

Reliability and Validity of Rubrics for Assessment Through Writing

Ali Reza Rezaei

College of Education, California State University, Long Beach, California, United States

Michael Lovorn

Department of Curriculum and Instruction, University of Alabama, Tuscaloosa, United States

Abstract

This experimental project investigated the reliability and validity of rubrics in assessment of students' written responses to a social science "writing prompt". The participants were asked to grade one of the two samples of writing assuming it was written by a graduate student. In fact both samples were prepared by the authors. The first sample was well written in terms of sentence structure, spelling, grammar, and punctuation; however, the author did not fully answer the question. The second sample fully answered each part of the question, but included multiple errors in structure, spelling, grammar and punctuation. In the first experiment, the first sample was assessed by participants once without a rubric and once with a rubric. In the second experiment, the second sample was assessed by participants once without a rubric and once with a rubric. The results showed that raters were significantly influenced by mechanical characteristics of students' writing rather than the content even when they used a rubric. Study results also indicated that using rubrics may not improve the reliability or validity of assessment if raters are not well trained on how to design and employ them effectively.

Keywords: Writing Assessment; Reliability; Validity; Rubrics; ESL writing

1. Introduction

The use of rubrics as a means of assessment of student performance is more popular now than ever in the United States. Classroom teachers of the 21st century use rubrics to assess everything from students' writing to their ability to follow prescribed directions. Although many

studies report that rubrics enhance and enrich assessment of student work, the validity of the tool is not without debate. Much of the debate on rubric effectiveness relating to students' writing is discussed in the field of English composition or English proficiency exams rather than writing in other academic disciplines. Additionally, most of these studies have been limited to descriptive or argumentative articles rather than experimental investigations (Meier, Beverly, & Cady, 2006). The purpose of this study was to investigate the reliability and validity of rubrics as effective tools for assessment of graduate students' written response to a social study "writing prompt". Therefore, this paper intends to investigate reliability of assessment *through* writing rather than assessment *of* writing.

Students' skills and abilities relating to writing and reading are measured as much as, if not more than, any other in today's assessment-driven public education system. A student's writing is not only used to evaluate her/his English proficiency, but also to assess her/his understanding of other subjects such as social studies, law, economics, and physical and natural sciences. Writing is also considered an important part of almost all university level courses. Decades of research have shown that there is a significant correlation between verbal ability and academic success factors such as intelligence, critical thinking, and self-esteem (Skaalvik & Skaalvik, 2004; Munoz, Frick, Kimonis, & Aucoin, 2008; Follman, 1993)

The use of rubrics for evaluating students' writing emerged from a general dissatisfaction, among teachers and administrators, with traditional essay grading strategies. In today's educational environment of high stakes assessment, many educators regularly and confidently employ rubrics as a way to assess students' work. This is an indication that rubrics are highly regarded as tools that increase reliability and validity in assessment. It should be noted; however, that simple implementation of rubrics may not guarantee effective assessment (Breland, 1983; Ross-Fisher, 2005; Tomkins, 2003).

Several researchers have reported that teachers' assessment is more reliable if a rubric is used (Silvestri & Oescher, 2006; Jonsson & Svingby, 2007). No research has been found to show a negative effect of using rubrics (decreasing the reliability). Consequently, many teachers have used rubrics with the assumption that they increase grading objectivity, particularly, regarding students' writing submissions. As a result, another assumption exists that assessment without a rubric tends to be more subjective because it is based only on each grader's subjective judgment, and her/his overall impressions of the writer's style. With this in mind, teachers often resolve that using a rubric is better than not using one (Spandel, 2006). Researchers have asked; however, if this assumption about rubrics is based on false claims to objectivity, or if they simply make subjectivity more visible (Turley & Gallagher, 2008).

2. Background

In this paper, a rubric is defined as a set of criteria for grading assignments. A rubric can be either holistic or analytical, or it may be a combination of the two. A holistic rubric is used to assess the overall quality of a student's response. Holistic rubric is more product-oriented than process-oriented, and are primarily concerned with the total performance or product rather than with the individual steps taken to arrive at the final product (Finson, 1998). An analytical rubric consists of multiple, separate scales, and therefore provides a set of scores rather than just one. For example, a given writing assignment could be assessed with an analytical rubric made up of three scales wherein five points is given for creativity, four points is given for reasoning or critical thinking, and six points is given for sentence structure.

According to Hamp-Lyons (2002) essay testing has been around for thousands of years. As early as the 1970s, perceptions arose in the United States that holistic essay grading, when left to the individual teacher, can be overly subjective, often resulting in a lack of reliability, validity, and predictive assessment (Breland, 1983). Mechanical aspects of writing, particularly

spelling and grammar, have been found to be highly influential factors relating to how an essay is rated. Research indicates that a student's skill level with writing conventions such as writing style, vocabulary command, verbal ability, spelling, punctuation, and grammar can have a significant impact, whether positive or negative, upon how a grader values a student's work (Ross-Fisher, 2005; Read, Francis, & Robson, 2005).

Rater's bias is evident in many forms (Ross-Fisher, 2005; Tomkins, 2003; Read, Francis & Robson, 2005). It is noteworthy that students who write neatly and display better basic writing mechanics regularly receive higher marks on their essays than students who lack these skills, even though their attention to content is otherwise identical (James, 1927; Sheppard, 1929; Briggs, 1970; Markham, 1976; Bull & Stevens, 1979; Chase 1968; Marshall & Powers 1969). Gage and Berliner (1992) investigated reliability and validity of essay grading by having a group of 100 teachers grade an essay without a rubric. All teachers graded the same essay, scoring it on a 100-point scale. Despite the fact that each teacher was given the same instructions, grades varied greatly. Their scores ranged from 60 to the upper 90s. As a part of this investigation, researchers also asked participating teachers to estimate the grade level of the student who wrote the paper. Again, answers varied significantly. Estimates ranged from as high as the junior year in college to as low as 5th grade elementary school level. Reasons for this variance may be found in studies in assessment bias. According to recent findings, graders' biases have related to gender, language command, and even physical attractiveness of students (Malouff, 2008), and bias effect is accentuated when raters tend to rate unusually harshly or leniently (Knoch, Read, & Randow, 2007). Additionally, essays written in good penmanship are frequently assigned higher marks than essays written in poor penmanship (Chase, 1968; Marshall & Powers, 1969). According to Scannell and Marshall (1966), essays with several composition errors, including punctuation, spelling, and grammar mistakes, often resulted in lower scores than essays free of these mistakes.

This held true even when graders had been prompted to score on content alone. Related research by Marshall and Powers (1969) also found an inverse relationship between scores assigned to an essay and the number of composition errors (spelling, grammar, and punctuation) in the paper.

In contrast, rubrics have been lauded as tools that have effectively leveled the playing field for all (Jonsson & Svingby, 2007). However, do rubrics lead to a more reliable and less biased assessment in comparison with traditional essay grading? Today many teachers feel more confident about their assessment of student writing when rubrics are employed (Silvestri & Oescher, 2006). Rating scales used in performance assessment have been repeatedly criticized for being imprecise and therefore often resulting in holistic marking by raters. This means even if a rubric is used the grade might be mainly assigned based on teacher's overall impression. These studies have also determined that criteria that use impressionistic terminology are more open to subjective or ambiguous interpretations (Knoch, 2009; Weigle, 2002).

Teachers, schools, and school systems have adopted rubrics for more accurate assessment in every discipline. Recently, however, some educators have challenged the collective assumption that simply implementing rubrics increases inter-rater reliability and validity, and the overall accuracy and quality of assessment (Kohn, 2006; Wilson, 2006). In steadily increasing numbers, educators are coming to realize that no rubric can be completely effective in evaluation of students' individual writing idiosyncrasies or their unique understanding of the concepts. Some have even found that rubrics prematurely narrow and cement their visions of good writing (Wilson, 2007).

This paper investigates the extent to which the use of rubrics helps prevent raters from paying too much attention to writing mechanics over a focus on the reasoning, content knowledge, and logical thinking in graduate student writing.

2.1. Reliability and Validity of Rubrics

The reliability and validity of holistic writing assessment strategies have been studied for years; however, analytical rubric-based assessment has not been adequately and *experimentally* studied. Although analytical rubrics have emerged as one of the most popular assessment tools in progressive educational programs, there is an unfortunate dearth of information in the literature quantifying the actual effectiveness of the rubric as an assessment tool (Hafner & Hafner, 2003). More study also needs to be conducted into the degree to which raters are affected by superficial factors such as structural, spelling, grammar and punctuation errors when they use rubrics (Elliot, 2005).

As stated previously, the bulk of existing research on rubrics has been done in the area of English writing and composition, and these studies have typically focused primarily on holistic essay grading rather than analytical rubric-based assessment (Meier, Beverly, & Cady, 2006). Brookhart (2005) suggested that additional study needs to take place to ensure that assessment in all subjects be reliably judged with the rubrics. Nevertheless, quantitative and experimental research methods have rarely been used in this regard (Andrade, Du, and Wang, 2008).

There has been a longstanding, widespread assumption among academics that rubrics tend to improve inter-rater reliability in terms of how likely different raters will award similar scores. Researchers point out; however, that this assumed level of consistency is not commonly achieved because, among other reasons, raters do not all receive rubric training (Knoch, Read, & Randow 2007). Consequently, without proper, thorough training, a rubric may become little more than a checklist.

Some educators make strong arguments against the application of rubrics for writing assessment. According to some of these educators, a rubric, in and of itself, is not a good or a bad thing. Rather, it depends on how it is used (Wilson, 2007; Turley & Gallagher, 2008). Although many articles of this vein regularly describe the basic rubric as a highly effective and ultimately

consistent grading tool, it may be asserted that problems can and do exist regarding their use. As mentioned earlier, little empirical research has been performed to study whether rubrics may decrease the reliability of teachers' assessment. It should be mentioned that at least one research study indicated that in spite of rubric training, readers' judgments were strongly influenced by salient though superficial characteristics of writing samples (Charney, 1984).

A major trend on the effectiveness of rubrics centers on inter-rater reliability and intra-rater reliability. Traditionally, inter-rater reliability has been measured by the correlation between two sets of grades assigned to a group of students by two different raters (Hafner & Hafner, 2003; Newell, Dahm, & Newell, 2002). Some studies have found very high inter-rater reliability scores for their rubrics (Penny, Johnson, & Gordon, 2002), while others have reported a low or moderate (less than .70) reliability. Hamp-Lyons (1991) reported some relatively high reliability for her "multiple trait instrument" which is similar to current analytical rubrics. However, unlike many other rubrics multiple trait scoring methods are developed on-site for a specific purpose and a specific group of writers and they do not focus on trivial features of text (grammar, spelling, handwriting). The intra-rater reliability (consistency of grading a given writing by the same rater twice) is reported to be higher (Cronbach's alpha about .70) than the inter-rater reliability when a rubric is used (Jonsson & Svingby, 2007).

The problem with inter-rater reliability; however, is that this index is calculated based on z-scores, and z-scores depend on the distance between the mean and each individual score. To clarify, consider a case in which one rater is considerably harsh in her assessment, while another one is relatively lenient. One would expect their scores for a given paper to be quite different; however, it is conceivable, perhaps even likely, that raw scores converted into z-scores may turn out equal. It must, therefore, be considered that the reported inter-rater reliability might be artificially high simply because the assigned relative rankings are the same, and not because the

measurement is accurate (Johnson, Penny, & Gordon, 2000). Considering this, it could be concluded that inter-rater reliability is more appropriate for a norm-referenced assessment than a criterion-based environment. Therefore, in this study we use a different method to evaluate the reliability and validity of rubrics.

This study intends to investigate the reliability and validity of rubrics in assessment through students' writing. The goal is to see if using a rubric leads to a more reliable assessment in comparison with the case where no rubric (holistic assessment) is used.

3. Method

3.1. Subjects

A total of 326 participants, all of whom were college students, took part in this study. The authors divided participants into four groups. The first two groups (N=71 and N=108) primarily consisted of graduate education majors. These participants' ages ranged from 20 to 60, and about 70% of them were female. Most of them were relatively new classroom teachers, averaging between two and three years of work experience in schools. Groups three and four (N= 85 and N=72) were primarily business and marketing majors whose ages ranged from 20 to 55. About 55% of them were female.

3.2. Experiments

This study used standard error of measurement to evaluate the reliability of rubrics because it is considered more appropriate for a criterion-based assessment (Feldt & Qualls, 1999). For the purposes of this study, reliability is defined as the accuracy of measurement as measured by consensus among several raters on assessment of the same essay. The standard error of measurement, which has been widely used in the literature to evaluate the accuracy of measurement, is reported to be a more robust index of reliability (Feldt & Qualls, 1999). This

index is expected to be more powerful because so many raters evaluate the same essay using the same rubric. The authors presented a large group of participants with the same completed essay and had each of them grade it accordingly (Appendix A). The degree of consensus as measured by standard deviation of the assigned grades is considered to be the index of variability which in turn is used as an index of reliability. Ideally, the standard error of measurement should be zero; however, any SD less than the margin of error (less than 5% of the range of any given scale) was considered to be acceptable in this study.

3.2.1. Experiment 1

Group 1 (education students) and Group 3 (business and marketing students) participated in this experiment. Participants were asked to grade a written essay (Appendix A) on the topic of economic globalization on a 100-point scale. They were asked to grade the essay once without a rubric and then with a rubric. Participants were not privy to the writer's name, specific age, or level of education, but were instructed to assume the writer was a student of an advanced social studies class and that the essay was the student's response to the following prompt:

“In an essay, discuss economic globalization in terms of its history; economic, social, and political impacts; and how information technology has influenced the speed of globalization in terms of outsourcing and off-shoring.”

Although the essay response was well-written in terms of skills and mechanics, it covered only a broad description of economic globalization, and it did not fully address any element of the above prompt. The goal of this experiment was to evaluate how participants were influenced by and/or impressed with the mechanics and superficial characteristics of the essay, rather than the correctness or the accuracy of the answer.

3.2.2. Experiment 2

Group 2 (education students) and Group 4 (business and marketing students) participated in this experiment. The research design was similar to the first experiment (the same prompt and the same rubric were used); however, the essay given to these participants was different (Appendix B). Unlike the first essay, this second essay accurately addressed all parts of the prompt, and according to the rubric, deserved a high score because the writer answered all questions and used a variety of proper sources (references) in a complete and concise response. The penmanship of this essay; however, was obstructed by 20 structural, mechanical, spelling and grammar errors. In fact the authors inserted these 20 spelling and grammar errors in the second essay to investigate how mechanics influenced the raters' grading.

3.3. Instruments

The rubric used in this study (Appendix C) was very similar to writing assessment rubrics currently being employed in several programs at the College of Education at California State University, Long Beach (where this study was conducted). It is designed to conform to the 100-point scale, and contained the following criteria:

- *Structural organization and clarity (25 points)*
- *Understanding and synthesis of argument (25 points)*
- *Understanding the goals and implications of globalization (25 points)*
- *Support and citation of sources (15 points)*
- *Writing mechanics (10 points)*

As mentioned earlier, two sample essays were designed specifically for use in this study. The first response used in the first experiment is referred to in this paper as the "Wrong Essay". Although it was eloquently and professionally written, it contained only a broad description of globalization and did not fully address or answer any of the above prompts. The second response used in the study is referred to in this paper as the "Correct Essay". The authors inserted 20

structural and mechanical errors into this essay, yet it was written to adequately answer the questions outlined in the prompt .

4. Hypotheses

Regarding the literature review the authors predicted that using the rubric will lower the range and the variability of scores and therefore, increase the reliability of grading. In contrast with the holistic assessment, the rubric was clearly designed to evaluate the writer's success at answering the essay prompts, so it was predicted that participants' assessments will be predictably reliable, valid, and consistent. In the first experiment, the authors predicted that using the rubric would lead to high scores in "mechanics" and "structural organization and clarity" and low scores on "understanding and synthesis of argument" and "understanding the goals and implications of globalization". The authors also predicted that using the rubric would lead to a very low score on citations and references because the writer failed to cite any sources for the essay. It should be noted that the essay prompt did not specifically indicate that citations or references were required; however, researchers understood that as graduates, study participants are expected to support their arguments.

In the second experiment, the authors expected that the participants would be negatively influenced by the poor spelling and grammar, and that they would assign an overall low score particularly to the writer's mechanics. Finally, since education students are trained in assessment and rubrics and they are familiar with the issue of teacher bias the authors expected this group to be less influenced by the mechanics of the essay. The authors predicted that their rating would be more reliable than the ratings of students from the college of business.

5. Results

Table 1 shows the distribution of the participants in this study categorized in four groups. This table also shows a comparison of assigned grades with and without rubrics for each of the

above four groups. As shown in the table, the assigned scores with the rubric are lower than the assigned scores without the rubric.

Table 2 shows that for all groups this difference is significant. Table 3 shows that in both experiments the range and the variance of assigned scores increased significantly after using the rubrics. Table 4 shows details of the grading using the rubric in each experiment. The lowest range was observed in mechanics and the widest range was observed in organization. The participants from the College of Education were compared with those in the College of Business and Marketing using MANOVA. Table 5 shows that participants from College of Education rated the first essay (the more eloquently written essay that failed to answer the prompts) significantly higher than the second essay (with the answer with correct content, which also included twenty spelling and grammar errors). By contrast, the participants from the College of Business and Marketing rated both essays similarly.

Table 1- Assigned scores with and without a rubric

Groups		without	with
Wrong Answer Ed	Mean	79.55	68.00
	N	71	71
Correct Answer Ed	Mean	72.78	58.50
	N	106	108
Wrong Answer Bus	Mean	67.83	58.38
	N	84	85
Correct Answer Bus	Mean	73.1	59.19
	N	71	72
Total	Mean	73.63	64.57
	N	356	467

Table 2- Testing the difference between assigned scores with and without rubrics

Groups	without	with	t	sig
Wrong Answer Ed	79.55	68.00	-9.085	.000
Correct Answer Ed	72.78	58.50	-12.544	.000
Wrong Answer Bus	67.83	58.38	-7.723	.000
Correct Answer Bus	73.1	59.19	-9.832	.000

Table 3 -Range and variance of assigned scores with and without rubrics

Groups		Without	With
Wrong Answer Ed	Std. Deviation	10.50	15.05
	Minimum	49.00	32.00
	Maximum	96.00	100.00
Correct Answer Ed	Std. Deviation	10.03	14.98
	Minimum	27.00	12.00
	Maximum	98.00	98.00
Wrong Answer Bus	Std. Deviation	10.19	14.31
	Minimum	40.00	21.00
	Maximum	90.00	86.00
Correct Answer Bus	Std. Deviation	12.59	14.86
	Minimum	25.00	27.00
	Maximum	100.00	100.00

Table 4 - Assigned scores for each of the 5 categories of the rubric

Essay	N	Minimum	Maximum	Mean	SD
Wrong Answer					
Organization (25 points)	287	.00	25.00	16.55	5.09
Synthesis (25 points)	287	3.00	25.00	17.86	4.66
Specific answers (25 points)	287	2.00	25.00	18.15	4.55
Citation (15 points)	287	.00	22.00	8.65	4.58
Mechanics (10 points)	287	.00	10.00	7.00	3.17
Rubric (100 points)	287	21.00	100.00	68.21	16.01
Correct Answer					
Organization	180	2.00	25.00	13.52	4.95
Synthesis	180	3.00	25.00	15.43	4.16
Answer	180	.00	25.00	16.05	4.66
Citation	180	1.00	19.00	9.29	3.83
Mechanics	180	.00	10.00	4.47	3.17
Rubric	180	12.00	100.00	58.77	14.90
Total					
Organization	467	.00	25.00	15.38	5.25
Synthesis	467	3.00	25.00	16.92	4.63
Answer	467	.00	25.00	17.34	4.70
Citation	467	.00	22.00	8.90	4.315
Mechanics	467	.00	10.00	6.02	3.40
Rubric	467	12.00	100.00	64.57	16.24

Table 5 – Comparing teachers enrolled in the college of education (Ed) with masters students from the college of business and marketing (Bus).

Group	Subgroup	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Lower Bound
With rubric	Wrong Answer Ed	68.00	1.76	64.53	71.47
	Correct Answer Ed	58.58	1.44	55.74	61.42
	Wrong Answer Bus	58.354	1.62	55.16	61.54
	Correct Answer Bus	58.965	1.76	55.49	62.43
Without rubric	Wrong Answer Ed	79.554	1.28	77.04	82.07
	Correct Answer Ed	72.78	1.04	70.72	74.84
	Wrong Answer Bus	67.833	1.17	65.52	70.14
	Correct Answer Bus	73.099	1.28	70.58	75.61

6. Summary of findings

Four hypotheses were tested in this project. First, the authors predicted that the rubric would lower the range and the variability of scores and, therefore, increase the reliability of grading. The results did not support this hypothesis. Findings indicated that using the rubric did not lessen the range of assigned scores to a given essay. In fact, on the contrary, the authors found that the assigned grade variance increased significantly after implementing the rubric. Furthermore, it showed that participants (particularly from college of education) were strongly influenced by the trivial mechanics and superficial aspects of students' writing. The authors came to this conclusion because the only problem with the second essay was the grammar and spelling errors, therefore, higher grades were expected for this essay (in comparison with the first essay). However, the authors observed that many raters assigned low scores to this essay because the writer had demonstrated poor sentences structure and mechanics. When the authors asked the participants to justify their assigned grades, several responded with explanations to this effect: "The student's poor writing style indicated she/he did not take time (or perhaps even care enough) to spell check or to revise the sentence structure, and so she/he deserves the poor grade".

Statements of this nature elicited a response from the authors (among ourselves) questioning the very nature and purpose of rubrics. The authors found it interesting that although rubrics were designed to reduce or eliminate rater's bias such as this, and although this rubric placed very little assessable attention on writing mechanics (only 10 points out of 100), it was obviously still a significant factor in raters' assessments.

Second, it was predicted that in the first experiment using the rubric would lead to high scores in "mechanics" and "structural organization and clarity," and that low scores would result from "understanding and synthesis of argument" and "understanding the goals and implications of globalization". Results did not support this hypothesis. This is evident from the fact that although the first version of the essay did not address the prompts, it not only received a passing grade (overall 68.2 out of 100), but also it received a passing grade (18.1 out of 25) on the item specifically evaluating the writer's attempt to adequately address the prompt. The authors also predicted that using the rubric would lead to a very low score (ideally zero) on citations and references because the writer failed to cite any sources for the essay. It was particularly interesting to note that raters gave credit (8.6 out of 15) to some aspects (citation) of the paper that it did not contain at all. It should be noted that lack of any citation in the essay is not a subjective matter. Giving credit to something that did not exist indicated that the raters were extremely influenced by their overall impression due to the very impressive writing style of the first essay. As noted by Lumley (2002), although raters try to remain close to the rubric, they are heavily influenced by the complex intuitive impression of the text obtained when they first read it.

Third, it was predicted that in the second experiment, the participants would be negatively influenced by essay writer's spelling errors and grammatical mistakes, and that they would assign an overall low score, particularly to the writer's mechanics. This was confirmed by the

above results; however, the authors did not expect a low score on “understanding and synthesis of argument” and “understanding the goals and implications of globalization”. Surprisingly, the results showed that the participants rated the second essay lower than the first essay even in these two categories.

Finally, and surprisingly, the results did not show a significant difference between education students and business/marketing students. The authors expected education students to be less influenced by the mechanics of the essay because of their familiarity with rubrics, fair assessment, and raters’ bias; however, the results did not support this hypothesis. This perhaps indicates that they did not have enough training. It also could be interpreted that a general familiarity with rubrics is not enough to change the reliability of assessment and there is a need for special training for using rubrics for a specific assignment. These findings not only shows low reliability for the rubric used in this study but also doubts its construct and criterion based validity.

7. Discussion and limitations

The reliability and validity of rubrics have been studied from several perspectives. One group has studied the objectivity of rubrics (Spandel, 2006; Breland, 1983; Wolfe, 1977; Coffman, 1971; McColly, 1970), and another group has claimed they are overly reductive (Kohn, 2006; Pula & Huot, 1993; Mabry, 1999; Flower & Hayes, 1981).

Some educators believe that rubrics have not led to a more objective or more reliable grading. Ideally, the feedback given by employment of a rubric is better than the assignment of a simple letter grade; however, untrained users of rubrics may simply use it to justify their biased assessment. According to this group, the irrelevant variables affecting holistic assessment of a performance or an essay may still affect rubric-based assessment. Judgments ultimately turn on adjectives that are murky and end up being left to the teacher’s discretion (Kohn, 2006; Lumley,

2002). The results of the present research confirmed this observation and showed that the raters graded the essay based on their overall impression rather than following the rubric. For example, we observed teachers gave points for citation in the essay that did not include any citations. We believe this happened because the raters were influenced by the overall quality of writing and so did not pay attention to this particular weakness.

However, some educators criticize rubrics due to their reductive nature (Kohn, 2006; Pula & Huot, 1993; Mabry, 1999; Flower & Hayes, 1981). This group argues that performance on an essay should be evaluated as a whole. In other words, a paper that scores high on isolated factors of good writing does not necessarily add up to good writing or vice versa. The authors think this might be a legitimate concern regarding rubrics; however, this concern applies more to composition or writing courses than written essays in other disciplines. For example, in this study (or for a social science professor) the overall quality of writing was not as important as responding to the essay prompts. Therefore, using a rubric in this study could not be considered a reductionist element. Furthermore, as Hamp-Lyons (1991) argues, sometimes, it is the holistic scoring that is reductive by reducing the writer's complex responses to a single score.

Although the results of this study were quite unexpected, the authors trusted them because the sample size was large and similar results were found across all four groups. There may be several reasons to justify these unexpected findings. Several limitations also make the project conducive to further study. Some possible explanations/limitations are listed here, and the authors agree there is a need for more investigation.

1. Virtually none of the participating raters in this study had received any specific training in the application of this particular rubric. Although the majority of raters in this study were teachers, and there was an expectation that they had been trained in effective use of rubrics as an assessment tool, few seemed to show an understanding of rubrics, as observed by the

authors. According to the literature, raters who receive appropriate rubric training are taught to focus on the language and components of the rubric, and to minimize or avoid personal judgments when grading (Knoch, Read, & Randow 2007; Dempsey, PytlikZillig, & Bruning 2009). Nevertheless, the effectiveness of rater training has not been studied enough and needs further investigation, particularly because some research studies suggested that the results of training might not endure for long after training session (Lumley & McNamara, 1995).

2. The authors used a different method to evaluate the reliability of the rubric. As mentioned earlier the inter-rater and intra-rater methods reported low to moderate rates of reliability. The limited number of researchers who studied the reliability of rubrics primarily used the “percentage of total agreement” or “percent of adjacent agreement” (within one score point) as a measurement probably because it is easy to calculate and it allows to use nominal data such as “pass-fail” or “not acceptable-acceptable” (Jonsson & Svingby, 2007). Inter-rater reliability is typically used to evaluate the consistency of the ratings.

As noted by Mabry (1999), consistency is not achieved because rubrics provide a vehicle for expressing naturally occurring agreement. On the other hand, Kohn (2006) pointed out that agreement among scorers is more easily achieved with regard to such matters as spelling and organization. Further, rubric scales (e.g., 1-5) artificially limit the scope of variability of scores. That is probably why Hamp-Lyons (1991) suggests using Spearman Brown’s prophecy formula, known as correction for attenuation, to correct for the artificial increase in reliability.

As mentioned earlier, this method of evaluating inter-rater reliability is more appropriate for a norm-referenced assessment. The correlation coefficient is directly influenced by the z-score (the standard difference between a score and the mean), therefore, use of this index only reflects the agreement on the relative ranks rather than the agreement on achieving criteria.

3. Raters often sacrifice validity for reliability, insight for efficiency, and authenticity for easier scoring (Wiggins, 1994). The analysis of the raters' comments and the authors' discussions with the raters indicated that they believed the rubric used in this study was too broad and needed to be more specific. Some educators have reported that rater reliability is substantially higher when more detailed descriptors are used (Knoch, 2009). The authors plan to incorporate a more specific rubric in subsequent research to investigate its impact on reliability; however, the authors wonder if making the rubric more specific might negatively impact the assessment by making it more reductive, and thus limiting the room for creativity and critical thinking in writing.

4. The goal of this study was to investigate how the surface features of writing influenced the raters' judgment of graduate students' writing performance. The results of this experiment should be examined very carefully. One should note that, in real life, the quality of writing and the knowledge of the writer about the topic are correlated to some extent. In other words, successful students are usually skilled writers as well. This experiment investigated an extreme situation where the quality of writing and the correctness of the answer were deliberately manipulated to have a negative correlation.

This situation is more likely to happen among minority language students (ESL students) who might understand the subject matter, but are not as skilled in writing style and conventions. As Hamp-Lyons (1991) explains in details, ESL writers' problems are not with lack of ideas or lack of knowledge, but with the means of expression for those ideas in English. Therefore, she argues that there are several reasons for developing separate writing assessment measures for first language and second language writers. Huang (2008) raised a potential question about the fairness of the writing scores assigned to ESL students. A non-native student who displays marginal command of English, for instance, may be scored lower on a written

assignment, but not in the same way a native speaker is evaluated. Hamp-Lyons (1991) explains that readers (raters) who are used to read reading in a midrange (typical L1 student writings) may overreact (underestimate) to a poorer paper, not being able to judge how much lower it is.

Researchers note that ESL students' errors differ markedly depending on the linguistic features of their first language. Their errors and/or discourse style might make it hard for the raters to understand ESL student's argument in the paper. This is particularly important when the raters are assessing an essay in academic programs other than English. In those cases the student may fully comprehend the course content; however, her/his grade may still suffer as a result of mechanical writing errors.

8. Conclusions

In conclusion, it appears that many teachers use rubrics without considering their reliability. Research indicates that at least some teachers believe decisions in the classroom, made on the basis of an assessment, can easily be changed if they appear to be wrong (Jonsson & Svingby, 2007). Many educators are now suggesting using portfolio assessment instead of a one-day or one-time (snapshot) writing sample (Elliot, 2005; East, 2006; Hamp-Lyons, 2002). As mentioned earlier, many teachers use rubrics simply because they believe using *any* rubric is better than assessing without a rubric. The authors understand the many benefits of using rubrics in assessment, and realize that rubrics should be well-designed, topic-specific (contextual), analytic, and complemented with exemplars to be effective. However, if a rubric like the one used in this project, which was designed by a group of professors in a college of education, is shown to be unreliable, then what does this say about the thousands of rubrics being used every day in schools? What does this say particularly about those rubrics downloaded from the Internet and implemented without any training? The unexpected results indicated that making a quality rubric, and using it effectively, are not as easy as one originally assumes. We learned that

rubrics should be developed locally for a specific purpose and a specific group of students. Like any tool, improper use is sometimes worse than not having used the tool at all. In the same way, using a rubric may not necessarily be better than not using one. The history of writing assessment shows that achieving high reliability in writing assessment is not easy, and we should be careful not to sacrifice validity to achieve higher rates of reliability.

References

- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3–13.
- Breland, H. M. (1983). *The direct assessment of writing skill: A measurement review*. New York: College Entrance Examination Board.
- Briggs, D. (1970). The influence of handwriting on assessment. *Educational Research*, 13, 50–55.
- Brookhart, S. M. (2005). The quality of local district assessments used in Nebraska's school-based teacher-led assessment and reporting system (STARS). *Educational Measurement: Issues and Practice*, 24 (2), 14 – 21.
- Bull, R., Stevens, J. (1979). The effects of attractiveness of writer and penmanship on essay grades. *Journal of Occupational Psychology*, 52, 53-59.
- Charney, D. A. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 18, 65-81.
- Chase, C. I. (1968). The Impact of some obvious variables on essay test scores. *Journal of Educational Measurement*, 5, 315-318.
- Coffman, W. E. (1971). On the reliability of ratings of essay examinations in English. *Research in the Teaching of English*, 5(1), 24-36.
- Dempsey, M. S., PytlikZillig, L. M. & Burning, R. H. (2009). Helping preservice teachers learn to assess writing: Practice and feedback in a web-based environment. *Assessing Writing*, 14(1), 38-61.
- East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing*, 11(3), 179-197.
- Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. New York: Peter Lang.
- Feldt, L. S. & Qualls, A. L. (1999). Variability in reliability coefficients and the standard error of measurement from school district to district. *Applied Measurement in Education*, 12(4) , 367–381.

- Flower, L., & Hayes, J. R. (1981, April). *Process-based evaluation of writing: Changing the performance, not the product*. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA.
- Follman, J. (1993). Critical thinking and verbal ability. *ACEHI Journal*, 19, 71-76.
- Gage, N. L., & Berliner, D. C. (1992). *Educational psychology*. Wentzville, MO: Borgasorus Books.
- Feldt, L. S. & Qualls, A. L. (1999). Variability in reliability coefficients and the standard error of measurement from school district to district. *Applied Measurement In Education*, 12(4), 367-381
- Finson, K. D. (1998). Rubrics and their use in inclusive science. *Intervention in School and Clinic*, 34(2), 79-88.
- Hafner, J. C. & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25 (12), 1509 - 1528.
- Hamp-Lyons, L. (1991). *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, 8 (1), 5-16.
- Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing*, 13, 201-218.
- James, H. (1927). The effect of handwriting on grading. *English Journal*, 16, 180-205.
- Johnson, R. L., Penny, J., Gordon, B. (2000). The relation between score resolution methods and inter-rater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, 13 (2), 121-138.
- Jonsson, A. Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(20), 275-304.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing* 12, 26-43.
- Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12-15.
- Lumley, T. (2002) 'Assessment criteria in a large-scale writing test: what do they really mean to the raters?' *Language Testing*, 19 (3), 246-276.
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12, 54-71.
- Mabry, L. (1999). Writing to the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan*, 80(9), 673-79.
- Malouff, J. (2008). Bias in grading. *College Teaching*, 56(3), 191-192.
- Markham, L. R. (1976). Influences of handwriting on teacher evaluation of written work, *American Educational Research Journal*, 13, 277-283.

- Marshall, J. C. & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, 6, 97-101.
- McColly, W. (1970). What does educational research say about the judging of writing ability? *Journal of Educational Research*, 64, 148-56.
- Meier, S. L., Rich, B. S., & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy and Practice*, 13(1), 69-95.
- Munoz, L.C., Frick, P.J., Kimonis, E.R., & Aucoin, K.J. (2008). Verbal ability and delinquency: Testing the moderating role of psychopathic traits. *Journal of Child Psychology and Psychiatry*, 49, 414-421.
- Newell, J. A., Dahm, K. D., Newell, H. L. (2002). Rubric development and inter-rater reliability: Issues in assessing learning outcomes. *Chemical Engineering Education*, 36(3), 212-15.
- Penny, J., Johnson, R.L., & Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, 68(3), 269-87.
- Pula, J. J., & Huot, B. A. (1993). A model of background influences on holistic raters. In M.M. Williamson & B.A. Huot (Eds.), *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations* (pp. 237-265). Cresskill, NJ: Hampton Press.
- Read, B., Francis, B., & Robson, J. (2005). Gender, bias, assessment and feedback: Analyzing the written assessment of undergraduate history essays. *Assessment and Evaluation in Higher Education*, 30(3), 241-260.
- Ross-Fisher, R. L. (2005). Developing effective success rubrics. *Kappa Delta Pi*, 41(3), 131-135.
- Scannell, D. P., & Marshall, J. C. (1966). The effect of selected composition errors on grades assigned to essay examinations. *American Educational Research Journal*, 3, 125-130.
- Sheppard, E. M. (1929). The effects of quality of penmanship on grades. *Journal of Educational Research*, 19, 102-105.
- Silvestri, L. & Oescher, J. (2006). Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education*, 9, 25-30.
- Skaalvik, E. M. & Skaalvik, S. (2004). Self-concept and self-efficacy: A test of the internal/external frame of reference model and predictions of subsequent motivation and achievement. *Psychological Reports*, 95(3), 1187-1202.
- Spandel, V. (2006). In defense of rubrics. *English Journal*, 96(1), 19-22.
- Tomkins, M. (2003). Trouble comes in threes. *Times Educational Supplement*, 4547, 23.
- Turley, E. D., & Gallagher, C. G. (2008). On the uses of rubrics: Reframing the great rubric debate. *English Journal*, 79(4), 87-92.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wiggins, G. (1994). The Constant Danger of Sacrificing Validity to Reliability: Making Writing Assessment Serve Writers. *Assessing Writing* 1(1), 129-39.
- Wilson, M. (2006). *Rethinking Rubrics in Writing Assessment*. Portsmouth: Heinemann.

Wilson, M. (2007). Why I won't be using rubrics to respond to students' writing. *English Journal*, 96 (4), 62-66.

Wolfe, E. W. (1977). The Relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, 4 (1), 83-106.

Appendix A

The Wrong Essay

The following essay was adopted from a sample essay written about globalization by professional writers from Social Science Research Council. However, this essay was not written to answer our experimental essay prompts. That is why we called it the Wrong Essay in this study. While it is well written it does not answer the questions in the essay prompt. The raters assumed this was written by a student in a course (ETEC 657). The source of this essay (retrieved from the Internet) is

http://essays.ssrc.org/sept11/essays/teaching_resource/tr_globalization.htm

ETEC 657 –Midterm Exam

Essay Question

Write an essay about globalization in which you explain;

- a- a brief history of globalization**
- b- its economical, social, and political impact**
- c- how information technology has influenced the speed of globalization**
- d- outsourcing and offshoring as the implications of globalization**

Although globalization is often thought of in economic terms (i.e., "the global marketplace"), this process has many social and political implications as well. Many in local communities associate globalization with *modernization* (i.e., the transformation of "traditional" societies into "Western" industrialized ones).

There are heated debates about globalization and its positive and negative effects. While globalization is thought of by many as having the potential to make societies richer through trade and to bring knowledge and information to people around the world, there are many others who perceive globalization as contributing to the exploitation of the poor by the rich, and as a threat to traditional cultures as the process of modernization changes societies. There are some who link the negative aspects of globalization to terrorism. To put a complicated discussion in simple terms, they argue that exploitative or declining conditions contribute to the lure of informal "extremist" networks that commit criminal or terrorist acts internationally. And thanks to today's technology and integrated societies, these networks span throughout the world.

Increasingly over the past two centuries, economic activity has become more globally oriented and integrated. Some economists argue that it is no longer meaningful to think in terms of national economies; international trade has become central to most local and domestic economies around the world. Economists project that, in the U.S., more than 50 percent of the new jobs created in this decade will be directly linked to the global economy.

The recent focus on the international integration of economies is based on the desirability of a free global market with as few trade barriers as possible, allowing for true competition across borders.

International economic institutions, such as the World Trade Organization (WTO) and the International Monetary Fund (IMF), facilitate this increasingly barrier-free flow of goods, services, and money (capital) internationally. Regionally, too, organizations like the North America Free Trade Association (NAFTA), the European Union (EU), and the Association of South East Asian Nations (ASEAN) work towards economic integration within their respective geographical regions.

Many economists assess economic globalization as having a positive impact, linking increased economic transactions across national borders to increased world GDP, and opportunities for economic development. Still, the process is not without its critics, who consider that many of the economies of the industrial North (i.e., North America, Europe, East Asia) have benefited from globalization, while in the past two decades many semi- and non-industrial countries of the geopolitical South (i.e., Africa, parts of Asia, and Central and South America) have faced economic downturns rather than the growth promised by economic integration. Critics assert that these conditions are to a significant extent the consequence of global restructuring which has benefited Northern economies while disadvantaging Southern economies. Others voice concern that globalization adversely affects workers and the environment in many countries around the world.

Though there are many social and cultural manifestations of globalization, here are some of the major ones:

- **Informational services:** On the one hand, the electronic revolution has promoted the diversification and democratization of information as people in nearly every country are able to communicate their opinions and perspectives on issues. On the other hand, this expansion of information technology has been highly uneven, creating an international "digital divide" (i.e., differences in access to and skills to use Internet and other information technologies due predominantly to geography and economic status). Often, access to information technology and to telephone lines in many developing countries is controlled by the state or is available only to a small minority who can afford them.
- **News services:** In recent years there has been a significant shift in the transmission and reporting of world news with the rise of a small number of global news services. This process has been referred to as the "CNN-ization of news," reflecting the power of a few news agencies to construct and disseminate news. Thanks to satellite technology, CNN and its few competitors extend their reach to even the most geographically remote areas of the world. This raises some important questions of globalization: Who determines what news? What is "newsworthy?" Who frames the news and determines the perspectives articulated? Whose voice(s) are and are not represented? What are the potential political consequences of the silencing of alternative voices and perspectives?
- **Popular culture:** The contemporary revolution in communication technology has had a dramatic impact in the arena of popular culture. Information technology enables a wide diversity of locally-based popular culture to develop and reach a larger audience. For example, "world music" has developed a major international audience. Old and new

musical traditions that a few years ago were limited to a small local audience are now playing on the world stage.

On the other hand, globalization has increased transmission of popular culture easily and inexpensively from the developed countries of the North throughout the world. Consequently, despite efforts of nationally-based media to develop local television, movie, and video programs, many media markets in countries of Africa, Asia, and Latin America are saturated with productions from the U.S., Europe and a few countries in Asia (especially Japan and India). Local critics of this trend lament not only the resulting silencing of domestic cultural expression, but also the hegemonic reach of Western, "alien" culture and the potential global homogenization of values and cultural taste.

Appendix A

The Correct Essay

The following essay was written by the authors using some online resources. The essay was written to answer all 4 parts of the essay prompt correctly. It was well organized and the references and citations were included. That is why we called it the Correct Essay in this study. However, the authors deliberately inserted 20 spelling and grammar errors in it to see how these errors will affect raters' assessment of the essay. The raters assumed this was written by a student in a course (ETEC 657).

ETEC 657 –Midterm Exam

Write an essay about globalization in which you explain;

- e- a brief history of globalization
- f- its economical, social, and political impact
- g- how information technology has influenced the speed of globalization
- h- outsourcing and offshoring as the implications of globalization

Answer

a- a brief history of globalization

I know that early forms of globalization existed during the Roman Empire, the Arab Empire and Islamic Golden Age, when Muslim traders and explorers established an early global economy across the Old World resulting in a globalization of crops, trade, knowledge and technology; and later during the Mongol Empire, when there was greater integration along the Silk Road. Global integration continued through the expansion of European trade, as in the 16th and 17th centuries, when the Portuguese and Spanish Empires reached to all corners of the world after expanding to the Americas. I should say this Globalization became a business phenomenon in the 17th century when the Dutch East India Company, which is often described as the first multinational corporation, was established. Because of the high risks involved with international trade, the Dutch East India Company became the first company in the world to share risk and enable joint ownership through the issuing of shares: an important driver for globalization. (Harvey, 2005).

Some say Globalization in the era since World War II was first the result of planning by economists, and politicians who recognized the costs associated with protectionism and declining international economic integration. Their work led to the Bretton Woods conference and the founding of several international institutions intended to oversee the renewed processes of globalization, promoting growth and managing adverse consequences. These were the International Bank for Reconstruction and Development (the World Bank) and the International Monetary Fund. It has been facilitated by advances in technology which have reduced the costs of trade, and trade negotiation rounds, originally under the auspices of GATT, which led to a series of agreements to remove restrictions on free trade. The Uruguay Round (1984 to 1995) led to a treaty to create the World Trade Organization (WTO), to mediate trade

disputes and set up a uniform platform of trading. Other bi- and trilateral trade agreements, including sections of Europe's Maastricht Treaty and the North American Free Trade Agreement (NAFTA) have also been signed in pursuit of the goal of reducing tariffs and barriers to trade grand. (Sachs, 2005).

b- its economical, social, and political impact

Globalization has Very Many aspects which affect the world in several different ways such as:

Industrial – emergence of worldwide production markets and broader access to a range of foreign products for consumers and companies. Financial - emergence of worldwide financial markets and better access to external financing for corporate, national and subnational borrowers. Economic - realization of a global common market, based on the freedom of exchange of goods and capital. Spread of local consumer products (e.g. food) to other countries (often adapted to their culture). (Capra, 2002).

Political – As our teachers said globalization is the creation of a world government which regulates the relationships among nations and guarantees the rights arising from social and economic globalization. Increase in the number of standards (rules & laws) applied globally; e.g. copyright laws, patents and world trade agreements. The push by many advocates for an international criminal court and international justice movements.

Informational - increase in information flows between geographically remote locations. Development of a global telecommunications infrastructure and greater transborder data flow, using such technologies as the Internet, communication satellites, submarine fiber optic cable, and wireless telephones. Cultural - growth of cross-cultural contacts; advent of new categories of consciousness and identities such as Globalism - which embodies cultural diffusion, the desire to consume and enjoy foreign products and ideas, adopt new technology and practices, and participate in a "world culture". Ecological- the advent of global environmental challenges that can not be solved without international cooperation, such as climate change, cross-boundary water and air pollution, over-fishing of the ocean, and the spread of invasive species. Many factories are built in developing countries where they can pollute freely. Social - the achievement of free circulation by people of all nations. Spreading of multiculturalism, and better individual access to cultural diversity (e.g. through the export of Hollywood and Bollywood movies). However, the imported culture can easily supplant the local culture, causing reduction in diversity through hybridization or even assimilation. The most prominent form of this is Westernization, but Sinicization of cultures has taken place over most of Asia for many centuries.

Transportation - fewer and fewer European cars on European roads each year (the same can also be said about American cars on American roads) and the death of distance through the incorporation of technology to decrease travel time. Greater international travel and tourism. Greater immigration, including illegal immigration. (Croucher, 2004).

c- how information technology has influenced the speed of globalization

I no wonder that the 1990s witnessed the emergence of new information technologies that have had a substantial impact on both commerce and society in general. Digital technologies have opened the way towards global networks. Global networks are the networks in which all

information and knowledge – also the ideology- necessary for the realization, maintenance and the reproduction of the system – basically the capitalist system. The term “New Economy” is the clearest explanation of how all these information, knowledge and ideology are in close relation to capitalism. <http://mediaif.emu.edu.tr/pages/atabek/GCS7.html>

Given that approximately 50 percent of economic production in OECD countries is now generated by knowledge-based industries, it is no surprise that advanced telecommunications are increasingly viewed as requisites for economic and human development (Crenshaw & Robinson, 2006). It is believed that networking of distributed computing systems not only reduce costs, but also improve the efficiency of resource acquisitions. Since the information and communication technology knows no boundary, it might be also accessed, theoretically, by every ventures using Internet. On the other hand, commerce on the Internet opens not only new forms of trade relationships among world trade participants, but it also restructures the whole market system, mainly as electronic market system, thereby such a system could certainly give opportunities for small firms specially those of third world countries to enter the market, which otherwise was hardly possible. The new information technology redefines the relationship between buyer, seller and middleman, allowing new ways of accessing and tapping information, and price arrangements. The information and communication development in the developing countries has given many positive external effects to the third world countries. No doubt that the information technological revolution has reached African countries too. World-wide fads and pop culture such as Pokémon, Sudoku, Numa Numa, Origami, Idol series, YouTube, Orkut, Facebook, and MySpace. World-wide sporting events such as FIFA World Cup and the Olympic Games are just some examples. Formation or development of a set of universal values. http://www.uis.unesco.org/template/pdf/cscl/IntlFlows_EN.pdf

d- outsourcing and offshoring as the implications of globalization

Off course technology has provided new opportunities for globalization of economy and international trade. Offshoring is defined as the movement of a business process done at a company in one country to the same or another company in another country. Production offshoring of established products involves relocation of physical manufacturing processes to a lower-cost destination. Examples of production offshoring include the manufacture of electronic components in Taiwan, production of apparel, toys, and consumer goods in China, Vietnam etc. Almost always work is moved due to a lower cost of operations in the new location. Offshoring is sometimes contrasted with outsourcing or offshore outsourcing. Outsourcing is the movement of internal business processes to an external company. Companies subcontracting in the same country would be outsourcing, but not offshoring. A company moving an internal business unit from one country to another would be offshoring, but not outsourcing. A company subcontracting a business unit to a different company in another country would be both outsourcing and offshoring. (Hunter, 2001).

Conclusion

I don't surprise that supporters of free trade claim that it increases economic prosperity as well as opportunity, especially among developing nations, enhances civil liberties and leads to a more efficient allocation of resources. One of the surprisings of the recent success of India and China

is the fear that success in these two countries comes at the expense of the United States. These fears are fundamentally wrong and, even worse, dangerous. Globalization advocates such as Jeffrey Sachs point to the above average drop in poverty rates in countries, such as China, where globalization has taken a strong foothold, compared to areas less affected by globalization. (Sachs, 2005).

In the other hand critiques of the current wave of economic globalization typically look at both the damage to the planet, in terms of the perceived unsustainable harm done to the biosphere, as well as the perceived human costs, such as increased poverty, inequality, injustice and the erosion of traditional culture which, the critics contend, all occur as a result of the economic transformations related to globalization. They challenge directly the metrics, such as GDP, used to measure progress promulgated by institutions such as the World Bank, and look to other measures, such as the Happy Planet Index (an index of human well-being and environmental impact, designed to challenge well-established indices of countries' development, such as (GDP) and the Human Development Index (HDI). In particular, GDP is seen as inappropriate, as the ultimate aim of most people is not to be rich, but to be happy and healthy and it is critical to understand what effect the pursuit of those goals has on the environment. They believe most people want to live long and fulfilling lives, and the country which is doing the best is the one that allows its citizens to do so. (United Nations Development Program, 1992).

References

Croucher, S. L. (2004). *Globalization and Belonging: The Politics of Identity in a Changing World*. NY: Rowman & Littlefield.

Sachs, J. (2005). *The End of Poverty*. NY: The Penguin Press.

http://www.uis.unesco.org/template/pdf/cscl/IntlFlows_EN.pdf

<http://mediaif.emu.edu.tr/pages/atabek/GCS7.html>

Capra, F. (2002). *The Hidden Connections*. NY: Random House.

Hunter, W. R. (2001). The Rising Inequality of World Income Distribution. *Finance & Development*, Vol 38,(4), 1-4.

United Nations Development Program (1992). *Human Development Report*. Oxford: Oxford University Press.

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford UP.

Appendix C- Grading Rubric

	Not Passing	Not Passing	Passing	Exceptional	Your Score
Structural Organization and Clarity 25 points	There is no clear purpose; Essay lacks logical progression of ideas; Essay addresses topic but loses focus by including irrelevant ideas; Ideas are unclear and/or not well-developed	Attempts to communicate the purpose throughout; Essay includes brief skeleton (introduction, body, conclusion) but lacks transitions; Essay is focused on topic and includes few loosely related ideas; Unelaborated ideas that are not fully explained or supported; repetitive details	Generally maintains purpose; Essay includes logical progression of ideas aided by clear transitions; Essay is focused on the topic and includes relevant ideas; Depth of thought supported by elaborated, relevant supportive evidence provides clear vision of the idea; contains details	Establishes and maintains clear purpose; Essay is powerfully organized and fully developed; The essay is focused, purposeful, and reflects clear insight and ideas; Depth and complexity of thought supported by rich, pertinent details; supporting evidence leads to high-level idea development	
	<i>0-6</i>	<i>7-13</i>	<i>14-19</i>	<i>20-25</i>	
Understanding and Synthesis of material: Argument 25 points	Apparent misunderstanding of material; Lack of confidence with subject matter which leads to unconvincing argument	Limited understanding of material displayed by vague, unclear language; Some confidence with material; does not present a convincing argument	Developing understanding of material; Confidence with most material, thus presenting fragmented argument	Clear understanding of material displayed by clear, concrete language and complex ideas; Confidence with all material which leads to strong, convincing, consistent argument	
	<i>0-6</i>	<i>7-13</i>	<i>14-19</i>	<i>20-25</i>	
Understanding the goals and implications of globalization 25 points	Demonstrates a lack of knowledge about the history of globalization, its implications, the	Demonstrates a little knowledge about the history of globalization,	Demonstrates a general knowledge about the history of globalization,	Demonstrates explicit and extensive knowledge about the history of globalization, its	

	role of information technology, outsourcing and offshoring.	its implications, the role of information technology, outsourcing and offshoring.	its implications, the role of information technology, outsourcing and offshoring.	implications, the role of information technology, outsourcing and offshoring. Promotes engagement and demonstrates a deeper conceptual understanding of key concepts. Critically discusses the pros and cons of globalization.	
	0-6	7-13	14-19	20-25	
Support and Citing Sources 15 points	Few to no solid supporting ideas or evidence for the essay content; Little to no source citation, or inaccurate citations with no adherence to standard format; difficult to follow; No references or incorrect references	Some supporting ideas and/or evidence for the essay content; Some source citation but somewhat inaccurate; no adherence to standard format; difficult to follow; Few references or some incorrect references	Support lacks specificity and is loosely developed; Mostly accurate source citation according to standard format; Use of references indicate some research	Specific, developed details and superior support and evidence in the essay content; Cites sources accurately and according to standard format; person who engages product will easily be able to access sources based on citation; Use of references indicate substantial research	
	0-3	4-7	8-11	12-15	
Mechanics 10 points	Frequent errors in spelling, grammar, and punctuation	Errors in grammar and punctuation, but spelling has been proofread	Occasional grammatical errors and questionable word choice	Nearly error-free which reflects clear understanding and thorough proofreading	
	0-2	3-5	6-8	9-10	