The Effect of Front-loaded Expectations on Cumulative Assessment Scores

---

A Special Project

Presented to

Dr. Gordon Martinen

Heritage University

In Partial Fulfillment

Of the Requirement for the Degree of

Master of Education

Sammi L. Muecke

2010

FACULTY APPROVAL

The Effect of Front-loaded Expectations on Cumulative Assessment Scores

Approved for the Faculty

_____, Faculty Advisor

_____, Date

ABSTRACT

The research project was intended to determine the effects of front-loaded expectations on scores achieved on a cumulative assessment. Literature was read regarding the use and correct construction of effective rubrics, the theory of multiple intelligences and its effect on student learning, and Differentiated Instruction (DI) as it applied to student learning.

The study was conducted in a high school in Southwest Washington State. Student population consisted of four first year Spanish classes, two in the control group and two in the treatment group. A general rubric was used as an assessment guide for students in the treatment group.

Values were entered using the Statpak (Gay, Mills & Airasian 2003) and data were analyzed. Scores were compared and statistics were entered to determine what effect, if any, front-loading had on test scores. The scores were entered using the independent *t* test and significance was sought at $p \geq .05, .01, .001$ levels.

PERMISSION TO STORE

I, Sammi L. Muecke, hereby irrevocably consent and authorize Heritage University Library to file the attached Special Project entitled, *The Effect of Front-loaded Expectations on Cumulative Assessment Scores,* and make such Project and Compact Disk (CD) available for the use, circulation and/or reproduction by the Library. The Project and CD may be used at Heritage University Library and all site locations.

I state at this time the contents of this Project are my work and completely original unless properly attributed and/or used with permission.

I understand that after three years the printed Project will be retired from the Heritage University Library. My responsibility is to retrieve the printed Project and, if not retrieved, Heritage University may dispose of the document. The Compact Disc and electronic file will be kept indefinitely.

_____, Author

_____, Date

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Page

CHAPTER 1

Introduction

"What is a grade? A grade is an inadequate report of an imprecise judgment of a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown portion of an indefinite amount of materials." Paul Dressel

"I want my children to understand the world, but not just because the world is fascinating and the human mind is curious. I want them to understand it so that they will be positioned to make it a better place…An important part of that understanding, is knowing who we are and what we can do." (Howard Gardner, 1999, pp.180-181)

"A different way to learn is what kids are calling for….All of them are talking about how our one-size-fits-all delivery system- which mandates that everyone learn the same thing at the same time, no matter what their individual needs-has failed them." Seymour Sarason, (The Predictable Failure of Educational Reform.)

Background for the Project

Students at all levels of education were expected to increase information and abilities possessed by fulfilling criteria. Usually the criteria in question were known only to the instructor(s) while students rarely had opportunity to obtain prior knowledge of expectations until after the learner did or did not fulfill said expectations. Students often failed to reach goals to which they were not privy.

Statement of the Problem

In the view of the author, teachers in training were taught to construct rubrics for assessment. However, beginning teachers were not trained in the use of parallel language, front-loaded expectations or student input regarding what criteria were assessed. There existed a considerable gap between how teachers were trained and the realities of the classroom. Additionally, no real thought was given to the vast differences which existed with regards to individual student ability. There was vague mention of disparity between individuals but no more than a quick acknowledgement that differences existed. Students, on the other hand, were very aware that inequality in the classroom not only existed, it was the primary constant. Even very young children knew that true equality was unjust, unrealistic and unachievable. Equity, on the other hand was possible, indeed vital, to the competent management of classrooms. Only through constant vigilance on the part of the instructor and students did equitable and just treatment of students exist. Especially at the secondary levels and into college, the impression was fostered that the needs of the individual were of little or no importance, which in turn led to apathy and disinterest on the part of instructors as well as students. The author held that apathy and disinterest led to poor scores on assessments as well as low numbers of successful students.

Purpose of the Project

The project was intended to discover if front-loading of expectations at the beginning of the learning process engendered interest in self-improvement as demonstrated by higher scores on formal assessments. If students were given the desired

outcomes and expectations, would the students be more likely to reach the goals as outlines?

Delimitations

Population for research was necessarily limited to a convenience sample which consisted of first-year, non-native speakers of Spanish in a high school class. Participants were required to be present in class a minimum of 75 of the first 90 school days. Class size was between 25-30 students.

Population for control group consisted of students from the academic year, 2008/2009. Number grades based on points received for examinations were compared with post-treatment grades of students from the academic year 2009/2010 for the same period of time, September 1st to November 14th, as the scores from 2008/2009. The control group included students from two separate first year, non-native students. The two classes, in total, contained 33 boys, 24 girls. Two boys and one girl were removed from the group because they each attended less than 75 days of the first 90 days of classes. The test was administered on the same day in November of 2008 to both classes of the control group.

Treatment was given to students of the 2009/2010 sample beginning with a teacher generated rubric that allowed students to familiarize themselves with rubric language and usage. All students were given detailed explanations and opportunities to practice rubric application. All terms and vocabulary were defined and explained for maximum comprehension. The test group included two first year, non-native classes

which, in total consisted of 21 boys, 32 girls. Three of the boys and one girl were removed from the test group due to non-participation in the test. The same test was given to the experimental group as the control group. It was also administered on the same day in November to both classes of the treatment group.

Assumptions

Assumptions were made that some students were able to comprehend spoken Spanish as the population of the school consisted of approximately 68% Hispanic students. There was a necessary distinction made between bilingual and bi-literate students. Students who were able to converse in Spanish but were unable to read or write in Spanish were included. All students in the class whether non-Spanish speakers, bilingual or bi-literate received the treatment and the researcher later removed the students the author determined to be bilingual as well as bi-literate. Students included in the study were assumed to possess at minimum eight years of formal education and to be able to perform all skills necessary to be considered at the ninth grade level or higher dependent on the grade level of the individual student. All classes in both the treatment group and the control group consisted of students from ninth through twelfth grades.

Research Hypothesis

If trained in the use of rubrics before being assessed students were able to achieve significantly higher scores than students who were not so trained. Front-loading expectations would allow students to be aware of expectations and keep the goal in mind before beginning a project, activity or assessment.

Null Hypothesis

If students were not given training in rubric use the students would not receive equal or significantly higher scores that the students who were trained in rubric use or front-loaded the expectations. Significance was determined for $p \geq .05, .01, .001$.

Significance of the Project

Federal and state law required students to achieve increasingly higher scores in mathematics, science, reading and writing. Students educated in the knowledge and skills to self-monitor and self- assess were equipped to transfer such skills in Spanish to other areas of curriculum. Through use of rubrics with clear, parallel language, students and instructors were prepared to measure individual and class-wide progress toward unambiguous expectations.

Procedure

Student scores from the first quarter, September 1, 2008 to November, 14, 2008 were compared in a quasi-experimental, non-equivalent control group study to the scores of students for the same level of curriculum for the same amount of time in the 2009/2010 school year. Scores from 2008 reflected student achievement with no front-loading of expectations or student training regarding scoring criteria. The scoring criteria were not given to the students before the assignment, assessment or activity. The scores students received were then compared to the scores of students at the same level, during the corresponding period of time September 1, to November 4, 2009. Teacher-produced rubrics were front-loaded and then used for assessment during the first quarter of 2009

for comparable activities, assignments or assessments. Using an independent *t* test, the two sets of scores were compared.

Definition of Terms

assessment. Criteria, examinations, quizzes or activities used to measure student knowledge, abilities and/or performance.

bench mark. Standard the learner reached in order to be considered sufficiently competent to continue to the next level of education.

bilingual. Students able to speak, read, write and comprehend English as well as comprehend spoken Spanish and respond in Spanish but unable to read or write in Spanish.

bi-literate. Students able to comprehend, speak, read and write in Spanish as well as English.

differentiated instruction. Instruction that was adjusted according to the specific needs, preferences and/or abilities of students.

double-barrel questions. Questions which called for the rater to assess multiple tasks or characteristics at the same time.

front-load. Expectations and/or criteria were given to students before students were assessed with those expectations or criteria.

general use rubric. Rubric which was applicable and transferable to many activities, assignments and tasks.

intelligence. Preferred method in which an individual acquired and applied knowledge and information.

Likert scale. An ordered, one-dimensional scale of the number of times a specific behavior occurs in an assignment.

parallel language. Wording in categories of rubrics or other assessment tools, which was alike, allowing for clarity and consistency.

rubric. A written instrument of assessment which included criteria and measurable expectations for completion of a given assignment, project or activity.

Acronyms

DI. Differentiated Instruction

MI. Multiple Intelligence

CHAPTER 2

Review of Selected Literature

Introduction

Traditionally, students were the passive receivers of knowledge while teachers were the fountain from which that knowledge flowed. Students, by tradition, were considered unequipped to fill the role of educator and as such, had little or no say in the management of his/her learning.

Such thinking was in the process of being altered by researchers and scientists such as Carol Ann Tomlinson, Howard Gardner and others. These researchers maintained that students who knew what was coming were better prepared than students who were taught in the traditional manner. The purpose of this project was to test the theory that students were more successful if prepared in advance by the front-loading of expectations.

Differentiated Instruction

All students were not created equal. Each one possessed a unique manner with which to view the world, acquire information and apply skills using that information.

Carol Ann Tomlinson, considered one of the pioneers of Differentiated Instruction (DI), held that learners possessed distinct abilities, preferences and skills for the acquisition and assimilation of new information. Further, Tomlinson maintained that learners were more likely to retain target concepts and information if allowed to implement personalized instructional methods. Tomlinson stated: “Even though students

may learn in many ways, the essential skills and content they learn can remain steady. That is, students can take different roads to the same destination." (Tomlinson, 2000, p. 12) It remained the responsibility of instructors to establish methods and means by which students of distinct abilities and preferences were to achieve target bench marks. Tomlinson questioned the validity of single textbook adoption as sending 'inaccurate messages about the sameness of all learners' and established the school of thought which encouraged teachers to foster differences rather than ignore them.' (Tomlinson, 2000, p. 12)

In order for DI to function correctly it was first necessary to reach the understanding that "(teachers) have to know where we want to end up before we start out- and plan to get there. That is, we must have solid curriculum and instruction in place before we differentiate them." (Tomlinson, 2000, p. 12)

While most instructors instinctively knew that learners were different and therefore had distinct requirements and abilities, those same instructors lacked formal training in the how-to of equitable instruction and assessment for such diverse students. If students were to be treated equally, such distinctions could not be made. If instructors maintained equitability rather than equality, differentiation was a must. "To make differentiation work- in fact, to make teaching and learning work- teachers must develop an alternative approach to instructional planning beyond 'covering the text' or 'creating activities that student will like' " (Tomlinson, 2000, p. 13). While a teacher may have needed to teach key concepts during a specific period of time, it did not serve student

needs to simply introduce, discuss and then administer assessments in the same manner to all. If the goal was to ensure students retained information and skills, it was vital that students ingested, digested and integrated said concepts and information in a manner that best suited the individual's preferences for information acquisition, development and retention.

> Rick Wormeli stated, Differentiated Instruction is doing what is fair for students. It's a collection of best practices strategically employed to maximize student learning at every turn including giving them tools to handle anything that is undifferentiated. It requires us to do different things for different students some, or a lot, of the time in order for them to learn when a general classroom approach does not meet student needs. It is not individualized instruction though that may happen time to time as warranted. It's whatever works to advance the students. It's highly effective teaching.
>
> If we accept this premise then every aspect of our teaching, including grading and assessment practices, should be fair to students, and should maximize the students' learning. Anything that does not adequately provide for such is suspect. (Wormeli, 2006, pp3-4)

Multiple Intelligences

Dr. Howard Gardener of Harvard University developed the theory of multiple intelligences in 1983. Defining 'intelligence' as 'the capacity to solve problems or to

fashion products that are valued in one or more cultural settings', Gardener maintained the existence of at least seven:

> Linguistic intelligence involves sensitivity to spoken and written language, the ability to learn languages, and the capacity to use language to accomplish certain goals. This intelligence includes the ability to effectively use language to express oneself rhetorically or poetically and language as a means to remember information. Writers, poets, lawyers and speakers are among those Howard Gardener sees as having high linguistic intelligence.
>
> Logical-mathematical intelligence consists of the capacity to analyze problems logically, carry out mathematical operations, and investigate issues scientifically. In Howard Gardener's words, it entails the ability to detect patterns, reason deductively and think logically. This intelligence is most often associated with scientific and mathematical thinking.
>
> Musical intelligence involves skills in the performance, composition, and appreciation of musical patterns, it encompasses the capacity to recognize and compose musical pitches, tones and rhythms. According to Gardener musical intelligence runs in an almost structural parallel to linguistic intelligence.

Bodily-kinesthetic intelligence entails the potential of using one's whole body or parts of the body to solve problems. It is the ability to use mental abilities to coordinates bodily movements. Howard Gardener sees mental and physical activity as related.

Spatial intelligence involves the potential to recognize and use the patterns of wide space and more confined areas.

Interpersonal intelligence is concerned with the capacity to understand the intentions, motivations and desires of other people. It allows people to work effectively with others. Educators, salespeople, religious and political leaders and counselors all need a well-developed interpersonal intelligence.

Intra-personal intelligence entails the capacity to understand oneself, to appreciate one's feelings, fears and motivations. In Howard Gardener's view, it involves having an effective working model of ourselves, and to be able to use such information to regulate our lives. (As cited by Smith, 2002, 2008)

Gardener later concluded that there existed an additional intelligence which needed to be accounted for:

"Naturalist intelligence enables human beings to recognize, categorize and draw upon certain features of the environment. It 'combines a description of the core ability

with a characterization of the role that many cultures value.'" (as cited by Smith, 2002, 2008)

Dr. Gardener held that our schools and culture focused most attention toward linguistic and mathematical-logical intelligences while, for the most part, the remaining six intelligences were ignored. This led to inequity when educating most of society and placed all other types of intelligences at a disadvantage. Gardener stated: "many of these kids, in fact, end up being labeled 'learning disabled', 'ADD' (attention deficit disorder), or simply underachievers, when their unique ways of thinking and learning aren't addressed by a heavily linguistic or logical mathematical classroom." (Gardner, 1987 as cited by Thomas Armstrong)

The basic premise behind Gardener's theory of multiple intelligences (MI) indicated an individual was more likely to acquire, retain and apply information or skills when allowed to use his or her specific intelligence. Adults in the work force were happier and more productive if they held a position which allowed them to use their preferred intelligence rather than a position that did not. It followed, then, that students who were encouraged to focus on target concepts and information through the individual's personal brand of intelligence would enjoy the learning process, became more deeply engaged in knowledge acquisition and retain far more than if forced to acquire, assimilate and manipulate information and concepts using intelligence that fit others.

The use of a general rubric such as the rubric used in the research project was to establish a baseline for student achievement and growth. Each student, according to Howard Gardener, possessed different preferences for the acquisition of new knowledge and skills. The general rubric applied in the study was intended to measure students in a more individualized way which allowed the researcher to measure individual student growth compared only to the student under scrutiny. In this manner, student growth was measured on personal growth and development. While all students were held to the same high standard, the starting place of each student was considered when the student's daily work was assessed. The use of Marzano's general use rubric allowed students to become familiar with the expectations for performance and what was considered quality, acceptable work. The use of the rubric for daily assignments and formative assessments permitted the students to have a graphic demonstration of what constituted quality work. Though not assessed with the rubric on the final assessment, students were never-the-less thoroughly familiar with the expectations. Appendix A included the general use rubric created by Robert J. Marzano.

Tomlinson and Gardener agreed that students must acquire specific knowledge, concepts and abilities. What was more, each student was unique, possessed strengths and weaknesses which shaped the world for that individual. If students were to be able to accomplish a simple task, and know certain facts by the end of a given segment in time, it did not matter how one acquired or applied the necessary knowledge. On the other hand, if the goal was for the student to understand the information and have the ability to apply

skills to other, unconnected tasks, it became imperative that students were allowed to work toward their strengths, using the intelligence by which they viewed the world.

Tomlinson wrote,

> The core of *what* the students learn remains relatively steady. *How* the student learns- including degree of difficulty, working arrangements, modes of expression, and sorts of scaffolding- may vary considerably. Differentiation is not so much the "stuff" as the "how". If the "stuff" is ill conceived, the "how" is doomed. (Tomlinson, 2000, p. 16)

Use of Rubrics to Aid in Assessment

"Rubrics, scoring guides, and performance criteria describe what to look for in products or performances to judge their quality." (Arter, 2000, p.3) When quality was judged, there necessarily existed a device or method of measurement. Provision of scoring criteria, rubrics or guidelines provided such a device.

Each of the following articles of this subtopic focused on the various uses and application of rubrics. Each met with varied levels of success, depending on the intended purpose for the use of the rubric.

The intent of the researcher was to discover a wide variety of populations as well as purposes, for which rubrics were used. The researcher read existing research on construction and application of various rubrics and was given a wide spectrum of successes and failures which allowed the author to focus the project.

Programs from a nursing college, an engineering school and an elementary reading program all were included in the research of literature. Each program met with measurable results that demonstrated successful application at all levels.

Reid and Cooney explained the use of a variety of rubrics. Time was taken to define characteristics of effective rubrics as well as the necessity for clarity, brevity, and adaptability. Some characteristics of effective rubrics included:

> Language that is understandable to the learner and teacher,
>
> Terms which are clearly defined and measurable,
>
> Descriptions encourage a 'continuous improvement' mindset (indicate what can be done to improve),
>
> Avoid double-barrel questions (questions that ask the rater to assess multiple characteristics at one time),
>
> Avoid duplication of questions. (Reid, 2008, p. 893)

Reid and Cooney further explained that once the assessment instrument was developed it could be used for a considerable period of time. The rubrics aided in improvement of instruction and were intended to be adapted for other programs. Users of the rubrics had to pay particular attention to the requirements of each specific task and ensure that requirements defined in the rubric fit expectations and objectives of the task being assessed. "Successful implementation requires attention to the specific program objectives rather than simply adopting rubrics designed as part of other programs." (Reid and Cooney, 2008, p.893)

One type of rubric was the survey. Surveys were difficult to use to generate data as surveys were an "indirect measure of student performance" (Reid, 2008, p. 894) and therefore were used only as one part of the assessment plan. Students used the survey to report on their own mastery of target concepts. Self-reporting was necessarily subjective so not a true measure of student learning.

Standardized tests were a more accurate measure of actual student learning as standardized tests were not used to measure one's opinion of what the individual learned. Rather, specific criteria or standards were or were not met. These standards or criteria became the measure of student learning or lack thereof. However, other, less accurately measured but necessary skills, needed to be assessed as well. Accurate and clear rubrics were constructed to measure such skills as time usage, ability to work in a team, design and application of components, systems and software. All such skills were necessary to the program being taught and so needed instruments of measurement to determine student success at attaining these skills. Skills such as team building and function were considered a 'soft skill' and so were not included in a final exam yet were needed components in the students' repertoire. It followed, then, the students' level of ability in soft skills needed to be assessed in some manner. The use of written rubrics allowed students to be assessed in the soft skills and supplied opportunity and focus for self-improvement. "These rubrics also enable an instructor to provide feedback to the student, thus allowing the student to improve." (Reid and Cooney, 2008, p. 896)

Rubrics used for the case study of Reid and Cooney "were not designed to be used for grading, although this is certainly a common use." (Reid and Cooney, 2008, 896) Each criterion could have been weighted as desired, if they were used in grading.

The rubric used for Reid and Cooney's case study of 2008 was agreed upon by faculty committee. Staff used the rubrics, such as the one on the following page, to assess each area on a number scale from 5 (excellent) to 1 (poor) while the specific wording varied depending on the course to which the rubric was applied.

Table 1. ECET Design project assessment rubric

To be completed by the instructor regarding each project, may be individual students or a team design
Used to evaluate ABET items d & k:

- Apply creativity in the design of systems, components or processes appropriate to program objectives
- Have a commitment to quality, timeliness and continuous improvement.

May be used in student grading, but this is not required.

	Excellent	Average	Poor	ex…ave…poor
Identification of Problem or Definitio of Project	Clear and complete ID of design goals &objectives.	Adequate ID of problem: any lack of specifics does not impair solution or design.	Insufficient ID of problem, inadequately id's objective.	
Technical design	Exceeds specs if appropriate; meets specs with efficient design.	Meets nearly all specs.	Missing significant specs.	
Complexity of project design.	Exceeds typical technical complexity for course level	Meets typical technical complexity for course level.	Below typical technical complexity for course level.	
Appropriate choice & use of resources (e.g. Computer apps., intern sources, lab equipmen	Innovative selection of resources: expert use.	Appropriate resources used (such as demonstrated in class): resources limited to faculty-provided materials/tools.	Inadequate use of suggested resources	
Time management	Identified plans/timeline & worked to it: consistently met deadlines	Goals accomplished: most milestones met: some schedule defined; inconsistent use of time; misses some deadlines despite reasonable effort.	Missed significant milestones or project not completed.	
Information management: Log books, status reports, documentation.	Detailed, appropriate and timely entries: collected & distributed to appropriate parties.	Adequate entries in journal or log book: only critical data/information collected & distributed.	Insufficient data collection/ recording: existing documentation not shared/utilized.	
Conclusions & result interpretation	Obtained & adequately interpreted meaningful results with appropriate, insightful conclusions.	Produced some results, but struggled with interpretations or lacked sufficient support for their conclusions.	Generated few results with little meaningful interpretations; conclusions ore absent, wrong, trivial or unsubstantiated	

(Reid and Cooney, 2008, 894)

Reid and Cooney concluded, "The use of these rubrics alone cannot be a complete assessment plan: effective assessment requires additional direct measures to allow for

triangulation of data. "Data collection is a necessary step in an assessment plan but actions taken based on data analysis are more important than collection of raw data." (Reid and Cooney, 2008, p. 899)

Stewart, Choate, & Poteet defined performance evaluations: "Essentially, performance evaluation takes into account students' performance of behaviors, with minimal interference, often in real-life settings" (as cited in Widerman, 2003, p.110). Further, Busching stated, "Rubrics are guides for grading that outline assessment criteria and define successful achievement." (as cited in Widerman, 2003, p.110). Eileen Widerman, Ph.D. and Assistant Professor of Social work, Temple University, sought methods for evaluating her students' behaviors and performances in the social work program. It became increasingly difficult to rate students' ability to develop and apply concepts and behaviors necessary to social work without reducing it all to a letter grade. The letter grade became a reflection of subjective judgment rather than a true reflection of student ability.

Widerman stated "I began exploring and experimenting with nontraditional approaches to grading. I found that performance evaluation, using a rubric, allowed me to operationalize assignments, weight the importance of assignment components, and grade complex tasks. I could involve students in the assessment process and receive feedback on my teaching effectiveness." (Widerman, 2003, p. 110)

Indeed, student input in the evaluation process aided Widerman in becoming a more effective instructor while at the same time, giving her students clear goals and

guidelines by which they were judged. As cited by Owens, 1995: “Students, as well as faculty, experience pressure related to grading. Practice teachers have different interpretations of the concept of competence and how best to measure performance, often leaving students unsure of expectations.” (Widerman, 2003, p. 111)

In response to the vast differences in interpretations, Widerman investigated the value and applications of performance-based rubrics. Widerman’s findings were consistent with the findings of Huff & Johnson in 1998, (as cited in Widerman, 2003, p. 123) which indicated “that nontraditional teaching and evaluative approaches, such as learning contracts, formative evaluations, and narrative assessments, are positively associatied with students’ perceptions of empowerment.

Reeves and Stanford stated “a good rubric for assessing writing can serve (at least) two purposes evaluating students' knowledge and measuring teaching.” (Reeves, Stanford 2009, p. 25)

The expectations, or the vision of what the written work should look like, may be described in terms of "look fors" (for example, "look for five paragraphs"). These descriptors become the criteria or characteristics that will be used on a Likert scale for measuring student growth in their writing. The Likert scale may be very basic and in three levels (e.g., under the heading "Creative Word Choices" the following scale is used "Not Found (0 points)," "Found Less Than Two Times (1 point)," and "Found More Than Two Times (2 points)”.

Once the criteria were established, the educator was encouraged to ask the students what the students valued most for the success of the project. This allowed the students ownership of work and a voice in the manner in which the work was assessed.

By determining what criteria or characteristics were present in the highest level of performance and clearly describing these characteristics, the indicator of performance was understood by teachers, students and parents (Andrade, 2000). Likewise, by describing the lowest level of scorable student work or bare minimum of performance, students and teachers were often able to distinguish real differences in performance based on specific writing needs, goals, and differentiated instruction. (Reeves, Stanford 2009, pp. 25-26)

Through the employment of rubrics and front-loaded expectations, students as well as teachers, and parents were given the bar against which all student work would be measured. The use of such information allowed for clear and concise guidelines and expectations for all stakeholders.

Summary

According to Howard Gardener (1983) students possessed a wide variety of intelligences. These intelligences were directly related to student learning in that the intelligence, or method of acquiring and processing data had direct influence on the student. Tomlinson (1999) held that differentiation was necessarily made in order to reach students in the manner best suited to the individual's intelligence. While the goal or

purpose of teaching was the same for all intelligences and levels of ability, the manner in which students reached that goal varied. Assessment was performance based rather than standards based. Student who were front-loaded with criteria were more likely to reach the expectation or standard than those who were not given criteria beforehand.

Arter (2000), Ramey (2007), Reid (2008), Widerman (2003) and Wormeli (2006) all discovered that rubrics which were composed to clear, consistent, parallel language produced consistently higher student scores. This was especially true when student expectations and grading criteria were front-loaded.

CHAPTER 3

Methodology and Treatment of Data

Introduction

Students in the control group were given a minimum amount of information regarding expectations before being tested. No outlines or lists of expectations were given and students did not know in advance what the criteria were by which assessments were to be evaluated, as was typical of modern educational practices.

Scores were compared between the control group and the treatment group. The experimental group was given Marzano's general use rubric at the beginning of the academic year which outlines the criteria by which the examinations were assessed. Students in the treatment group knew the expectations in advance. Both the control group and the treatment group were assessed using identical examinations and scores were compared using an independent *t* test.

Methodology

The investigator employed a quasi-experimental, nonequivalent group study using an independent *t* test to discover the effectiveness of employing front loaded rubrics and grading criteria before the evaluation of student learning. The researcher compared two groups, one group to whom the treatment had not been given while the second group was given the treatment. Comparison was made between percentage grades received on a cumulative assessment with no clear guidelines on grading criteria and the group which received front-loaded criteria.

Participants

Participants in both the control group and the experimental group included first year, non-native speakers of Spanish as well as some bilingual students. Bilingual, bi-literate students were not included. Students who were bi-literate and bilingual were given the treatment but the scores were later removed from the test scores. Student population in the district in which the study was conducted included 68% Hispanic students, 25% white, with the remaining population including Asian, Pacific Islander, Black and American Indian making up the remaining 8% of student population. (Office of Superintendant for Public Instruction, School report card 2008) All participants were assumed to possess a minimum of eight years of formal education and all required skills to have reached ninth grade level. Population for research was necessarily limited to a convenience sample which consisted of first-year, non-native speakers of Spanish in a high school class. Participants were required to be present in class a minimum of 75 of the first 90 school days. Class size was between 25-30 students.

Population for the control group consisted of students in the classes from August to November of 2008. Number grades based on points received for examinations were compared with post-treatment grades of students from August to November of, 2009. The control group included students from two separate first year, non-native students. The two classes, in total, contained 33 boys, 25 girls. Combined the two classes consisted of 16 boys and 16 girls of Hispanic descent, 11 White boys, eight White girls, four African-American boys, one Ukrainian girl, two Ukrainian boys. One Hispanic boy

and one Hispanic girl were removed from the group because they each attended less than 75 days of the first 90 days of classes. The test was administered on the same day in November of 2008 to both classes of the control group. There were no criteria or rubrics given to the control group prior to testing.

Treatment was given to students of the 2009/2010 sample beginning with Marzano's rubric which allowed students to familiarize themselves with rubric language and usage. All students were given detailed explanations and opportunities to practice rubric application. All terms and vocabulary were defined and explained for maximum comprehension. The treatment group included two first year, non-native classes which, in total consisted of 23 boys, 32 girls. The population included 13 girls and 10 boys of Hispanic descent, eight White boys, 16 White girls, one Samoan boy, one Ukrainian boy, one Vietnamese girl, and five African-Americans, two girls and three boys. Two boys and one girl were removed from the test group because they did not participate in the test. The same test was given to the experimental group as the control group. The test was also administered on the same day to both classes in November 2009.

The researcher had three years of experience teaching first year high school Spanish at the time the study was conducted and had taught the curriculum for two consecutive years 2006/2007 and 2007/2008. For continuity of the study, the same assessments were administered to both groups, though the criteria by which the students were graded had been front-loaded in the case of the second, experimental group and not

in the case of the control group. Practice tests were administered to both the control group and the experimental group.

Instruments

The test administered to both the control group and the experimental group included eight parts. Part one included a listening element composed of a message left on an answering machine by a native Spanish speaker for another native speaker. Students were required to answer true or false to five statements about the listening selection. The listening portion was worth 10 points.

The next part of the test required students to write a dialogue between one of three groups of people in a drawing. Elements assessed included vocabulary use, grammar, style and creativity. This extended response portion was worth 15 points.

Section three required students to change nouns from singular to plural or from plural to singular. Students needed to use both definite and indefinite articles and be able to distinguish the uses of each type of article. No examples were given but the pattern was demonstrated as follows:

-Hay un lápiz - No. Hay dos _____

Students needed to fill in the blank with the plural form of the noun “lápiz”. The number/gender of nouns portion was valued at 10 points, with students being required to complete five sentences in the manner shown above. Students were expected to use a complete sentence for each response.

The fourth part of the exam required students to use learned vocabulary to answer questions regarding time. Once again test takers needed to use complete sentences as well as write out words for numbers in the time-telling phrase. Students were given 5 sentences such as the following:

-Hola, Reyes. ¿Qué hora es?

-Hola. *(It's 9:30 a.m.)* _____

The time-telling section valued 10 points.

Part five of the test was a fill-in-the-blank section which required the student to complete a conversation by filling in the missing words. Additionally, if a verb was required, the test taker must use the correct form of the verb. This section focused primarily on logical responses to a memorized pattern of conversation and so, of necessity, was in dialogue form. There were 10 missing words or phrases in the dialogue so part five was worth 10 points.

Section six required students to write complete, logical responses to personal questions using vocabulary acquired in the chapter. The five responses were valued at 15 points with a point each for correct grammar, target vocabulary use, and proper punctuation.

Part seven assessed reading comprehension. Test takers needed to read two short advertisements from a newspaper. Students then answered four questions about the reading. Test takers had to use context cues as well as word recognition from target

vocabulary in order to create logical, complete responses. Each response was worth two points, one for proper grammar and the other for understanding the question.

The eighth portion of the test was an extended essay. No clues were given in the target language though topics to address were listed in English. Writers needed to construct a dialogue which included a self-introduction, ask the person they were addressing how they felt, ask where each conversant was from, mention what time it was and say goodbye. This section of the test was valued at 22 points with eight points each for proper grammar and vocabulary use and six points for style.

There were two similar versions of the test administered to the control group and the experimental groups. Students were seated close together due to limited space so cheating was made more difficult if not impossible by giving each student a different version of the test from that of the person seated on either side. Students were allowed to use vocabulary lists. Test takers were not allowed to use notes, ask questions, talk to anyone or leave the room during the test. Tests remained in the room after the test and tests were not graded until after all examinees were finished. No students tested in either the control group or the experimental group different day from other students in that respective group. No electronic devices were allowed in the room and no answers were given to any students until after the tests were corrected and grades were entered into the permanent record. There were no retakes of the test by any participant in either the control group or the experimental group. No copies of the test left the room with any student at any time either in the control group or the experimental group.

Design

An independent *t* test was used to discover if a significant difference existed between front-loaded rubrics and student success on cumulative finals. Scores were included from two separate academic years using first year Spanish classes for non-native speakers. The control group included scores from two classes at the first year, non-native speakers' level. Treatment group was two classes of first year, non-native speakers' level. A comparison was made between the two sets of raw scores using a quasi-experimental design. A quasi-experimental design, the non-equivalent control group design, was employed because the classes were pre-selected with no input from the researcher. A comparison was made between the two sets of raw scores.

Procedure

The investigator used the scores from the first quarter final over the first chapter of the text book Descubre 1. The chapter final was the pre-written final which came as part of the curriculum produced by the authors of the text book. Identical tests were employed in both the control group and the experimental group. Each time the examination was administered, students were given either version A or B, alternating between the two, so that no student was seated next to a person with the same version of the test. The alternation between the test versions was intended to prevent the possibility of cheating or copying of answers as each version asked different yet similar questions. Identical scoring criteria were used when correcting examinations. Criteria included:

1. A listening section in which students were required listen to a message in Spanish left on an answering machine. Test takers were then required to decide whether the statements given on the paper were true or false according to what they heard.

2. An extended response section in which students were required to write a conversation between one of the groups of people shown in a picture. Points were given for vocabulary use, grammar, style and creativity.

3. A section in which students were required to change number and gender of nouns using complete sentences.

4. Written responses to questions about telling time and telling at what time certain events occurred. This was primarily a translation from English to Spanish.

5. Use of vocabulary introduced in the chapter. This was a fill-in-the-blank section which employed memorized sentences and phrases acquired through practice and repetition.

6. The answering of questions in the target language. Students were required to use memorized, rote responses to memorized questions.

7. Reading comprehension. This section required students to employ reading strategies such as word recognition and contextual clues while reading in Spanish in order to form logical Spanish responses to the questions using correct grammar and vocabulary.

8. An extended, primarily unguided written response to a specific task. Students needed to construct a conversation between three imaginary students. Criteria included:

Self-introduction, a question about how each was feeling, ask place of origin, mention what the time was, and took their leave of one another. Grading requirements included, once again, correct grammar, punctuation and vocabulary as well as style.

Treatment of the Data

The author employed the Statpak (Gay, Mills & Airasian, 2007) when calculating and comparing scores in the independent *t* test. Scores from both first-year, non-native classes from 2008/2009 were entered in as the values for the y (control) group while the scores from the two classes of non-native, first year classes from 2009/2010 were entered as the x (experimental) group. Scores for students who did not participate in the quarter final were not included. Scores for students receiving a zero who did take part in the test were included. Students who were in class for less than 75 of the first 90 days of school were not included. Bi literate, bilingual students were removed from both the control group and the experimental group within the first 10 days of the school year and so were not included.

Summary

It was hypothesized that students taken from a convenience sample and given training in rubric use would achieve significantly higher tests scores that the students not so trained. Based on the writings of Howard Gardener (1983), Carol Ann Tomlinson (1999), Grant Wiggins & Jay McTighe (2003) and others, the investigator held that student test scores were significantly higher if students had been given the expectations before attempting an assignment, assessment or supplemental task.

Student scores on a multiple-part, task-based assessment were compared using a quasi-experimental, non-equivalent group test. Using an independent *t* test, test scores from the control group, students who received no training or front-loaded expectations were compared to the tests scores of the treatment group. Treatment group students were given training in rubric use and expectations were front-loaded for the treatment group. The scores were entered in the Statpak (Gay, Mills & Airasian, 2007) using the independent *t* test and significance was sought at $p \geq .05, .01, .001$ levels.

CHAPTER 4

Analysis of the Data

Introduction

The project was intended to determine if front-loading of expectations at the beginning of the learning process result in higher scores on formal assessments. If students were given the desired outcomes and expectations, would students be more likely to reach the goals as outlined?

Description of the Environment

Parameters of the project were limited to the first quarter of the academic school years for both the control group and the treatment group. Due to the fact that the researcher had no control over the populations of either the control group or the treatment group, the samples used were convenience samples in a quasi-experimental, non-equivalent control group design. The tests for both the control group and the treatment groups were administered at the end of the first quarter, at the beginning of November, using identical sets of examinations. The exams administered were the pre-written, publisher-produced tests that were included in the curriculum for Descubre 1 and were graded using the criteria set forth by the publishers.

The researcher had a Bachelor of Arts degree in Education for Spanish and three years of teaching experience in first and second year Spanish classes for non-native speakers at the high school level. Students were assumed to have at least eight years of formal education and to possess the reading and writing skill necessary to students entering the ninth grade. Students in both the control group and the treatment group were between 14 and 18 years of age, and were in the ninth through twelfth grades.

Demographics of the control group included 19 White students, 32 students of Hispanic descent, as well as one Vietnamese, four African-Americans and three Ukrainians. There were 33 boys and 25 girls with a total student population of 58. Students included in the control group attended at least 75 of the first 90 days of school. One student was removed for non-participation in the test and two because the student attended less than the required 75 days.

Treatment was given to classes consisting of 24 White students, including eight boys and 16 girls. The 23 students of Hispanic descent included 10 boys and 13 girls. The remaining students included one Vietnamese, one Ukrainian, five African-Americas and one Pacific Islander. There were 23 boys and 32 girls in the treatment group, all of whom attended classes at least 75 of the required 90 first days of school. There were 55 students in the treatment group. Three students, one girl and two boys, were removed for not participating in the test.

Research Hypothesis

If trained in the use of rubrics before being assessed, students were able to achieve significantly higher scores than students who were not so trained. Front-loaded expectations would allow students to be aware of expectations and keep the goal in mind before beginning a project, activity or assessment.

Null Hypothesis

If students were not given training in rubric use, they would not receive equal or significantly higher scores than students who were trained in rubric use or front-loaded the expectations. Significance was determined for $p \geq .05, .01, .001$.

Results of the Study

The results of the study provided the data in order to address the hypothesis of the research. The treatment group and the control group completed the formal assessment. The resultant scores were analyzed using the Statpak (Gay, Mills & Airasian, 2007), which produced the statistics and values.

Table 2.

Student Scores for Control Group

Students	Scores	Students	Scores
C1	100	C28	83
C2	86	C29	100
C3	94	C30	98
C4	84	C31	92
C5	96	C32	86
C6	99	C33	96
C7	80	C34	93
C8	96	C35	71
C9	93	C36	88
C10	98	C37	85
C11	94	C38	99
C12	90	C39	84
C13	92	C40	99
C14	99	C41	79
C15	55	C42	81
C16	97	C43	72
C17	85	C44	91
C18	96	C45	100
C19	93	C46	80
C20	97	C47	92
C21	82	C48	89
C22	93	C49	99
C23	73	C50	95
C24	0	C51	61
C25	99	C52	97
C26	97	C53	93
C27	100	C54	88

Scores were entered into the Statpak (Gay, Mills & Airasian, 2007) using the scores from the control group as the *y* group. The highest score was 100 and the lowest score was 0, while of the 54 total scores the median was 85.72. The sum of the scores in *y* group was 4643.00 with a mean of 89.29.

Table 3.

Student Scores for Treatment Group

Students	Scores	Students	Scores
T1	97	T27	64
T2	88	T28	97
T3	100	T29	98
T4	97	T30	93
T5	100	T31	93
T6	94	T32	93
T7	100	T33	69
T8	96	T34	89
T9	76	T35	90
T10	96	T36	95
T11	95	T37	95
T12	81	T38	98
T13	96	T39	89
T14	78	T40	94
T15	98	T41	61
T16	97	T42	99
T17	93	T43	87
T18	97	T44	100
T19	91	T45	97
T20	98	T46	88
T21	70	T48	97
T22	0	T49	97
T23	88	T50	96
T24	74	T51	100
T25	100	T52	79
T26	94	T53	91

A *t* score of 0.38 was determined in the statistical analysis (Gay, Mills, & Airasian, 2007). The means of the control and experimental group's exam scores determined the value of *t*. The mean of the treatment group was 89.29, and the mean of the control group was 86.13. The degrees of freedom were 104. The evidence suggested the treatment group did not score significantly higher than the experimental group. Obviously, front-loaded expectations and/or guidelines had no significant effect on student learning in the assessment.

Values were taken from the Statpak program (Gay, Mills & Airasian, 2007) after all test scores were entered. The scores for the treatment group were entered as group x scores and the scores for the control group was entered for group y scores.

Table 4

Analysis Using *T* test and Independent Samples

Statistics	Values
No. of scores in Group x	52
Sum of Scores in Group x	4643.00
Mean of Group x	89.29
Sum of squared scores in Group x	427453.00
SS of Group x	12886.67
No. of scores in Group y	54
Sum of scores in Group y	4759.00
Mean of Group y	86.13
Sum of squared scores in Group y	432507.00
SS of Group y	13098.09
t- value	0.38
Degrees of freedom	104

$$t = \frac{\overline{X1} - \overline{X2}}{\left(\frac{SS1 + SS2}{n1 + n2 - 2}\right)\left(\frac{1}{n1} + \frac{1}{n2}\right)}$$

$$0.38 = \frac{89.29 - 86.13}{\left(\frac{427453.00 + 432507.00}{52 + 54 - 2}\right)\left(\frac{1}{52} + \frac{1}{54}\right)}$$

$t = 0.38$

Significance was sought at levels .05, 01 and .001. Degrees of freedom were 104 and the *t* value was 0.38. The sum of the scores in *y* group was 4759.00 with a mean of 88.13. Distribution of *t* at .05 was 1.980, at .01 the distribution of *t* was 2.618 and at .001 the distribution of *t* was 3.373.

Table 5.

Distribution of *T*

	p		
df	*.05*	*.01*	*.001*
104	1.980	2.618	3.373

Findings

Students who received the front-loaded expectations did not achieve significantly higher scores at any of the three distributions of *t.* The Statpak analysis calculated a value of *t,* 0.38, which was lower than the threshold value at .01 which was 1.658. This suggested that students who received front-loaded expectations were no more likely to achieve higher scores than students who did not receive the expectations ahead of time. Thus the null hypothesis was accepted and the hypothesis was not supported. Null hypothesis stated if students were not given training in rubric use, they would not receive significantly higher scores than students who were trained in rubric use or front-loaded the expectations.

The hypotheses stated if trained in the use of rubrics before being assessed, students were able to achieve significantly higher scores than students who were not so trained. Front-loaded expectations allowed students to be aware of expectations and keep the goal in mind before beginning a project, activity or assessment. There was no support for the hypotheses.

Discussion

The researcher was unable to locate literature that paralleled the treatment performed. Available research existed that supported the use of rubrics for a wide variety of purposes yet none was located that promoted the front-loading of expectations or grading guidelines. According to literature located, the use of rubrics allowed students a clear picture of expectations after the assessment, activity or assignment was completed. Presentation of the grading criteria provided students with ability to understand in what areas improvement was needed after the fact. However, no research had yet been performed that supported the idea that students needed the guidelines prior to fulfillment of the required assignment or assessment.

The author expected that scores would be significantly higher after the treatment. The researcher held that if students were given the parameters or expectations before the group was asked to complete a task, assignment or assessment, the students would receive significantly higher scores after the treatment. The scores students achieved after the treatment did not bear the theory out.

Statistical analysis demonstrated that front-loading of expectations did not have a significant effect on student performance on a cumulative assessment. While student

performance on daily assignments and other, related activities were better, test scores showed no significant improvement.

Summary

When the scores of the control group were compared with the scores of the treatment group and calculated using the Statpak (Gay, Mills & Airasian, 2007), there was no significant difference between the scores. Significance was sought at .05, .01 and .001. Degrees of freedom were 104 and the t value was 0.38. The null hypothesis was accepted and the hypothesis, which stated that student scores would be higher if the learners were provided the guidelines before the performance of a given task, was not supported.

CHAPTER 5

Summary, Conclusions and Recommendations

Introduction

The researcher held that teachers were not trained in matching class expectations to student needs. As each individual possessed different intelligences or preferences with regard to acquiring new information and skills, it behooved teachers to adjust instruction to accommodate as many of eight intelligences as outlined by Howard Gardener (1983) as possible.

Through the use of clear, front-loaded expectations and consistent, concise rubrics with parallel language, students were able to achieve significantly higher scores on assessments than students who were not so trained. Writings by a variety of scientists and researchers such as Rick Wormeli (2006), Judith Arter (2000), Eileen Widerman (2003) and others suggested that the use of rubrics had a significant positive impact on student scores.

The project was intended to discover if front-loading of expectations at the beginning of the learning process engendered interest in self-improvement as demonstrated by higher scores on formal assessments. If students were given the desired outcomes and expectations, would the students be more likely to reach the goals as outlines?

Summary

In chapter 1, the researcher outlined the demographics of the region in which the experiment took place. Delimitations were included as were the assumptions which were made prior to the commencement of the experiment. A quasi-experimental, nonequivalent group study was employed to test the hypothesis and data were entered in the Statpak (Gay, Mills &

Airasian, 2007) to determine whether or not there existed a significant difference between the scores of the treatment group and the scores of the control group. Definitions were listed as were the acronyms used.

Chapter 2 included observations by Howard Gardener which stated students possessed a wide variety of intelligences. These intelligences were directly related to student learning in that the intelligence, or method of acquiring and processing data had direct influence on the student. Tomlinson (1999) held that differentiation was necessarily made in order to reach students in the manner best suited to the individual's intelligence. While the goal or purpose of teaching was the same for all intelligences and levels of ability, the manner in which students reached that goal varied. Assessment was performance based rather than standards based. Student who were front-loaded with criteria were more likely to reach the expectation or standard than those who were not given criteria beforehand.

Arter (2000), Ramey (2007), Reid (2008), Widerman (2003) and Wormeli (2006) all discovered that rubrics which were composed to clear, consistent, parallel language produced consistently higher student scores. This was especially true when student expectations and grading criteria were front-loaded.

It was hypothesized in chapter 3 that students taken from a convenience sample and given training in rubric use would achieve significantly higher tests scores that the students not so trained. Based on the writings of Howard Gardener (1983), Carol Ann Tomlinson (1999), Grant Wiggins & Jay McTighe (2005) and others, the investigator held that student test scores were significantly higher if students had been given the expectations and were trained in the use of the

rubric by which students were scored before attempting an assignment, assessment or supplemental task.

Student scores on a multiple-part, task-based assessment were compared using a quasi-experimental, non-equivalent group test. Using an independent *t* test, test scores from the control group, students who received no training or front-loaded expectations were compared to the tests scores of the treatment group. Treatment group students were given training in rubric use and expectations were front-loaded for the treatment group. The scores were entered in the Statpak (Gay, Mills & Airasian, 2007) using the independent *t* test and significance was sought at $p \geq .05, .01, .001$ levels.

The author employed a quasi-experimental, non-equivalent group study, in chapter 4, to determine if there was a significant difference between the two groups, the control group and the treatment group. Scores were entered in the Statpak (Gay, Mills, & Airasian, 2007). When the scores of the control group were compared with the scores of the treatment group and calculated using the Statpak (Gay, Mills & Airasian, 2007), there was no significant difference between the scores. Significance was sought at .05, .01 and .001. Degrees of freedom were 104 and the *t* value was 0.38. The null hypothesis, which stated that student scores would not be significantly higher after the treatment, was accepted and the hypothesis, student scores would be significantly higher after treatment, was not supported.

Conclusions

The short period of time in which the study was conducted was not sufficient to effect change. Students and the teacher were attempting to counteract decades of habits and practices

using only one ten-week period. The theory of multiple intelligences as outlined by Gardner was not as thoroughly addressed as was necessary. Students needed more time to adapt to teaching which allowed students to employ each individual's personal preferences before the assessment was administered. Further, the assessment was graded in a traditional style while all previous assessments and assignments were rated using Marzano's general rubric. The inconsistency may have had an unintended effect on the outcome of the study.

Recommendations

The researcher recommends that a longer time be used to establish norms for any further tests. Due to the shortness of time for the study, the researcher used scores from the previous year and compared them to the scores of the year after in which the treatment was given for the first time. For a more comparable study it is recommended that students and teachers receive additional training and the same general rubric be employed throughout the study. In other words, the students should be rated using the same rubric in all assignments and assessments and the scores compared between the front-loaded group and the group which did not get expectations ahead of time. The variables were too many for an accurate comparison of resulting scores.

REFERENCES

Armstrong, Thomas. *Multiple intelligences* Retrieved on 09/05/09 from www.thomasarmstrong.com/mulitple_intelligences.htm. Google search engine.

Arter, Judith A., (2000). Paper presented at the Annual Meeting of the American Educational Research Association *Rubrics, scoring guides and performance criteria: Classroom tools for assessing and improving student learning.* New Orleans, LA. Retrieved on 08/29/09 From Educational Data bases from EBSCO.

Gardener, H. (1983) *Frames of Mind: The theory of multiple intelligences.* New York: Basic Books. Basic Books Paperback, 1985. Tenth Anniversary Edition with new introduction, New York: Basic Books, 1993.

Gay, Mills & Airasian. (2007) *Statpak.* Retrieved from http://www.prenhall.com/gay

Marzano, R. J., Pickering, D.J., & Pollock, J. E. (2001) *Classroom instruction that works: research-based strategies for increasing student achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Ramey, Sandra, VanderVusse, Leona, Gosline, MaryBeth. (2007). Using a written communication rubric to improve student writing. *International Journal of Learning, 13,* 67-74. Retrieved on 08/27/09 from Educational Data-bases from EBSCO.

Reid, Kenneth J., Cooney, Elaine M. (2008). Implementing rubrics as part of an assessment plan. *The International Journal of Engineering Education, 5,* 893-900. Retrieved on 08/29/09 from Educational Data-bases from EBSCO.

Tomlinson, Carol Ann, (1999). Mapping a route toward differentiated instruction. *Educational Leadership,* 57, 1, 12-16.

Widerman, Eileen. (2003) Performance evaluation using a rubric: Grading student performance in practice courses. *The journal of baccalaureate social work,* 8 (2) 109-125.

Wiggins, Grant, McTighe, Jay. (2005) *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wormeli, Rick. (2006) *Fair isn't always equal: Assessing and grading in differentiated classrooms.* Stenhouse Publishers, 3-4.

APPENDIX

General Use Rubric for topics Addressed in Class

4. The student has a complete and detailed understanding of the information important to the topic AND the student can perform the skills and processes important to the topic fluently and without error.
3. The student has an understanding of the information important to the topic but not in great detail AND the student can perform the skills or processes important to the topic without significant error.
2. The student has some misconceptions or is missing some information important to the topic but still has a general understanding of the topic AND/OR the student makes significant errors when performing the skills or processes important to the topic but still performs a rough approximation of these skills and processes.
1. The student has major misunderstandings or is missing critical information about the topic AND/OR the student cannot perform even a rough approximation of the skills and processes important to the topic.