TRADITIONAL VERSUS NARRATIVE ASSESSMENT ITEMS:

A QUASI-EXPERIMENTAL COMPARISON

by

SARAH GRABOWSKI and THOMAS MACAULAY

A Paper Submitted to Professor Thomas C. Reeves

in Partial Fulfillment of the Requirements of EDIT 6900

THE UNIVERSITY OF GEORGIA

ATHENS, GEORGIA

FALL 2003

Table of Contents

Introduction 3
Statement of the Problem 3
Theoretical and Conceptual Framework 4
Literature Review 5
Methodology 6
Timeline 9
Results 10
Interpretation 12
Discussion 13
Recommendations 14
Conclusion 15
Appendix 16
- Appendix A: Research Proposal 17
- Appendix B: Teacher Instructions 25
- Appendix C: Assessment Part A 26
- Appendix D: Assessment Part B 27
- Appendix E: Data Collected 28
- Appendix F: Legend for Table of Data Collected 29
- Appendix G: Statistical Analysis 30

References 31

Introduction

Technology and new types of assessments are being considered in classroom education more and more. Now with the No Child Left Behind Act that mandates testing, assessment strategies have become even more important to the K-12 classroom. However, multiple-choice tests have been criticized for their lack of substance in demonstrating the transfer of knowledge (Abrams, Pedulla & Madaus, 2003). An alternative to these traditional assessment items is testing that includes a story-based scenario. Assessments that contain story-based scenarios engage the student, offer a relevant or real-world situation and enhance the student's interest and motivation to achieve. Our objective was similar in examining alternative assessment designs that would sharpen and elaborate knowledge and ability distinctions in fields such as math (Snow, 1995, p. 17).

Feedback is fundamental to learning, but opportunities to receive it are often scarce in classrooms. Students may receive grades on tests, but these are summative assessments that occur at the end of projects; also needed are formative assessments that provide students opportunities to revise and improve the quality of their thinking and learning (Bransford, Brown, & Cocking 1999). If the goal is to enhance understanding, it is not sufficient to provide assessments that focus primarily on memory of facts and formulas but rather on the application of the concepts. It is our hope that teachers and schools will change their approach to classroom assessments after seeing how students are able to apply knowledge with the use of story-based scenarios. This will enhance the classroom learning environment that allows teachers to evaluate learning, remediate or adjust their teaching style before moving on.

However, it is not known if story-based scenario test items are as effective in measuring learning as traditional test items. This research project set out to compare traditional test questions to story-based scenario questions in hopes that a realistic problem might engage a student more because critical thinking and decision making are utilized in order to solve that problem.

Statement of the Problem

This research project is focused on assessment strategies for the K-12 classroom. The study evaluates the effectiveness of traditional tests questions versus an alternative assessment

tools. The research conducted specifically compares fill-in the blank test questions to story-based scenario questions.

Since a story-based scenario would require the student to use higher order thinking, we hypothesized that students would demonstrate a greater comprehension of the material in the assessment items that included story elements. Students would have to apply what they have learned in order to successfully complete the assessment. This requires students to rely on their higher-order thinking skills: critical thinking and decision-making.

The study was conducted to determine if there is a greater comprehension based on the type of assessment strategy used in the classroom. The project also hoped to shed light on the purpose and objective of an assessment strategy, in general. A sound assessment strategy evaluates the learning achieved by the student, the ability to apply those concepts and the effectiveness of the instruction (Senk, Beckmann, & Thompson, 1997). The strategy used to determine the assessment should also consider the level of instruction delivered and match instruction accordingly.

Because technology and new assessments types are being considered in K-12 classroom education more, our project can be considered as socially responsible with respect to the investigation of these important new evaluations of learning. It is necessary for new classroom techniques to be tried and investigated in order to allow others to realize their mass appeal and effectiveness. Teachers and researchers investigating the new assessment strategies are exploring new classroom techniques for use by others (Bocij & Greaskey, 1999). Building a knowledge-base of classroom experiences that are successful as well as those that are failures is important work. It is necessary work if new assessment strategies are to be readily accepted and easily delivered by practitioners.

Theoretical and Conceptual Framework

Our research relied heavily upon Bloom's Taxonomy as a basis for developing assessments that match the lesson. Constructionism is another theory that was considered as we tried to find alternative assessment techniques. At the core of our efforts, we were investigating whether these learning theories can assist in building better assessment strategies and techniques. Our literature review and research focused on using assessments that demand higher-order

thinking as evidence that learning is taking place. When this evidence is present, it is not just a question of if students gain new knowledge, but if they are able to apply it.

Much of our literature review also revealed to us the motivation and anxiety caused by assessments. Alternative forms of assessment in some cases can empower the student and lower anxiety and give them a stronger sense of control. Effective assessment strategies should consider the Six C's of Motivation (choice, challenge, control, collaboration, constructing meaning, and consequences) in attempting to achieve flow (Wang & Han, 2001).

Another theoretical and philosophical focus for our research was the objectives of an assessment. Our literature review and research looked at the goals of evaluation. It is important to analyze the purpose for classroom assessments as a tool for evaluating the instruction, the students' progress and their ability to apply that knowledge.

Literature Review

The literature reviewed focused on the studies and theory used to develop assessments. We researched the various impacts assessments have on the student as well as different assessment strategies. Much of our literature review focused on the amount of control and anxiety levels a student experiences during the assessment and throughout the course. Technology course enhancements and new assessment methods were also reviewed. These new computer-based options are being implemented with varying success, but break ground into new possibilities for assessment. Existing research evaluates the effectiveness of these assessments on various issues such as comfort, relaxation, anxiety, time, and scores (Butler, 2001). Many studies in our literature review focused their attention on strategies to implement new assessment technologies to improve existing courses, not a wholesale change in the course delivery.

As for the theory portion of our literature review, it was necessary to draw upon some of the effective learning taxonomies and learning theory to evaluate the topic of assessment. Some of the topics that we evaluated in our literature review were Bloom's taxonomy, Constructivism, and Montessori based methods (Elkind, 2003). These topics are important to our research topic because they are widely used as alternatives to Instructionist teaching methods.

Of these topics, the research on Bloom's taxonomy as a framework for classroom assessment conducted by Kastberg (2003) had the greatest impact on our study. In this study, Kastberg attempts to apply Bloom's Taxonomy to the preparation of testing materials in an

attempt to find a more systematic approach to assessing learning and then implement an assessment strategy that provides better evidence of students' content knowledge and their ability to use that knowledge.

This study helped us recognize two important features of an assessment. The first is the evaluation of what the student learned. By using Bloom's taxonomy, students' learning is assessed in a manner that can yield particular insights into their thought process, which then helps perfect a teacher's ongoing performance. The second is that assessment is a tool by which teachers can evaluate their teaching strategies and classroom activities in order to revise its future delivery.

The literature reviewed also raised the fundamental question of the purpose of assessment tools (Henderson, 2001). It is a matter of developing alternative assessment types and questions that allow the student to express their knowledge. By using higher order thinking, the students show their grasp for the content rather than simply reciting it. Using higher order evaluation also gives the teacher a chance to reflect on the connection between the content and its application, aligning their teaching strategy in an appropriate fashion (Kastberg, 2003). This will help to guide our research goals in determining the effectiveness of the teaching, the measure of comprehension by the student and their ability to apply that knowledge (Byers, 2002).

Methodology

At the core of our efforts, we are investigating whether learning theories, such as Bloom's Taxonomy and Constructionism, can assist in building better assessment strategies and techniques. We are interested in using assessments that demand higher-order thinking as evidence that learning is taking place. When this evidence is present, it is not just a question of if the students understand but if they can effectively apply the knowledge. Since we want to investigate the effectiveness of traditional tests questions versus an alternative assessment tools, the selection of quantitative methods is appropriate.

Our research was conducted at the Branch Christian Academy in Lawrenceville, Georgia, specifically the second grade mathematics class. All second grade students attended the same two-day lesson about how to read the time from and draw the time on an analog clock.

A two-part assessment has been designed for this research project. Special attention was given to the measure of criterion validity of the assessment. All students involved first took

assessment part A and then immediately took assessment part B. Part A, located in Appendix C, tested students on their knowledge of time using traditional fill-in the blank questions. There are four questions in this section. Each question has three sections. Section one asks the student to read the time from the clock; section two asks the student to write the new time given a certain amount of time has passed or a certain amount of time earlier; section three asks the student to draw the new time on a new clock.

Part B, however, uses a story-based scenario to ask questions similar to those found in part A. After all students had completed part A, the teacher next read a scenario about a field trip to the zoo to the students. Then students were asked to answer the three accompanying questions based on the scenario. The first question required students to add a certain amount of time to the given time and then find the difference between this named time and another given time. This question was designed to relate to question four of part A. The second question asked students to find a new time given a certain time interval had passed. This question was very similar to question one of part A, except that a story accompanied the question in part B. The final question, asked the student to draw the hands on the clock based on the previous question's answer. Again, this question was designed to mirror question one of part A. This method of similar content was used in order to compare the results.

Collecting the student-written responses generated the data for this research project. Responses are in the form of written time (xx:xx), amount of time in minutes and drawing the hour and minute hands on an analog clock. Students were also encouraged to draw a picture based on the scenario in the story, so there are also pictures included. The pictures, however, were not used in the data analysis. Privacy was respected by numbering of the assessments where gender was the only item requested from the student's identity. We gave the same test to each student. This is in part because the pool of students was limited. We also believe that this is an experiment that can be replicated in other classroom settings to highlight the effectiveness of assessments. Teachers and practitioners can assess the level of understanding by using Bloom's Taxonomy to create assessments that test knowledge at various levels. If it is created carefully, it becomes an evaluation for the student with a plan for remediation.

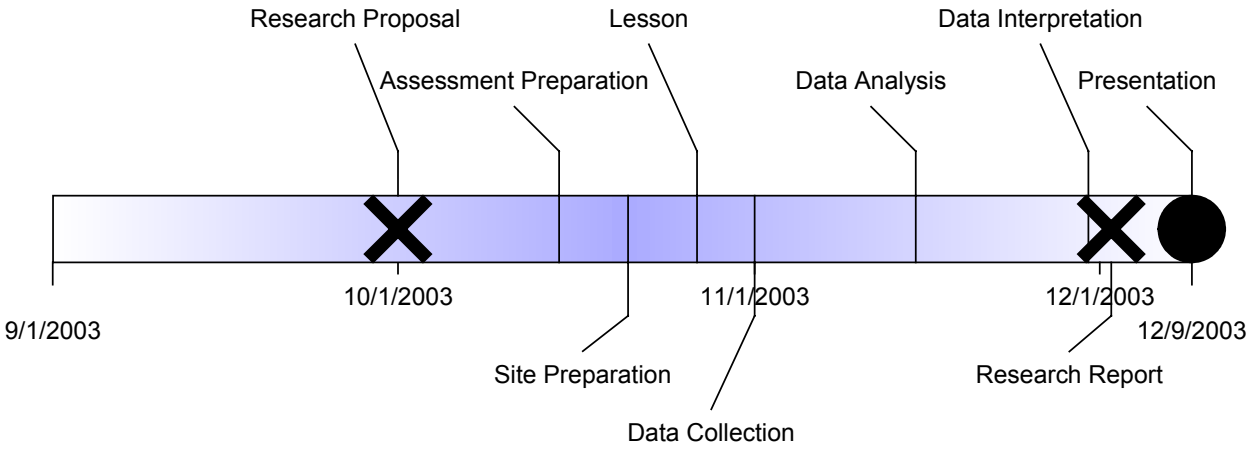
For every section of every question, the data was judged to fit into one of three possible categories: correct answer, incorrect answer, or not attempted/ left blank. Totals were found for

each category of every section of every question. For example, there were twenty responses total for question 1, section a: thirteen correct, five incorrect, and two not attempted.

Comparisons were then made between the questions designed to be similar in parts A and B. Statistical methods were used to seek out any statistical differences between the two assessment methods. Charts and diagrams were also constructed to show any visual relationships.

Timeline

The project began during the first week of September 2003. The timeline for this project was, as follows.



Task to Complete	Estimated Amount of Time Needed	Target Date for Completion	Date Completed
Research Proposal	4 weeks	October 1, 2003	October 1, 2003
Assessment Preparation	2 weeks	October 15, 2003	October 15, 2003
Site Preparation	1 week	October 21, 2003	October 21, 2003
Lesson	3 days	October 27, 2003	November 7, 2003
Data Collection	3 days	November 1, 2003	November 14, 2003
Data Analysis	2 weeks	November 15, 2003	November 21, 2003
Data Interpretation	2 weeks	November 30, 2003	December 2, 2003
Research Report	4 weeks	December 2, 2003	December 7, 2003
Presentation of Report	1 week	December 9, 2003	December 9, 2003

Results

As stated above, the assessment was divided into two parts: part A contained the traditional fill-in the blank questions and part B used a scenario to introduce the questions. The results are as follows:

Of the thirty-five tests handed-out, twenty-seven were returned for data generation. The gender breakdown was 15 female (56%) and 12 male (44%). Of the twenty-seven returned responses, four were considered to be invalid due to incompletion. The teacher did not give out the additional eight blank assessments because those students had not yet received the lesson on time.

Of the 345 total questions asked in the research project (twenty-three tests times fifteen questions per test), 172 questions (50%) were answered correctly, 135 questions (39%) were answered incorrectly and 38 (11%) were left blank. See figure 1 below

Information	Fill in the blank	Story-based
Correct	55	117
Incorrect	10	125
Blank	4	34
Total	69	276

Figure 1 shows breakdown of 345 questions.

When students were asked a question which required writing their answer in minutes, they answered correctly 87% of the time when it was accompanied by a word problem (part B) but only 52.17% of the time without an accompanying story (part A). 45% of the questions that required a written time answer in part A were incorrect responses, while only 13% were incorrect in part B. Also, almost 2% of the part A written answers were left, while 0% of the part B written answers were left blank. See figure 2

Information	Fill in the blank/%		Story-based/%	
Correct	96	52.17%	40	87%
Incorrect	83	45%	6	13%
Blank	5	2.7%	0	0
Total	184		46	

Figure 2 shows answers that required written response

When students were asked to draw hands on the clock to represent their answer, they answered correctly 65.22% of the time when it was accompanied by a word problem (part B) and only 22.83% of the time as a fill-in the blank (part A). See figure 3.

Information	Fill in the blank/%		Story-based/%	
Correct	29	52.17%	15	65.22%
Incorrect	42	45%	4	17.39%
Blank	21	2.7%	4	17.39%
Total	92		23	

Figure 3 shows answers that required written response

Students answered the questions in part B correctly 82.61% of the time, when comparing to a similar question (adding two hours and forty-five minutes to the clock). In part A, students answered the question correctly only 56.52% of the time.

Students answered a comparable question in part two 91.3% correctly when it accompanied a word problem. In part one, students answered the same question correctly 28.26% of the time. See figure 4

Information	Fill in the blank/%		Story-based/%	
Correct	13	28.26%	21	91.3%
Incorrect	22	47.83%	2	8.7%
Blank	11	23.91%	0	
Total	46		23	

Figure 4 shows results of adding 30 minutes versus a word problem adding 30 minutes

Interpretation

When students were asked to draw hands on the clock to represent their answer, they answered correctly 65.22% of the time in part B, when the question was accompanied by a word problem, as in part B, and only 22.83% of the time as a fill-in the blank, as in part A. The word problem and the use of higher order processing were valuable to their success in answering questions correctly here as well. With the word problem, the students needed to figure out their answer first. They still achieved higher scores.

When students were asked a question which required writing their answer in minutes, they answered correctly 87% of the time when it was accompanied by a word problem (part B) and only 52.17% of the time without a story (part A). The same can be said about these questions. The word problem and the use of higher order processing was valuable to their success in answering questions correctly. They were required to use higher-order thinking to figure out the answer. Their scores were considerably higher than just answering the same question without any frame of reference or relationship to the story.

Students answered correctly 82.61% of the time in part B, when comparing a similar question (adding two hours and forty-five minutes to the clock). In part A, students answered the question correctly only 56.52% of the time. By reading the story to the students and offering a frame of reference in the story, the students were able to apply the knowledge and answer correctly more often. The story had several pieces of information that the student needed to distinguish as important to the answer. This created an even higher degree of difficulty. The students still faired better using higher order thinking.

Students answered a comparable question in part B 91.3% correctly when it accompanied a word problem. In part A, students answered the same question correctly 28.26% of the time. The students were far more likely to answer correctly when the question was delivered as a word problem. The students were asked to answer two questions in part A that required them to add 30 minutes to the time given. Their answer to one was fill-in the blank and one was drawing the hands on the clock. They were more likely to answer incorrectly without a word problem. Most students also attempted the part B questions. This may be because of a rush to complete the assessment as the comparable part A question was last.

Our last focus area of data interpretation was to perform a statistical analysis on the collected data. We decided to conduct a simple two-sample t-test in order to compare the percentages of correct responses between several of the similar questions in parts A and B. Our sample size was small, so a t-test was appropriate in letting us know what to expect if we had tested the entire population. A 95% confidence interval was used. In all cases, we were able to reject the null hypothesis that the percentage correct for problems in part A would be equivalent to the percentage correct for problems in part B (Appendix G). Looking at the analysis as a whole, the percentage of correct in part B would be at least 2% and at most 75% greater than the percentage correct in part A.

One-sample t-tests were also conducted on the percentage of students who got a particular question wrong in part A but got the similar question in part B correct. Again a 95% confidence interval was used, and the percentage of students who got the question wrong in part A but correct in part B will be at least 10% and at most 77% of the population. This implies that our findings are not due to chance or coincidence, but indeed there is a relationship between higher percentages correct and the questions in part B.

Discussion

There are several factors that may contribute to the success of part B. While the questions were more difficult, there were fewer questions in this section. The rate of incorrect results increased at the end of part A. This might be a result of boredom. The part B questions would have captivated their attention since they were interesting, less repetitive and shorter. This would point to a “flow” situation in part B versus boredom in part A.

Another possible contributing factor is that the directions made the word problems easier to understand and presented an interesting challenge to the student. Since the directions were read aloud, the students may have viewed the graphics and story in a more motivating way and understood the questions as more of a challenge. Even though they were required to apply the knowledge, they scored higher with a challenging request.

Our argument, however, is that part B required students to use higher order thinking and decision making skills and the accompanying story contributed to the increase in percentages correct. We feel that our statistical analysis implies that we can be 95% sure that percentages correct in part B will always be higher than the percentages correct in part A. While we may truly never know exactly why this is so, more research will help us zone in on those reasons.

We also believe that this is an experiment that can be replicated in other classroom settings to highlight the effectiveness of assessments. Teachers and practitioners can assess the level of understanding by using Bloom's Taxonomy to create assessments that test knowledge at various levels. If it is created carefully, it becomes an evaluation for the student with a plan for remediation.

Recommendations

While we find the findings statistically significant, we recommend a larger test population. Also, incorporating observation and surveys will help to find the factors in the increase of scores. In addition, it would be stronger to test several subject areas, topics and grades. The assessment could be expanded to a relatively equal number of non-scenario and scenario-based questions.

As for the recommendations for assessment classroom strategies, teachers should use varied assessments to make sure they are assessing many aspects of the topic and triangulating the results to determine if students are able to apply what they learn (Reeves, 2002). Applying higher order processes to assessments and using Blooms' Taxonomy to create assessments is an effective strategy for teachers (Kastberg, 2003).

Assessments provide a useful form of feedback as well as a measurement for what the student has learned at the end of some set of learning activities. Ideally, these two assessments are aligned to provide the framework for remediation and some degree of support for the school's testing standard accountability (Bransford, Brown, & Cocking, 1999). The use of story-based

assessments allows student's to apply the knowledge in the assessment. Teachers can use this information gathered from an application of knowledge to better determine if the student has learned the material or how they can be assisted with a follow on lesson.

Conclusion

In conclusion, it has become clear that analysis of the assessments is an essential part of the assessment process. In order to gather any conclusion about the data we collected, it was important to investigate the assessments for their ability demonstrate what learning took place. Our study did determine that the presence of story-based assessments did have a positive effect on the student's ability to perform the assessment. However, this investigation has led us to agree with Dr. Reeves' conclusion that various assessments types are useful in triangulating a learning outcome.

Appendix

Appendix A: Research Proposal

TRADITIONAL VERSUS NARRATIVE ASSESSMENT ITEMS:

A QUASI-EXPERIMENTAL COMPARISON

by

SARAH GRABOWSKI and THOMAS MACAULAY

A Proposal Submitted to Professor Thomas C. Reeves

in Partial Fulfillment of the Requirements of EDIT 6900

THE UNIVERSITY OF GEORGIA

ATHENS, GEORGIA

FALL 2003

Introduction

Technology and new types of assessments are being considered in K-12 classroom education more and more. Now with the No Child Left Behind Act that mandates testing, assessment strategies have become even more important to the K-12 classroom. However, multiple-choice tests have been criticized for their lack substance of demonstrating the transfer of knowledge (Abrams, Pedulla & Madaus, 2003). An alternative to these traditional assessment items is testing that includes a story-based scenario. Assessments that contain story-based scenarios engage the student, offer a relevant or real-world situation and enhance the student's interest and motivation to achieve. However, it is not known if story-based scenario test items are as effective in measuring learning as traditional test items. This research project will compare traditional test questions to story-based scenario questions in order to evaluate if a realistic problem engages a student in critical thinking and decision making to solve a problem.

Purpose of Study

Our research will focus on assessment strategies for the classroom. The study will evaluate the effectiveness of traditional test questions such as fill-in the blank test questions versus an alternative assessment tools.

A story-based scenario requires a student to use higher order thinking. Therefore, we propose that students will demonstrate a greater comprehension of the material in the assessment items that include story elements. That is, they will need to apply what they have learned in order to successfully complete the assessment (Kastberg, 2003).

This study is being conducted to determine if there is a greater comprehension based on the type of assessment strategy used in the classroom. The project also hopes to shed light on the purpose and objective of an assessment strategy, in general. A sound assessment strategy evaluates the learning achieved by the student, the ability to apply those concepts and the effectiveness of the instruction (Senk, Beckmann, & Thompson, 1997). The strategy used to determine the assessment should also consider the level of instruction delivered and match instruction accordingly.

Our research proposal is socially responsible with respect to the investigation of these important new evaluations of learning. It is necessary for new classroom techniques to be tried

and investigated in order to allow others to realize their mass appeal and effectiveness. Teachers and researchers investigating the new technology-based assessment strategies are exploring new classroom techniques for use by others (Bocij & Greaskey, 1999). Building a knowledge-base of classroom experiences that are successful as well as those that are failures is important work. It is necessary work if new assessment strategies are to be readily accepted and easily delivered by practitioners.

Literature Review

The literature reviewed focused on the studies and theory used to develop assessments. We researched the various impacts assessments have on the student as well as different assessment strategies. Much of our literature review revealed to us the amount of control and anxiety levels a student experiences during the assessment and throughout the course. Technology course enhancements and new assessment methods were reviewed. These new computer-based options are being implemented with varying success, but break ground into new possibilities for assessment. Existing research evaluates the effectiveness of these assessments on various issues such as comfort, relaxation, anxiety, time, and scores (Butler, 2001). Many studies in our literature review focused their attention on strategies to implement new assessment technologies to improve existing courses, not a wholesale change in the course delivery.

As for the theory portion of our literature review, it was necessary to draw upon some of the effective learning taxonomies and learning theory to evaluate the topic of assessment. Some of the topics that we evaluated in our literature review were Bloom's taxonomy, Constructivism, and Montessori based methods (Elkind, 2003). These topics are important to our research because they are widely used as alternatives to instructionist teaching methods. The literature review also raised the fundamental question of the purpose of assessment tools (Henderson, 2001). This thought will help guide our research goals in determining the effectiveness of the teaching, the measure of comprehension by the student and their ability to apply that knowledge (Byers, 2002).

Methodology

At the core of our efforts, we are investigating whether learning theories, such as Bloom's Taxonomy and Constructionism, can assist in building better assessment strategies and

techniques. We are interested in using assessments that demand higher-order thinking as evidence that learning is taking place. When this evidence is present, it is not a question of if the students understand but if they are able to apply the information. Since we want to investigate the effectiveness of traditional tests questions versus an alternative assessment tools, the selection of quantitative methods will be the most appropriate.

Our research will be conducted at the Branch Christian Academy in Lawrenceville, Georgia, specifically the second and third grade mathematics class. All second and third grade students will attend the same two to three day lesson where they will learn how to read the time from and draw the time on an analog clock.

A two-part assessment will be designed for this research project. Special attention will be given to the measure of criterion validity of the assessment. All students involved will first take assessment part A and then immediately take assessment part B. Part A will test students on their knowledge of time using traditional fill-in the blank questions. There will be four to six questions. Each question will have approximately three sections. Section one will ask to read the time from the clock; section two will ask to write that new time given a certain amount of time has passed or a certain amount of time earlier; section three will ask to draw the new time on a new clock.

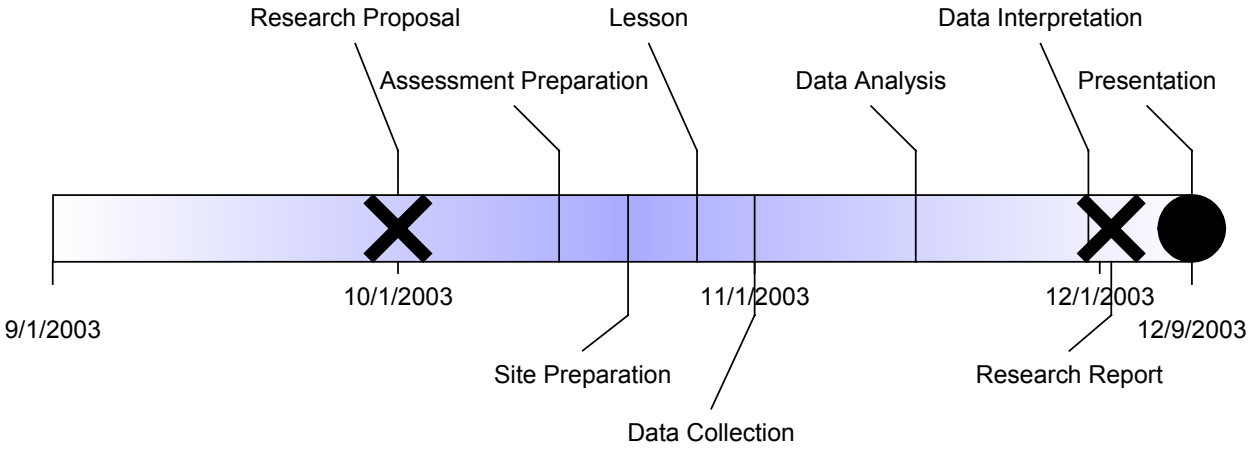
Part B, however, will use a story-based scenario to ask questions similar to those found in part A. First the teacher will read the zoo field trip scenario to the students. Then students will be asked to answer the three to five accompanying questions based on the scenario. Questions in part B will address the same time intervals as the questions in part A. This method will help compare the results by using similar content.

Collecting the student-written responses will generate data. Responses will be in the form of written time (xx:xx), amount of time in minutes and drawing the hour and minute hands on an analog clock. Students will also be encouraged to draw a picture based on the scenario will listening to the story. Privacy will be respected by numbering of the assessments where gender is the only item requested from the student's identity.

For every section of every question, the data will be judged to fit into one of three possible categories: correct answer, incorrect answer, or not attempted/ left blank. Totals will be found for each category of every section of every question. For example, there were twenty responses total for question 1, section a: thirteen correct, five incorrect, and two not attempted.

Comparisons will then be made between the questions designed to be similar in parts A and B. Statistical methods will be used to seek out any correlations between the two assessment methods. Charts and diagrams will also be constructed to show the any visual relationships.

Timeline



Task to Complete	Estimated Amount of Time Needed	Target Date for Completion
Research Proposal	4 weeks	October 1, 2003
Assessment Preparation	2 weeks	October 15, 2003
Site Preparation	1 week	October 21, 2003
Lesson	3 days	October 27, 2003
Data Collection	3 days	November 1, 2003
Data Analysis	2 weeks	November 15, 2003
Data Interpretation	2 weeks	November 30, 2003
Research Report	4 weeks	December 2, 2003
Presentation of Report	1 week	December 9, 2003

IRB Application

Student Application to Engage in Course-Directed Human Subjects Activity*

Student's Name: Sarah Grabowski and Thomas Macaulay **Course:** EDIT6900

Instructor: Dr. Thomas C Reeves **Date:** November 8, 2003

Project title (be descriptive): Traditional vs. Narrative Assessment Research

1. Project description (purpose, research design/perspective, and general overview of study):

The research is focused on classroom assessments. The purpose of the project is to evaluate the effectiveness of traditional fill in the blank tests versus an alternative assessment with a story. The study will evaluate whether a student will understand a lesson based on telling time if they are using higher order thinking to attain the answer.

2. Research Procedures (explain, in detail and in order, the steps in your study and what the participants and the researcher will be doing in each step):

The classroom teacher will hand out a two-page assessment to each student. The first page is a fill in the blank question on the topic of time. The second page will be a scenario for the student to read along with the teacher and a blank area for the students to put their answer. Each student will be asked to put a student number on the top of his or her page.

Next, they will be instructed to answer the fill in the blank questions on page one. After they have completed page one, the classroom teacher will read the scenario aloud to the class. They will then be asked to perform the task. The task is to draw a representation of the scenario and answer the questions.

3. Materials (describe all materials that participants will use, including any informal or formal tests that will be administered)

The materials consist of the two-page assessment and the instructions for the classroom teacher.

4. Interview Protocol (list the questions that you will ask or the questions that you will use to guide your interviews):

- Fill-in the blank questions about what time it is
- Draw the hands on the clock
- How long does it take to get to the zoo?
- What time will the animals be fed?

5. Confidentiality/anonymity (explain how you will make certain that all information about participants will remain confidential or anonymous: e.g., group data only, using pseudonyms for name of participants and locations, plans for storage, procedures for removing all traceable identifiers from data, etc.)

The classroom teacher will assign each student a student number. The assessment will be handed out with the student number, group number and sex on each page. There will be no name to correspond with the student number. However, if the classroom teacher wants to maintain this they can.

6. Certifications (Please initial each to indicate that you understand that that you will comply):

My study will involve only the voluntary participation of the subjects involved.

My study will involve only "minimal risk" or less than minimal risk to participants, which is the "probability and magnitude of harm that is normally encountered in the daily lives of healthy individuals".

This study is NOT part of a graduate/Master's thesis or doctoral dissertation.

It is not my intent to publish the results of this project.

I will follow all school and school district policies regarding the conduct of research, completing a research district research proposal if so required (Note: Clarke County requires School Board Approval for research conducted in connection with UGA, which includes course-related research. See Clarke County Guidelines)

When necessary**, I will obtain informed consent from the parents of minor participants in my research and written assent of the minors themselves.

I will obtain authorization from every institution that is involved in this project (e.g., schools, hospitals, social service agencies, businesses, etc.).

Student's Signature _____ Date_____

Student's Signature _____ Date_____

Instructor's Signature _____ Date_____

*See Section XIII, "Course Directed Human Subjects Activity" in the University of Georgia Institutional Review Board Guidelines for conducting Research with Human Subjects"

**The IRB requires written consent for interview research, observation research when individuals are being targeted as subjects, and for studies requiring information from a student's permanent file. Phone interviewers should use a phone script to obtain verbal consent. Mailed surveys should be accompanied by a cover letter that contains all the necessary information for informed consent. It is the researcher's responsibility to follow local policy in regards to parental consent for classroom-based studies that include an entire class or groups of students and which involve normal kinds of instructional activities. Students should refer to the required format for consent forms in the Guidelines for Investigators that are available on the web.

References

Abrams, L., Pedulla, J., & Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*(1), 18 – 29.

Bocij, P. & Greaskey, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology, 1*(1). Accessed September 13, 2003. http://www.ao.uiuc.edu/ijet/v1n1/bocij/index.html

Butler, D. (2003). The impact of computer-based testing on student attitudes and behavior. *The Technology Source, v. January/February.* Accessed September 14, 2003. http://ts.mivu.org/default.asp?show=article&id=1013.

Byers, C. (2002). Interactive assessment and course transformation using web-based tools. *The Technology Source, v. May/June.* Accessed September 14, 2003. http://ts.mivu.org/default.asp?show=article&id=9388.

Elkind, D. (2003). Montessori and constructivism. *Montessori Life, 15*(1), 26 – 29.

Henderson, T. (2001). Classroom assessment techniques in asynchronous learning networks. *The Technology Source, v. September/October.* Accessed September 16, 2003. http://ts.mivu.org/default.asp?show=article&id=908

Kastberg, S. (2003). Using Bloom's taxonomy as a framework for classroom assessment. *Math Teach, 96,* 402 – 405.

Leedy, P. & Ormrod, J. (2001). *Practical research and design*. Upper Saddle River, New Jersey: Merrill Prentice Hall.

Senk, S., Beckmann, C., & Thompson, D. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187 – 215.

Appendix B: Teacher Instructions

The purpose of this assessment is to determine if there is a difference student performance among various testing strategies. Please hand a copy of the assessment to each student.

For Part A on page 1, ask the students to answer the four multi-part exercise questions. Please tell them to wait until all students are finished before turning to the next page. Wait until all students are done with Part A and then ask the students to turn to page two.

For Part B on page 2, read the story aloud to the students and ask them to answer the three questions. Encourage students to draw a picture, while listening to the story about what is happening.

Each assessment has been given a number in order for the results to be blind. Please ask the students to write their gender next to the number at the top of the paper. If you prefer, you may also ask the students to put their name on the top of the page. Their names will not be used for research.

There are 40 two-paged stapled assessments. Also enclosed is an extra copy in case you need to make extras.

Thank you very much for participating in this research study. If you have any questions or suggestions, please contact us.

Thank you again,

Sarah Grabowski & Thomas Macaulay

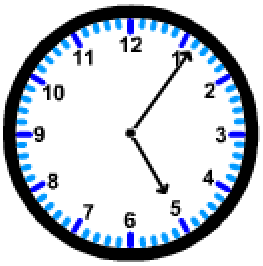
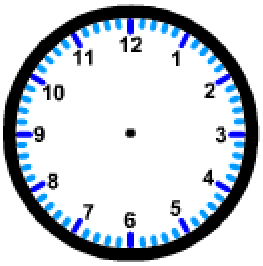
Appendix C: Assessment Part A

Name_____Number_____

Part A

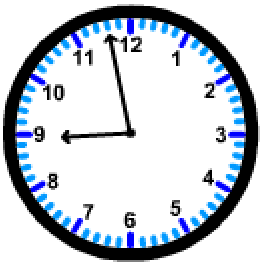
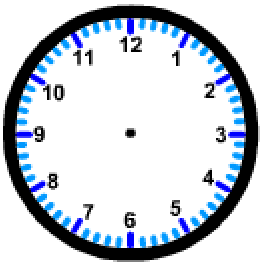
Directions:

Write the time shown. Then draw the hour and minute hands to show each new time. Write the new time.

1.  

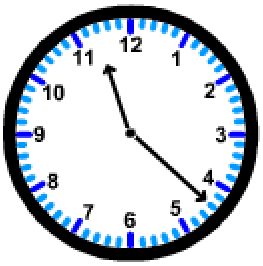
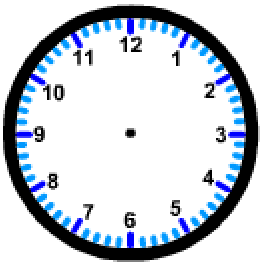
2 hours & 45 minutes later

_____ _____

2.  

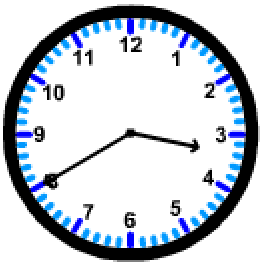
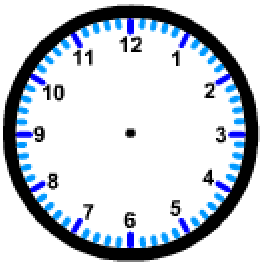
15 minutes earlier

_____ _____

3.  

11 hours earlier

_____ _____

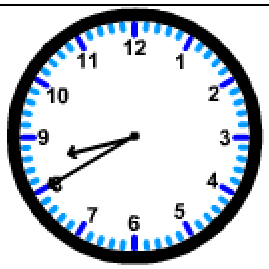
4.  

30 minutes later

_____ _____

Appendix D: Assessment Part B

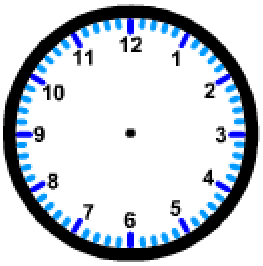
Tom and Sarah are going to the zoo today. Tom looked up at the clock and told Sarah that it was 8:40 AM. Sarah said that they would need to leave in 30 minutes so that they will get to zoo at 10:00 AM.



How long does it take to get to the zoo?

_____ Minutes

Tom and Sarah get to the zoo at 10:06 AM and want to visit the elephants, tigers, zebras, and polar bears first. All of the animals will be fed in 2 hours and 45 minutes.



Draw the hands on the clock to show the time.

What time will the animals be fed?

___ ___ : ___ ___

Write the time here

Draw a picture of Sarah and Tom feeding the animals.

Page 2

Appendix E: Data Collected

	Gender	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C	5A	5B	5C	Total 1	Total 0	Total blank
1	M	0		0	1		0	1		0	1		0	1	1	1	6	5	4
2	F	1	0	0	0	0	0	1	0	0	1	1	0	1	1	1	7	8	0
3	F	1	0	0	0	0		1	1		1	1		1	1	1	8	4	3
4	F	1	1	1	1	0	0	1	1	1	1	1	1	1	1		12	2	1
5	F	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	12	3	0
6	F	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	12	3	0
7	M	1	0	0	0	0		0	0		1	0		1	1	1	5	7	3
8	M	1	0		0	0		0	0	0	1	0		1	1	1	5	7	3
9	M	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	6	9	0
10	M	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	3	12	0
11	M	1	1	0													2	1	12
12	F	1	0														1	1	13
13	F	1	0	0	0												1	3	11
14	F	1	0	0	0	0	0										1	5	9
15	M	1	0	1	1	0	0	1	0	0	1	0	0	1	1	1	8	7	0
16	M	1	1	1	0	0	1	1	0	0	1	1		1	1		9	4	2
17	M	1	1		0	0		1	1		1	0		0	1	0	6	5	4
18	F	1	1	1	0	0	0	0	0	0	0	0	0	1	0		4	10	1
19	M	1	1		0	0		1	0		1	0		1	0	0	5	6	4
20	F	1	1	1	0	0		1	0	0	0	0		1	0	0	5	8	2
21	M	1	0	0	0	0	0	0	0		1	0		1	1	1	5	8	2
22	F	1		1	0	0		1	0		1	1	1	1	1	1	9	3	3
23	F	1	1	0	0	0	0	1	0	0	1	0	0	1	1	1	7	8	0
24	F	1	0	0	0	0		0	0	0	1	1	1	1	1	1	7	7	1
25	F	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	10	5	0
26	M	1	1	1	1	0	0	1	0		1	1		1	1		9	3	3
27	F	1	1	1	1	1	1	1	0		1	1		1	1	1	12	1	2
Total 1		22	13	11	7	3	3	17	4	3	21	9	4	21	19	15	172		
Total 0		1	8	9	16	19	12	6	18	12	2	13	9	2	4	4		135	
Total blank		0	2	3	0	1	8	0	1	8	0	1	10	0	0	4			38

Appendix F: Legend for Table of Data Collected

LEGEND

1 = correct response
0 = incorrect response
_ = left answer space blank

1A = Part A, Question 1, Answer 5:06
1B = Part A, Question 1, Answer 7:51
1C = Part A, Question 1, Clock of 7:51
2A = Part A, Question 2, Answer 8:58
2B = Part A, Question 2, Answer 8:43
2C = Part A, Question 2, Clock of 8:43
3A = Part A, Question 3, Answer 11:22
3B = Part A, Question 3, Answer 12:22
3C = Part A, Question 3, Clock of 12:22
4A = Part A, Question 4, Answer 3:40
4B = Part A, Question 4, Answer 4:10
4C = Part A, Question 4, Clock of 4:10
5A = Part B, Question 1, Answer 50
5B = Part B, Question 2, Answer 12:51
5C = Part B, Question 3, Clock of 12:51

*Grey shaded cells are considered to be invalid and therefore not used in totaling.

Appendix G: Statistical Analysis

Two-Sample T-Test and CI: 5B correct, 1B correct

```
Two-sample T for 5B correct vs 1B correct

             N   Mean  StDev  SE Mean
5B correct  23  0.826  0.388    0.081
1B correct  27  0.519  0.509    0.098

Difference = mu (5B correct) - mu (1B correct)
Estimate for difference:  0.307568
95% CI for difference:  (0.052049, 0.563087)
T-Test of difference = 0 (vs not =): T-Value = 2.42  P-Value = 0.019  DF = 47
```

Two-Sample T-Test and CI: 5C correct, 1C correct

```
Two-sample T for 5C correct vs 1C correct

             N   Mean  StDev  SE Mean
5C correct  18  0.778  0.428     0.10
1C correct  24  0.458  0.509     0.10

Difference = mu (5C correct) - mu (1C correct)
Estimate for difference:  0.319444
95% CI for difference:  (0.026600, 0.612289)
T-Test of difference = 0 (vs not =): T-Value = 2.21  P-Value = 0.033  DF = 39
```

Two-Sample T-Test and CI: 5A correct, 4B correct

```
Two-sample T for 5A correct vs 4B correct

             N   Mean  StDev  SE Mean
5A correct  23  0.913  0.288    0.060
4B correct  22  0.409  0.503     0.11

Difference = mu (5A correct) - mu (4B correct)
Estimate for difference:  0.503953
95% CI for difference:  (0.253781, 0.754124)
T-Test of difference = 0 (vs not =): T-Value = 4.10  P-Value = 0.000  DF = 33
```

One-Sample T: 1B wrong, 5B right

```
Variable            N      Mean     StDev   SE Mean         95% CI
1B wrong, 5B right  23  0.304348  0.470472  0.098100  (0.100900, 0.507795)
```

One-Sample T: 1C wrong, 5C right

```
Variable            N      Mean     StDev   SE Mean         95% CI
1C wrong, 5C right  16  0.500000  0.516398  0.129099  (0.224831, 0.775169)
```

One-Sample T: 4B wrong, 5A right

```
Variable            N      Mean     StDev   SE Mean         95% CI
4B wrong, 5A right  22  0.500000  0.511766  0.109109  (0.273096, 0.726904)
```

References

Abrams, L., Pedulla, J., & Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*(1), 18 – 29.

Bransford, J., Brown, A. & Cocking, R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press. Accessed December 8, 2003. http://www.nap.edu/html/howpeople1/notice.html

Bocij, P. & Greaskey, A. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology, 1*(1). Accessed September 13, 2003. http://www.ao.uiuc.edu/ijet/v1n1/bocij/index.html

Butler, D. (2003). The impact of computer-based testing on student attitudes and behavior. *The Technology Source, v. January/February.* Accessed September 14, 2003. http://ts.mivu.org/default.asp?show=article&id=1013.

Byers, C. (2002). Interactive assessment and course transformation using web-based tools. *The Technology Source, v. May/June.* Accessed September 14, 2003. http://ts.mivu.org/default.asp?show=article&id=9388.

Elkind, D. (2003). Montessori and constructivism. *Montessori Life, 15*(1), 26 – 29.

Henderson, T. (2001). Classroom assessment techniques in asynchronous learning networks. *The Technology Source, v. September/October.* Accessed September 16, 2003. http://ts.mivu.org/default.asp?show=article&id=908

Kastberg, S. (2003). Using Bloom's taxonomy as a framework for classroom assessment. *Math Teach, 96,* 402 – 405.

Leedy, P. & Ormrod, J. (2001). *Practical research and design*. Upper Saddle River, New Jersey: Merrill Prentice Hall.

Reeves, T. (2002). Keys to successful e-learning: Outcomes, assessment, and evaluation. *Education Technology, 42*(6), 23 – 32.

Roeser, R. & Shavelson, R. (2002). The concept of aptitude and multidimensional validity revisited. *Educational Assessment, 8(2)*, 191 – 205.

Senk, S., Beckmann, C., & Thompson, D. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187 – 215.

Wang, S. & Han, S. (2001). Six c's of motivation. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology.* Accessed October 22, 2003. http://itstudio.coe.uga.edu/ebook/6csmotivation.htm