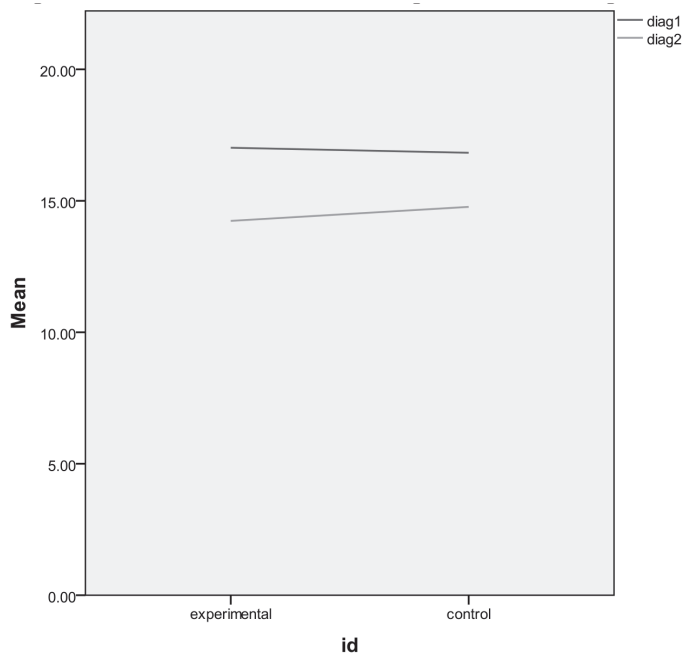
The ANOVA shows no statistically significant difference between the two groups measured as reason1 (Grader 1). The alpha level for the analysis was $\alpha = .05$. The between groups significance was $\alpha = .414$, thus failing to reject the null hypothesis of no significant difference between the two groups. This also held true for the groups designated reason2 (Grader 2). The between groups significance was $\alpha = .353$, thus failing to reject the null hypothesis of no significant difference between the groups.

Summary

Based on the results of these statistics it can be concluded that no statistically significant difference exists between the experimental group and the control group used in the analysis. This held true for the evaluations given by both Grader 1 and Grader 2.

Comparison of Effects of CLAQWA Rubric on Freshman College Student Writing / 461

Graphs of Scores for Control and Experimental Groups



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.