Creating a Moral Climate in the Classroom

Every generation seems to feel that the current generation is worse morally than previous generation. It may or may not be true, but what teachers sense in particular is that the classroom has changed from their experience as a student. Teachers want the best for their students, but often feel the roadblock of the “Wall of Separation”, if they are people of faith, or just the pressure of students who seem to have lost moral grounding.

The Nyack College graduate programs have required a course in Character Education to help address this moral gap. The course does not provide all the answers, but it does provide a full “tool belt” that the teacher can use in the classroom. The course also gives each teacher candidate ample opportunity to think about their classrooms and how to develop a climate where “character” is a vital part of the curriculum.

One of the most powerful ideas comes when teachers realize they are creating a moral climate in the classroom, intentionally or unintentionally. Teachers also realize they can use that climate to create a moral community, that not only involves students, but parents and guardians as well. The moral climate of the classroom is the grounding agent or source of authority for the student’s character development. Yes, some students bring their own source of authority like parents, faith and faith community. However, the classroom is where all this comes together into a working community.

It is often said that an excellent teacher is an excellent decision-maker. However, this statement often is interpreted as applying to content and pedagogical knowledge. However, Nyack School of Education, as a Bible-based program places heavy emphasis on the third element of excellent teaching: Dispositions. As Christian educators, Nyack teacher candidates recognize that the Bible places heavy emphasis on Christian Character in all areas, personal and professional. The Character Education course continues to be a major distinctive of a Nyack graduate program, recognizing that Christian character is reflected by our graduates every day in the classroom.

Teacher candidates learn to make decisions from their own background in creating a new moral atmosphere and community as an immediately noticeable distinctive in their classrooms. If we are to
Learning Curve

return our society to the Christian roots upon which it was founded, the Christian teacher as a role model of Christian character is an obvious place to begin.

The use of the Orff Approach of teaching music and the musical improvements associated with it with students with disabilities.

Eduardo N. Duque
EDG500, December 4, 2013

Abstract
The issue of actively engaging students with disabilities in a group setting has been a concern for many special education music teachers. The purpose of this quantitative, quasi-experimental study was to determine if there was a difference between using the Orff method as opposed to using what other special education music teachers use for teaching music, the Kodaly method. There will be a difference in singing and musical performance between the existing method of teaching music, the Kodaly method, and the new non-existent method, the Orff approach. Students exposed to the Orff approach received instruction about the instruments and a t-test was conducted to distinguish the difference between the pre-test and the post-test. Since the Group Statistics box revealed that both the Means for before and after the Orff method was introduced were relatively the same, we concluded that the participants’ singing did not significantly improve. The results of this study provided important information about the use of the Kodaly and Orff methods on students with special needs.

Introduction
There is a great concern on behalf of music educators to be informed on how to adequately teach music for those individuals with a range of disabilities in a self-contained or mainstreamed classroom. Hourigan states that, “there is little research into the effects of special needs experiences for preservice teachers in music education” (2007, para 4). Almost anyone with a valid music teaching certificate can accept a special music educator position without receiving any higher educational training for special music education. Music teachers who are instructed and educated on what to do with a special needs child in their classroom, will have a better attitude as to accepting that child into their classroom. Those music educators that have been taught music pedagogy and have incorporated students with special needs in their area of undergraduate studies will have a better success rate of achieving goals with these exceptional students. Those music teachers will be better prepared and positioned to achieve a favorable outcome with students that are struggling or are disabled.

This quantitative, quasi-experimental study was created with the Discovery Learning principle in mind, to inform music educators of the need of providing explicit instruction in the principles that are to be learned in any method of teaching music in a special education setting or a mainstreamed music classroom. Keller (1974) informs “Nothing requires more meticulous preparation than guiding and supervising lessons in discovery and improvisation” (p. 23) The purpose of this experiment is to find out whether the Orff method of teaching music is significantly contrasting from the pre-existing method the students have been exposed to, the Kodaly method. “The Kodaly philosophy, which originated in Hungary, is based on the assumption that the most effective
Arte into music is through the instrument available to us all - the human voice; The Orff approach, developed by composer Carl Orff just after World War II, uses rhythm in speech (word patterns, rhymes) and movement to explore the musical basics of note values, rhythm, phrasing, dynamics, etc. In contrast to Kodaly, however, Orff is more experiential (through actually playing instruments), and most of the music itself is improvised” (Campbell, 1997, para. 19).

The research hypothesis used in the study was that there will be a difference in musical performance and singing of special needs students between the existing method of teaching music, the Kodaly method, and the new non-existent method, the Orff approach. Does the Orff method of teaching music improve singing and its use of musical concepts? Is there a difference in singing with the xylophone and singing without the xylophone? The independent variable is the Orff teaching method that is being introduced after the pre-test. The dependent variable is the singing scores from the pre-test and the post-test.

Limitations of this research are the limited convenience sample, which was used for this research. There is also a limitation in that the pre-test and the post-test will be the same singing test used to measure any difference between the two methods and so there is no way to assure accurate post-assessment because students had already seen the pre-test. Teacher observation for the pre and post-tests is a reliability issue in that there is no real way to measure consistency in the teacher observation.

Literature Review

"Since there are few guidelines present for the special music educator, much experimentation and research still is needed” (Graham, 1975, p. 13). There is little to no research on the music teaching methods used in special educational situations. Most seem to give you activities and advice on how to teach special education children, but almost no literature has conveyed light on what teaching method or combination of teaching methods seem to work best for special music education. This comes to show that the majority of music educators are not adequately instructed on what to do or how to teach in an inclusive setting involving a special needs child, let alone a self-contained special education classroom. Hourigan (2007) communicates that, “there is a staggering absence of special needs preparation in preservice music education [and that] music teachers are under increasing demand to include students with disabilities in music classrooms” (para. 2). "Music educators are now faced with the problem of teaching the handicapped child, an issue that has been long ignored by members of the profession, [however] mainstreaming may come as a blessing in disguise forcing music educators to think and plan innovatively to develop better and more effective means of musically educating all of the children assigned to him or her” (Graham & Beer, 1980, p. 15).

The importance for music teachers to be trained for special music educational settings is an imperative veracity that cannot be avoided. "In recent years, music teachers in the school system have seen increasing numbers of students who have various mental and physical disabilities in their classrooms, [and] music teachers are under increasing demand to include students with disabilities in music classrooms” (Hourigan, 2007, para. 2). There is a great need not only for music educators, but educators in a whole to be trained in the area of special education considering the growth and outburst of students with disabilities.

The significance of the Orff approach in the classroom today is imperative. "All children have the need for expression through music, but for the
handicapped child, the physical, emotional, and intellectual benefits that can be derived from music make it an essential part of classroom activities” (Zinar, 1983, pg. 7).

With the Orff approach, any student will able to participate at any level or stage of their cognitive development, or academic performance. Keller informs his readers that the Orff method “does not restrict genuine talent, neither special music ability, nor previous training, instead, it allows children to try out and confirm their own abilities without sacrificing a common source of musical sound offering worthwhile activities for children with greater and lesser gifts, so no individual is pushed ahead too fast or held back unnecessarily” (1974, pg. 5). It has been stated that students with disabilities are at a disadvantage while learning in a group setting. For so long, educators sought after the correct method of teaching music to individuals with disabilities in a group setting. Although not much literature states any theories or the difference between using different music teaching methods to students with disabilities, Goodkin (2001) argues that “What distinguishes Orff from the other two [Dalcroze & Kodaly] is the experience of improvisation at the heart of the matter” (para. 12). A theory is that the students will enjoy playing the instruments, the improvisational aspect to it, and the change in instruction due to the fact that they have not seen or done anything like this before.

**Method**

This action research carried out a pre-test and post-test to judge the differences between the pre-existing method students have been exposed to and the new non-existing method by being executed in a convenience sample. The participants consist of a BOCES CABAS classroom made up of nine students from grades kindergarten to second grade with a range of disabilities composed of mild autism, apraxia, and other health impairment. All of the students were used in this study.

The instrumentation was composed of a singing test that examined the performance of the student singing a song. The singing test was comprised of a nursery rhyme, well known as “Twinkle Twinkle Little Star”. This required the accompaniment of the teacher’s guitar without the teacher phrasing the words to the student. The task asked students to sing a song they knew but have not yet mastered, or learned in some cases. The singing test was accomplished individually by all the students while the other students were at play in the toy area or at work with their cooperating teacher. The singing test assessed the percentage of the song they sung, their tone, if they were singing in tune, if they were on the beat, and their dynamic by rating them on a scale of one through ten. The higher the rating, the better they were singing in key, on the beat, and improvement dynamic.

To collect the data, the teacher would use observations, and record the data on a chart containing the different musical concepts being assessed after every child had attempted to sing the song individually. At the end of the pre-test, the teacher would gather all the data pertaining to every child, and create an average by finding out the mean.

To find out whether singing with the xylophone and singing without the xylophone was significantly different, the teacher first prompted the students to sing the song “Twinkle Twinkle Little Star” while they played that melody on the xylophone. The teacher then rated each student’s singing while they played and sung using the same chart to assess the pre-test. Second, the teacher prompted the students to sing without the xylophones by taking away their mallets. In the same manner as before, the teacher
assessed the students’ singing by recording the data. With all this data pertaining to every child, the teacher created an average by finding out the mean.

One way to exploit this experiment was to conduct a t-test to evaluate the student’s current state without the change in instruction, and a post-test to assess if the change in instruction resulted in a significant difference. The t-test took the overall grade the students got from the pre-assessment and compared them to the overall grade the students got from the post-assessment. The overall grade the students got from singing with the xylophone was then compared to the overall grade the students got from singing without the xylophone by creating a second t-test.

**Results**

An independent-samples t-test was conducted to compare singing performance before the Orff method was introduced and after the Orff method was introduced. There was not a significant difference before the Orff method was introduced (M=5.00, SD=3.23) and after the Orff method was introduced (M=5.20, SD=3.52) of singing; t (16) = -.125, p = .902. There is no statistically significant difference from the pre-test and the post-test. These results suggest that the Orff approach does not have a significant effect on singing performance. Specifically, this study suggests that there was not much difference from the Kodaly method and the Orff method of teaching music.

The Sig. (2-tailed) value is .902 and this value is greater than .05. Because of this, we can conclude that there is not a statistically significant difference before and after the Orff method was introduced. Since the Group Statistics box revealed that both the Means for before and after the Orff method was introduced were relatively the same, we can conclude that participants’ singing did not significantly improve.

A second independent-samples t-test was conducted to compare singing performance with and without the xylophone. There was not a significant difference singing with the xylophone (M=2.77, SD=3.44) and singing without the xylophone (M=3.18, SD=3.35); t (16) = -.257, p = .801. There is no statistically significant difference from singing with the xylophone and singing without it. There was a minuscule improvement but not enough for it to be significant. These results suggest that singing with or without a xylophone does not have a significant effect on singing performance.

### Pre-test and Post-test

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>VAR00001</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>1.00</td>
<td>9</td>
<td>5.0000</td>
<td>3.23844</td>
<td>1.07948</td>
</tr>
<tr>
<td>GPA</td>
<td>2.00</td>
<td>9</td>
<td>5.2000</td>
<td>3.52952</td>
<td>1.17651</td>
</tr>
</tbody>
</table>

### Singing With and Without the Xylophone

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Singing</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
</table>
Discussion

The results of this study provided important information about the use of the Kodaly and Orff methods on students with special needs. First it shows that these two methods are not too far apart when achieving the same common goal on students with disabilities. And second, it has displayed that there is not a statistically significant difference in singing performance with the xylophone and singing without the xylophone. Most of the students could only accomplish singing or playing but not both at the same time. Although, this study did not obtain any statistically significant difference, it has brought light on the fact that more needs to be said about the music pedagogy used on students with special needs. The results point out that the Kodaly and the Orff approach are both effective ways of teaching music and achieving similar results. This shows music educators that both approaches of teaching music are successful methods. This study provides many further options for research. A similar study could be done to determine the level of engagement the Kodaly and Orff approach offer. Although this study did not show a significant difference between the Kodaly and the Orff approach of teaching music, singing performance did stay relatively the same.

References


**The Difference in Gender and Reading Level on Types of Assessment Questions**

*Stephanie Leone*

*EDG 500, December 2013*

**Abstract**

The purpose of the study was to determine the differences between gender and reading levels on different types of questions. The two question types studied are multiple-choice and free recall questions. The research hypothesis for both variables is that there will be difference between groups: girls will perform better on free recall questions, as well students that are above average in reading. The study was conducted with 6th grade students at a private school in New Jersey. The students took a chapter assessment at the end of a three-week science unit that included both multiple-choice and short answer questions. The results of the study were that there was no difference between genders on either question type, but there was a difference between reading levels. Students that are below average in reading ability did not perform as well as above average students on the free recall questions.

**Introduction**

Males and females differ in many aspects of academics; including learning style, preferences, and test taking abilities. The differences experienced by males and females can also create testing bias in the construction of the assessment. Teachers have to be aware of bias in creating tests: reducing bias increases reliability and validity of the assessment (Gareis & Grant 2008). To remove bias, test questions have to be designed to meet both genders’ learning styles and needs. This requires study into different test questions and the two genders’ performance on each.

Science and literacy are two terms that need to be connected in the classroom. Students should be reading and writing in science to engage in inquiry, and thinking and reasoning like a scientist (Glasgow, Cheyne, Yerrick, & Keeley 2010). But a student’s reading level can impact how they comprehend scientific literature and are able to express his or her thoughts on it. Assessments must also be designed to eliminate bias for students at different reading levels. A study is required to test which questions are difficult or easy for students in the different levels to successfully answer.

**Problem**

The purpose of this study is to determine gender and reading level differences on free recall assessments vs. multiple-choice assessments.
Hypotheses

There will be a difference in scores of males and females on assessments with different types of questions. It is predicted that one gender will surpass the other in either multiple choice or free recall questions based on the literature. This research is intended to find the questions preferred by each gender in order to aid in eliminating bias in future assessments. The null hypothesis is that there will be no difference between genders on the types of questions.

There will be a difference between low and below average reading level students and average and above students on assessments with different types of questions. It is predicted that low average and below average students will have lower scores on both types of questions based on the literature. Due to better reading performance, higher level reading students will score higher in both questioning areas. This research is intended to seek deficiencies in creating assessments and implications for future examination questions. The null hypothesis is that there will be no difference between the reading levels and types of questions.

Research Question

Does a difference exist in test taking abilities of the different genders or students placed in lower or higher reading groups, in regards to type of assessment?

Literature Review

Males and females differ in their ability to self-regulate, or to control behavior and work ethic in school. Girls have better ability to sit still longer, delay gratification, and pay attention than boys (Eliot 2010). Girls in elementary school have the better ability to self-regulate: they can sit and complete longer assignments, delaying rewards in order to complete the task. Teachers have to create assessments to accommodate both male and female learning styles, preferences, and differences. Assessments must be designed to enhance future and current learning, to develop self-regulation, autonomy, and better lifelong learning practices (Crisp 2012).

Each gender typically demonstrates a different learning style in the classroom. This begins in early development and play preference in early childhood classrooms. Boys are more active and play with toys that require motion (trucks, balls), while girls are more verbal and prefer toys that require such relational skills (dolls, role-playing) (Eliot 2010). This continues into elementary age, as boys prefer kinesthetic learning activities and girls prefer auditory and visual learning activities. Assessments and instruction have to acknowledge learning styles and preferences, and differentiated accordingly (Wilson 2012).

Males and females have different types of ability and knowledge that impact their performance on assessments. Types of knowledge can be measured through demonstration of process skills or the product of such: females score better with process skills assessment and males score better when a product is desired (Bagele 2000). Male students tend to perform better when a direct answer is required, while female students demonstrate knowledge of the process of retrieving an answer. “Males tend to have the ability to extract spatial and logical relationships independently of the contextual components of the tasks [and] females tend to explain the meaning of a concept from connotative content” (Bagele 2000). Shorter question types such as multiple-choice or matching require such knowledge of relationships, while free recall and essay responses require explanations. Males would tend to score better on multiple-choice, while females short answer or essay questions.
This knowledge and ability difference is allocated to societal differences between males and females. Girls develop verbal skills sooner than boys, and reading and writing accounts for girls’ ability to explain concepts and demonstrate process skills better. Girls are statistically better at reading and writing because they read and write more often in daily activity (Eliot 2010). "Only 25 percent of teenaged boys around the world cite reading as one of their favorite hobbies, compared with 45 percent of teenage girls" (Eliot 2010). Girls report reading and writing more in everyday activity, which is why they tend to do better on tasks that require extensive reading and long writing explanations.

In regards to science achievement, males have historically performed better on assessments and the gap in achievement grew as grade level increased (Bursal 2013). In a study conducted by Kohlhaas, Chu, & Lin, male fifth grade students performed better on science assessments (2010). But a different shift towards better female performance is occurring in recent studies. Females are performing better in science due to a shift in attitude, where their attitudes towards science and academic studies are more positive than males (Bursal 2013). The female students will be expected to score as highly as the male students on science assessments.

A student’s reading level has an impact on their proficiency in science and ability to answer questions on a science assessment. In a study conducted by J.G. Cromley, students were given a science assessment with both multiple-choice and short answer questions and a high correlation was found between students’ reading comprehension level and science proficiency (2009). The higher the reading comprehension level of the student, the larger the correlation between reading and science proficiency (Cromley 2009). The ability to interpret scientific readings is important in learning scientific concepts, and students that have low reading comprehension levels will not be able to perform well on assessments.

Students on different reading levels differ in their level of self-efficacy, which impacts performance on different types of assessment questions. Self-efficacy is a person’s confidence level regarding their ability to problem solve or accomplish a task, and this directly influences their effort and choice of tasks to perform (Solheim 2011). Students in lower reading levels demonstrate low self-efficacy, and have less confidence on different types of assessment questions. The study conducted by Solheim suggested that students with low reading self-efficacy found multiple-choice questions to be more complex than the short answer questions (2011). The number of choices confused low reading level students, and their confidence in choosing the right answer waivered. Low reading level students are expected to perform better on free recall questions rather than multiple-choice.

Procedure

The sample size of the students is 36 sixth grade students from a private Christian school in Northern New Jersey. The students are a mixture of ethnicities, as well as genders and reading abilities. There are 21 girls in sixth grade and 16 boys. 49% of the students are Hispanic, 14% are African American, 11% are Caucasian, 11% are biracial, 8% are Portuguese or Brazilian, and 5% are other ethnicities. 16 students are low-average or below reading level, and 21 are average or above reading level. The sample was chosen conveniently; the researcher is the sixth grade science teacher at the school.

The students will be taught a science unit over a three-week period, and then given an assessment at the end. Instruction given will be a variety of methods; including reading and response activities, laboratory
activities, and problem solving strategies. The assessment contains a mixture of multiple-choice and free recall questions (or short answer).

Data Analysis

Multiple-choice responses will be scored out of 16, as well as the free recall. The test contains 16 multiple-choice questions and 4 free recall responses worth 4 points each. After the assessment is scored, data will be collected and analyzed using a t-test. The scores of each gender on each question type will be recorded, as well as the scores of each question type for the students in different reading levels. A t-test will be used to determine the difference in scores between the two groups. Measures of central tendency including mean, median, mode, range, and standard deviation will be calculated, as well as the significance score to ensure confidence in the results.

Findings

An independent-samples t-test was conducted to compare gender and scores on multiple-choice questions and scores on free recall questions. There was not a significant difference in male multiple-choice scores (M=77.27, SD=18.06) and female multiple-choice scores (M=74.52, SD=18.57); t (34) =.442, p=.661. There was no significant difference between male free recall scores (M=72.67, SD=21.057) and female free recall scores (M=66.19, SD=27.596); t (34) =.763, p=.451. These results indicate that there is no difference between male and female students and their scores on different types of assessment questions. Specifically, gender does not indicate what type of score a student will receive on multiple-choice or free recall questions.

An independent-samples t-test was conducted to compare reading level and scores on multiple-choice questions and scores on free recall questions. There was a not a significant difference in average or below students’ multiple-choice scores (M=69.67, SD=20.75) and above average students’ multiple-choice scores (M=79.95, SD=15.12); t (34) =1.72, p=.094. There was a significant difference between average or below students’ free recall scores (M=50.13, SD=23.75) and above average students’ free recall scores (M=82.29, SD=15.92); t (34) =4.91, p=.000. These results show that students that are average or below in reading achieve similar results on multiple-choice assessments as above average reading students. The results also indicate that reading level is a predictor of scores on free recall assessments: students that are average or below receive significantly lower scores on free recall questions than students who are above average.

Conclusion

The research hypothesis for gender differences is rejected and the null hypothesis is accepted. No significant difference is found between the genders on both types of questions. One explanation could be the limitations on the study, including the school and classroom selected. The convenience sample limits the study and its accuracy in predicting true differences between genders. Another explanation could be a new awareness of gender bias in assessment, and the curriculum tests are better suited for both genders and their learning styles. The types of lessons taught could have caused the results, which is an area for further research and study. Another area of further study is to test gender and differences in how they learn and their classroom preferences. This research could lead to the source of gender bias and enable educators to change how they teach their male and female students.

In regards to reading level and types of questions, the research hypothesis is accepted and the null hypothesis is rejected. T-test results displayed no
statistical significance for reading level and question type, but there is practical significance for classroom application. Students that struggle with basic reading skills also have trouble in decision-making, especially when given choices on exams. For example, one student in the sample has an IEP that states her struggle with decision-making and that multiple-choice questions are confusing for her. She scored low on multiple-choice, and there are other students on her academic level that struggled in the same way. Though the results were not statistically significant, educators should be aware of this struggle in order to modify multiple-choice assessments for their lower level reading students.

T-test results for reading level and free recall questions were statistically as well as practically significant. Lower reading level students that struggle with reading comprehension and fluency also struggle with free recall questions. The lack of comprehension causes a lack in ability to fully express ideas. These students had difficulty remembering the scientific concepts, as well as being able to explain them clearly. Areas for further study would include how to help students below level in reading achieve better results on their assessments. Practical solutions would further aid teachers in their pursuit to better their students’ education.

References


Learning Curve


**Test format effectiveness on student learning in an elementary school classroom**

Martha L. Guiza

EDG 500 Educational Research and Statistics

13 December 2013

**Abstract**

Educators have relied on classroom assessments in order to assess their students’ learning gains. However, numerous studies strongly suggest that the mere act of testing someone’s memory is an effective retention tool for the tested material. This study examined whether multiple choice and short answer classroom quizzing could enhance retention of the information tested in elementary age students in social studies. In a 3rd grade general classroom, a social studies unit was divided into six sessions. After each instructional session finished, students took either an open-ended/ fill in the blank, multiple-choice quiz or no quiz but instead were given a study guide. Five class days later, students took a pencil-and-paper final unit test had a total of 40 items whose format were evenly distributed in multiple-choice and open-ended / fill in the blank format. Fourteen participants’ scores were analyzed using a one-way within subjects (or repeated measures) ANOVA was conducted to compare the effects of quiz format (multiple-choice, open-ended/ fill in the blank and non quiz) on the final grade percentage on the final unit test, in order to determine if there exists a difference between quiz format and later retention. The results suggest that quiz format (multiple-choice, open-ended/ fill in the blank and non-quiz) does have an effect on later retention. Three paired samples t-tests were used to make post hoc comparison between the conditions. The results of the paired sample t-tests suggest that when elementary students are tested using multiple-choice quizzes, the retention of this content is greater than if they are tested using open-ended/ fill in the blank quizzes or no quizzes at all, but are instead given a study guide. However, there is no real difference in later retention if students were tested using no quiz or open-ended/ fill in the blank quizzes.

In conclusion, the results suggest that multiple-choice quizzing is an effective retention tool for the elementary classroom.

**Introduction**

Problem

In the past, educators have used testing for assessment purposes, to assign students a grade, to determine their prior knowledge or grouping them according to their abilities. Educators, now more than ever, are being held responsible for their student’s learning gains and have relied on classroom assessments in order to assess their students’ learning. Yet, there are numerous studies that strongly suggest that the mere act of testing someone’s memory will strengthen that memory and thus testing has a powerful effect on long-term retention. Therefore, testing is not only an assessment tool to calculate students learning, but also a retention instrument that
can greatly improve and promote student learning and performance (Hinze & Wiley, 2011; Kang, McDermott, & Roediger, 2007; Roediger and Karpicke, 2005; Roediger, Putnam, & Smith, 2011). Testing is actually a learning event, a vehicle of learning according to Bjork and Bjork (2011). Unfortunately is has been overlooked as a vehicle for learning and only seen as a vehicle for assessment (Bjork & Bjork, 2011). According to McDaniel, Anderson, Derbish, & Morrisette (2007) educational research has ignored the benefits of the testing effect to increase learning in the classroom. Furthermore, the benefits of testing are rarely taught in teacher educational courses or implemented in the general classroom (Kang, McDermott, & Roediger, 2007). In addition, empirical studies on the benefits of testing as a retention tool have been done with undergraduates in laboratories settings (Carrier & Pashler, 1992) and some in college course settings (Hinze & Wiley, 2011; Kang, McDermott, & Roediger, 2007; McDaniel et al., 2007). However, these studies are rarely done in an elementary school setting. Another concern is that laboratory studies give information just once, while in classroom settings learning does not take place like this. Teachers usually give a lesson and then do an activity, assign reading and homework related to the content learned (McDermott et al., in press). Also, laboratory studies usually give a short retention interval ranging from a couple of minutes, hours or 1 week. These short intervals are unlike the longer ones seen in a classroom setting where there is a final unit test at the end of a few weeks or months (Marsh, Fazio, & Goswick, 2012; McDaniel et al., 2007; McDermott et al., in press).

McDermott, Agarwal, D’Antonio, Roediger, & McDaniel, (in press) have concluded a study on the benefits of multiple choice and short answer quizzing in middle and high school classrooms. The only study where participants were elementary students was in Marsh, Fazio, & Goswick (2012). The participants of this study were second graders and the laboratory study focused on the potential benefits of multiple choice quizzing, negative testing effect and feedback on retention (Marsh, Fazio, & Goswick, 2012). However, there is no present research comparing the effects of multiple choice and short answers quizzing on elementary school students’ learning in a classroom setting.

**Purpose**

The purpose of this study was to investigate the effects of quiz format on elementary students’ later retention. The study explored whether, in an elementary school setting, recognition tests (multiple choice tests), production tests (opened end/ fill in the blank tests) or no test (study guide) was more effective on later retention. This later retention was measured using a final unit test which had half the question in multiple choice format and half the questions in short answer/ fill in the blank formats.

**Hypothesis**

The hypothesis of this study is that a difference exists in the third grade students’ later retention between the testing formats of multiple choice and short answer/ fill in the blank quizzing. The null hypothesis is that no difference will exist on the third grade students’ later retention between the testing formats of multiple choice, short answer/fill in the blank quizzing and the study guide.

**Variables**

The discrete independent variable of the study was the format of the quizzing, which consisted of no quiz, multiple choice, and short answer/fill in the blank. The continuous dependent variable is the students’ scores on the final unit test.

**Definition of Terms**
**Testing effect** – The mere act of testing someone on previously studied material leads to better long-term retention than restudying the material or not taking a test. (Carrier & Pashler, 1992; Karpicke & Grimaldi, 2012; Marsh, Fazio, & Goswick, 2012; Roediger & Butler, 2013; Roediger & Karpicke, 2005)

**Generation effect** – is the phenomenon where information that is generated, such as in responding an essay question, it is better recalled than if that same information were simply presented such as in reading or merely studying (Tigner, 1999).

**Multiple-choice format** – Recognition tests that rely on recognition processes (McDermott et al., in press; Roediger & Butler, 2013).

**Short answer/ Fill in the blank format** – Production tests that rely on recall (McDermott et al., in press; Roediger & Butler, 2013).

**Reliability and validity**

The limitations of this research lay in the limited convince sample that was used in this research. The validity of the multiple choice and short answer/ fill in the blank quiz questions have been checked by having these questions come directly from the content taught in the landform lessons’ PowerPoint presentation. The validity of the final unit test has been checked by having these questions come directly from the content taught in the Power Point presentation and from the quizzes and study guides given in class. Furthermore, the final unit test question formats were evenly distributed. The original thirteen multiple-choice questions were presented in the final test as seven multiple-choice and six open-ended questions. The thirteen open-ended quiz questions were presented in the final as six multiple-choice and seven open-ended questions. Finally, the thirteen facts in the study guides were presented as seven multiple-choice questions and seven open-ended questions in the final. The validity of final test scores have been checked by having the study guides as a restudy control so that the effects of quizzing format could be compared.

**Literature Review**

The testing effect is based on findings that suggest that the mere act of testing someone on previously studied material leads to better long-term retention than restudying the material or not taking a test. (Carrier & Pashler, 1992; Karpicke & Grimaldi, 2012; Marsh, Fazio, & Goswick, 2012; Roediger & Butler, 2013; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011). Studies have shown that the testing effect is not due to additional exposure of material because results suggested that taking a test led to better rendition than did restudying the material (Carrier & Pashler, 1992; Roediger & Butler, 2013; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011). Furthermore Roediger and Karpicke (2006) findings suggested that immediate assessment following exposure to the future test material promoted better long-term retention rather than repeated study of the material.

Although no single theory can explain the phenomenon of testing effect, there are three possible theoretical explanations that complement each other that involve the retrieval effort, transfer- appropriate processing and encoding variability (Roediger & Butler, 2013).

Studies have shown that production tests such as a cued recall, short answer and fill in the blank tests, have greater retrieval effort and also produce greater retention than recognition tests, such as multiple choice and true or false assessments. Therefore, a possible explanation for the testing effect is the effort involved in retrieving the studied information (McDermott et al., in press; Roediger & Butler, 2013; Roediger, Putnam, & Smith, 2011).
Another reason for this phenomenon is found in transfer-appropriate processing studies, which suggest that there is a correlation between how information is initially encoded and how it is later retrieved which, enhances memory performance. These studies have found that successful memory retrieval reinforces the encoding process, and enhances both the future retrieval and encoding function (Kang, McDermott, & Roediger, 2007; McDaniel et al., 2007; Roediger & Butler, 2013).

A final explanation for the benefits of the testing effect is encoding variability which suggests that studying and taking a test (which are two different encoding events) enhances learning by elaborating existing memory traces and or the creation of additional retrieval routes to the that trace which would result in a higher rate of recall in the future. Therefore, testing multiplies the number of retrieval routes to the stored information (Kang, McDermott, & Roediger, 2007; Roediger & Butler, 2013; Roediger & Karpicke, 2006).

Researchers point out that the effectiveness of testing depends on successful retrieval, feedback and repetition (Roediger & Butler, 2013; Roediger and Karpicke, 2005). Feedback is not always necessary in order for students to benefit from the testing effect as studies have shown. However, feedback can enhance learning from test by allowing students to correct their mistakes and increase the probability of successful retrieval in the future. Studies have shown that repeated testing leads students to space their study efforts and enhances their future retention (Kang, McDermott, & Roediger, 2007; Roediger & Butler, 2013; Roediger and Karpicke, 2006).

Another aspect of testing effect is the generation effect which consists of the notion that if a person comes up with information, such as in responding an essay question, it is better recalled than if that same information were simply presented such as in reading or studying or multiple choice questions (Tigner, 1999). Therefore there is a long-term benefit for generating our own answers or solutions then if the answers or solutions where presented to use like in a multiple-choice test. Bjork and Bjork stated that retrieval is thus, “a powerful memory modifier” (2011). Retrieval practice like free recall done in open-ended questions or essay questions improves the conceptual organization of that retrieved material. This kind of retrieval practice causes students to organize information more than does reading alone. Therefore open-ended tests can improve organization of knowledge and help later retention (Roediger, Putnam, & Smith, 2011). Hinze & Wiley (2011) tested the benefits of the generation effect and found that successful retrievals of fill in the blanks led to better performance on later multiple-choice questions of the same content. The researchers also noted that when they change the testing format from fill in the blank to open-ended questions there was a more clear impact on the ability to answer new questions on the final multiple choice test (Hinze & Wiley, 2011).

Teachers most often use multiple-choice tests in their classroom because it is easier to grade and it is perceived as more objective. However, the drawback to multiple-choice questions is that they require less retrieval effort than if the students’ were generating their answers (Kang, McDermott, & Roediger, 2007; Marsh, Fazio & Goswick, 2012). Another pitfall to multiple-choice questions is that only one answer provided is the correct one and therefore if there is no corrective feedback, the students could potentially remember the wrong answers. This phenomenon is called a negative testing effect (Marsh, Fazio & Goswick, 2012; Roediger & Marsh, 2005). Even though research has shown that the negative testing effect can be reduced by
corrective feedback, little is known about how effective feedback is on young elementary students because those studies have been done on fourth to six graders (Marsh, Fazio & Goswick, 2012). Marsh, Fazio & Goswick (2012) found that second graders who participated in multiple choice testing showed both positive and negative testing effects without feedback. These students chose the wrong answer or multiple-choice lures most often on hard questions. The researchers also found that corrective feedback eliminated the negative testing effect and concluded that multiple-choice test could be a viable learning event but that corrective feedback needed to be given in order to prevent a negative test effect (Marsh, Fazio & Goswick, 2012).

Method
Brief Overview of the Study
The research was done using a quasi-experimental method. The study was based on a third grade class in a public elementary school in Rockland County. Normal instructional procedures and classroom content were used for the study and the scores on the quizzes and final were part of the course grades. A convince sample of twenty participants were given three quiz formats over a period of three weeks. After each social studies lesson, the participants were given a multiple-choice, open-end / fill in the blank quiz or a study guide. No corrective feedback was given after the students finished the multiple-choice or open-ended/ fill in the blank quizzes due to time allotment. Unfortunately the social studies block was the last block of the day and there was not enough class time to give corrective feedback. During the three-week period, the participants received two multiple-choice quizzes, two open-ended/ fill in the blank quizzes and two study guides on their social studies unit on Landforms. The unit test was given five class days after the last lesson. The unit test was a pencil-and-paper test that had a total of 40 items, of which 20 were presented in multiple-choice format and 20 were presented in open-ended / fill in the blank format. Data was collected from the scores of two multiple-choice quizzes, two open-ended/ fill in the blank quizzes and the final unit test. Fourteen participants’ scores were analyzed using a one-way within subjects (or repeated measures) ANOVA in order to compare the effects of quiz format (multiple-choice, open-ended/ fill in the blank and non quiz) on the final grade percentage on the final unit test. Consequently, three Paired Samples T-Tests were conducted in order to find out which quiz formats are significantly different from each other.

Participants
Twenty 3rd grade students (M= 8.4 years, 6 girls and 14 boys) from a public elementary school located in Rockland County participated in this study. Out of the twenty students, six (1 girl and 5 boys) were excluded from this study because they were absent from at least one multiple-choice quiz session, one open-ended/ fill in the blank quiz session or were absent on the day of the final unit test. Three of these students received special education services and were pull out of their general education classroom during the social studies block. This situation did not allow them to finish the social studies lesson and were unable to do the quizzes.

Instrument
This study constituted of 2 quiz formats (multiple-choice, open-end /fill in the blank and no quiz), which were compared to 2 within subjects design (unit exam format: multiple-choice and open-end /fill in the blank). The material used to create the quizzes, study guides, and final unit test were from the class power point presentation on landforms. All
quizzes were paper-and pencil quizzes. There were two multiple-choice quizzes with a total of thirteen questions. Two open-ended / fill in the blank quizzes with a total of thirteen question and two study guides with 14 landform images and their facts. For the final unit test, the original thirteen multiple-choice questions were presented as seven multiple-choice and six open-ended questions. The thirteen open-ended quiz questions were presented in the final as six multiple-choice and seven open-ended questions. Finally, the thirteen facts in the study guides were presented as seven multiple-choice questions and seven open-ended questions in the final test. The unit test was a pencil-and-paper test that had a total of 40 items, of which 20 were presented in multiple-choice format and 20 were presented in open-ended / fill in the blank format. All quizzes, study guides and unit exams were in pencil-and paper format.

Procedure.

The study was conducted with third graders in order to observe if there was a difference in the effectiveness of later retention between different quiz formats (multiple choice, open-ended / fill in the blank quizzing and no quiz). The quiz material was from one social studies unit on landforms, which was divided into six sessions. Each session had approximately 6 - 7 items for a total of 39 items in the unit. The researcher administered each quiz or study guide to the students immediately after each instructional session was concluded. The sequence for the format of the quizzes was the following: multiple-choice quiz follow by an open-ended / fill in the blank and then a study guide. The quizzes and study guides were given in a period of three weeks, at intervals of approximately 2.5 days. The unit test was given five class days after the last study guide was distributed. The unit test was a pencil-and-paper test that had a total of 40 items, of which 20 were presented in multiple-choice format and 20 were presented in open-ended / fill in the blank format.

The experiment was implemented without altering the teacher’s typical lesson plans or classroom activities. The teacher used a Power Point presentation that had images, definitions and videos about the different landforms.

Data Collection

Data was collected from the scores of two multiple-choice quizzes, two open-ended/ fill in the blank quizzes and the final unit test. The researcher scored the quizzes and the unit test and the teacher checked the results. The researcher created a grading rubric for the open-ended questions, which included key phrases that each response needed in order to be correct.

Analysis of Date

Out of the twenty students, six students’ scores were not included in the data analysis. Six students were absent from the final unit test, all the multiple-choice quizzes or all the open-ended/fill in the blank quizzes. Fourteen participants’ scores were analyzed using a one-way within subjects (or repeated measures) ANOVA in order to compare the effects of quiz format (multiple-choice, open-ended/ fill in the blank and non quiz) on the final grade percentage on the final unit test. The results of the one-way ANOVA will be used to determine if a significant difference exists between quiz format and later retention.

Results

In general, students performed best on the final unit exam for questions that had occurred on multiple-choice quizzes (M= 88.64), the second best mean was found for the items that had appeared on the open-ended/ fill in the blank quizzes (M= 76.43) and the worst scores were found on items not previously tested (Study guide M= 71.21). Comparing
the means scores in the unit exam based on the content previously quizzed (multiple-choice, open-ended/ fill in the blank and no quiz) demonstrates the benefit of quizzing on later retention.

The distribution of the scores in the unit final test based on the content previously quizzed is displayed in Figure 1. The first graph shows the scores of the content in the final exam that was quizzed using multiple-choice format. The second graph displays the scores of the content in the final exam that was quizzed using open-ended/ fill in the blank format. The third graph shows the scores of the content in the final exam that was not quizzed but the students received a study guide. The last graph displays the overall score distribution of the final exam test.

Another way to compare each content score based on its quiz format is by visualizing the distribution of the scores through a Stem and Leaf Plot. A Stem and Leaf Plot of this data is provided in Figure 2. The graph displays how the multiple-choice content on the final exam had the highest retention, followed by open-ended/ fill in the blank content and finally the study guide content. The Stem and Leaf Plot however, shows some outliners in the study guide content scores. Using the Stem and Leaf Plot to compare the means scores in the unit exam based on the content previously demonstrates the benefit of quizzing on later retention.

A one-way within subjects (or repeated measures) ANOVA was conducted to compare the effect of quiz format on later retention measured by a unit final test in multiple choice quiz, open-ended quiz and no quiz but a study guide conditions. There was a significant effect of the quiz format, Wilks’ Lambda = .305, F (3, 11) = 8.354, p = .004.

After the results of the one-way within subjects ANOVA suggested a statistical significance, the researcher conducted post hoc tests in order to determine which conditions are significantly different from each other. Therefore, three Paired Samples T-Tests were conducted. Instead of using the value 0.05 to decide if there was statistical significant, the researcher used the value 0.017 as the cut off because three paired T-Tests were compared to find statistically significance.

Three paired samples t-tests were used to make post hoc comparisons between conditions. A first paired samples t-test indicated that there was a significant difference in the scores for content tested using multiple choice quizzes (M=88.64, SD=10.493) and open-ended quizzes (M=76.43, SD=13.682) conditions; t (13) =3.764, p = .002.

A second paired samples t-test indicated that there was a significant difference in the scores for content tested using multiple choice quizzes (M=88.64, SD=10.4) and no quiz but a study guide (M=71.21, SD=12.330) conditions; t (13) =4.845, p = .000.

A third paired samples t-test indicated that there was no significant difference in the scores for content tested using open-ended quizzes (M=76.43, SD=13.682) and no quiz but a study guide (M=71.21, SD=12.330) conditions; t (13) =1.768, p = .100.

Discussion

These results suggest that quiz format does have an effect on later retention. The results also suggest that when elementary students are tested using multiple-choice quizzes, the retention of this content is greater than if they are tested using open-ended/ fill in the blank quizzes or no quizzes at all, but are instead given a study guide. However, there is no real difference in later retention if elementary students are tested using either open-ended/ fill in the blank quizzes or no quiz at all. Therefore the null
hypothesis is rejected and the research hypothesis is accepted.

**Conclusion**

The results suggest that multiple-choice quizzing is an effective retention tool for the elementary classroom. Due to time restraints, no corrective feedback was given to the students after they finished their multiple-choice quizzes and therefore the negative test effect could not be eliminated. Despite not having corrective feedback, students were still able to benefit from multiple-choice quizzing. A surprising result was that there was no significant difference in later retention when students were quizzed using an open-ended/ fill in the blank format. Previous studies showed the production tests like short answer, open-ended and fill in the blanks improved later retention because of the generation effect (McDermott et al., in press; Hine &Wiley; Reedier &Butler, 2003; Tigner, 1999). Perhaps, the difference in the results lies on the fact that most production quiz questions, in this study, were fill in the blanks and only a few questions were open-ended.

The social studies content also relied heavily on visual identification of landforms on landscapes and maps. In addition, the classroom teacher requested that the quizzes have a scrambled word bank.

Further study should be done in order to test the effectiveness between the different formats of production tests on later retention. Possibly, there is a difference in the effectiveness of production test format (short answers, open-ended, fill in the blank) on later retention.

**References**


