2004

Computerized reading assessment using the star reading software

Brian Michael Bartlett

COMPUTERIZED READING ASSESSMENT USING THE STAR READING SOFTWARE

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Education:

Instructional Technology

---

by

Brian Michael Bartlett

June 2004

COMPUTERIZED READING ASSESSMENT USING THE STAR READING SOFTWARE

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Brian Michael Bartlett

June 2004

Approved by:

[redacted]

Eun-Ok Baek, First Reader

5/5/04

Date

[redacted]

Brian Newberry, Second Reader

ABSTRACT

This project focused on the use a computerized reading assessment program called Star Reading. Reading has been one of the most difficult areas across the curriculum to assess. Reading assessment differs widely from teacher to teacher, and has traditionally been very subjective.

One fourth grade class of thirty-two students was tested throughout the 2001-2002 school year using the Star Reading program, while following the district adopted Language Arts/Reading series. Scores from the first month and scores from the ninth month were collected, and at the end of the year the Star Reading NCE score was compared to the standardized test NCE score from the same year.

Results indicate that a very low correlation exists between NCE scores on the Star Reading and NCE scores on the standardized test. This is significant, in that the Star Reading computerized reading assessment provides an accurate, and standardized method of assessing reading.

ACKNOWLEDGMENTS

This Project could not have been completed without the support, encouragement, and sacrifice of my wife Stacie and my entire family.

I would also like to thank my principal, Melissa Wisner, for her support of my computer use and data collection, as well as the Yucaipa-Calimesa Joint Unified School District for their purchase of the Star Reading software.

Finally I would like to thank the technical support department of Renaissance Learning for their assistance with the Star Reading program, and data management.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

BACKGROUND

Introduction

Classroom teachers have historically struggled with assessment. Each subject offers their own challenges for teachers and students. While some subjects are easier to assess student understanding, others are more difficult. Assessment of math skills is far easier due to the objective nature of the subject. The answer to a math problem is either correct or incorrect. But when it comes to writing or reading, the subject itself becomes less objective and more subjective.

With the implementation of standardized testing in education, the assessment process has come into focus more than ever. These norm-referenced tests are administered only once a year, generally in the spring. At the elementary level, they cover the areas of math, spelling, reading comprehension, and reading vocabulary. The test is a multiple-choice style, which takes an average of five hours to complete.

In 1999 California began to use the Academic Performance Index (API) to rank schools according to their

performance on these standardized tests. Each school is given a number, which corresponds with their overall test scores. This policy has put added pressure on students, teachers, and administrators.

On one hand, there is an understandable need for standardized testing. It remains a means for policy makers to address the current state of education. On the other hand, it has been argued (www.fairtest.org/facts/ACHIEVE.html) that the practice of standardized testing is fundamentally unfair for students. It provides a very short glimpse into their academic ability. Much of the standardized test relies on a student's ability to take a timed, multiple choice style test.

Teachers often ask the question "How will my class do on the standardized test?" They want to know what they can do in the classroom to get a glimpse of student performance on this test. Will the assessment used in the classroom correlate with the standardized test?

The elementary school is located in Southern California and is one of the five elementary schools in the school district. At the time of the study, the school was a year-round, multi-track K-5 elementary school with

820 continuously enrolled students. It is situated in an upper middle class neighborhood, with a student population of 86% white, 10% Hispanic, and 4% other.

The school is the most technologically advanced elementary school in the district. It maintains a student computer lab of 32 Pentium class Windows computers, a teacher workstation, a projector, and a Windows NT file server. Each classroom has one teacher workstation, and high-speed internet access throughout the campus.

Statement of the Problem

The problem was to address an increasing concern in the area of reading assessment. As pressure increases for students to achieve well on standardized tests, there arises a need for an accurate pre-test to any standardized reading test. This would help teachers understand if their classroom instruction was achieving the desired results. Most teachers rely on their own experience and expertise to grade reading, with little standardization from class to class, grade to grade, and school to school. (www.uakron.edu/colleges/educ/Syllabi/5500720.pdf) This could result in the same student scoring differently in reading from class to class.

One year prior to this study, the school acquired a piece of computer software called Star Reading from the company Renaissance Learning. This piece of software was developed to test the reading skills of students from the first or second grade through the twelfth grade. This was brought to the school as a method to standardize reading assessment, and bridge the gap that had developed between classroom assessment and standardized testing.

Purpose of the Project

The purpose of this project was to develop a system of reading assessment through the use of existing computerized assessment software, and to examine if a correlation exists between student achievement on the STAR Reading test administered in the classroom, and the STAR 9 standardized test administered by the State of California. The project included the development of a reading assessment plan, the implementation of the plan, and the evaluation of the plan.

The development of the reading assessment plan began with a review of the assessment needs. Assessment needs included the requirements of the district, the teacher, and the parents. This was followed by a review of the school's computer systems, both hardware and software. There was a

thorough review and mapping of the school site computer network, and individual desktop computer. This was followed by a complete examination of all computer software installed on student workstations and the file server. This was done in an effort to meet the needs of the new software, which was to be used for the project. Finally, the assessment software itself reviewed and information was collected on its reliability, compatibility, and effectiveness.

Next was the implementation of the assessment plan. This included the training of the students on the general use of the computer, as well as the use of the STAR Reading software itself. Time management for this project was of a definite concern. A considerable amount of time for testing was needed each month, and more time was needed to maintain the computer network, which stored the data from each student's test. Within the STAR Reading program, there exist templates for a number of different reports, which can be generated for each student, of the class as a whole.

Finally, the plan was evaluated for its overall effectiveness. The STAR Reading software was able to provide the district with a grade equivalent score for each

student. It was also able to provide the teacher with diagnostic assessments for the report card. It was also able to provide sufficient feedback to parents, as evidenced through a parent survey, which indicated 96% positive response to the use of computerized reading assessment.

To meet the goals of an effective computerized reading assessment program required the collaboration of various school personnel. The computer lab technician was an integral part of the network maintenance during this project, as was the input of other teachers regarding the format of the STAR Reading test. The administration was instrumental in the development and implementation of the plan, as certain information was only made available to them.

One objective of the project was to improve the general computer skills of the students. By improving their skills in the area of computerized reading assessment, their skills in other programs improved and their confidence in their ability to use a computer effectively increased.

A second objective was to have teachers begin using the STAR Reading assessment in their classrooms. As

teachers began to see the ease and effectiveness of the program, others experimented with it in their classrooms.

The third objective of this project was to standardize the classroom reading assessment. With the STAR in place as a guide, teachers would be able to score student reading abilities with a reference to the STAR scale. Teachers would no longer have to rely on their own subjective scale of "Good, OK, alright, or low".

Research Question

Will the use of classroom computerized reading assessment prepare students for state standardized reading tests, and will the results from classroom computerized reading assessment correlate to results of a state standardized reading test?

Hypotheses

The following research hypotheses were made regarding the project:

1. The use of regular classroom reading assessment through the use of computerized assessment software will prepare students for the state standardized reading test.

2. Scores from computerized reading assessment will correlate highly with scores from the state standardized reading test.

Significance of the Project

Student reading remains one of the most difficult subjects to assess objectively, and if a means of accurately assessing student reading through the use of computerized assessment was proved effective, reading assessment across grade levels could be standardized. This would then lead the way to a well organized reading program that would meet the needs of each student.

Limitations

During the development of the project, a number of limitations were noted. First and most notably, is the fact that test data is only accurate under perfect testing circumstances. No matter which test is administered there are certain environmental factors that may distract test takers. The second limitation of this project is the reading instruction in the classroom. While there was consistency within the classroom, students who did not receive regular classroom reading instruction (i.e.

resource students) were not as familiar with the format of the computerized reading assessment. Finally, this project was limited by the accuracy and form of data analysis within the STAR Reading program itself. The program does offer data in the form of a Normal Curve Equivalency score, which is a nationally accepted means to compare results from various test types.

This study was based on the use of the STAR Reading program only, which does not directly assess student reading comprehension. A student could have mastered the test taking strategies of the STAR Reading program and scored very well. The same student could have difficulty comprehending ordinary grade level text. Renaissance Learning, the maker of STAR Reading, does offer several other computer-adaptive assessments. Accelerated Reader is the reading comprehension assessment companion to the STAR Reading program. In most cases, the two pieces of software would work together and provide and accurate, overall assessment of student reading (strategies and comprehension).

Strengths

The STAR Reading computer-adaptive reading assessment has offered teachers a consistent means of assessing student reading. It provides unbiased, accurate student data with multiple report formats, and is preferred by parents over traditional reading assessment. With the STAR Reading software, teachers can test whole classes, small groups, or individual students.

Another strength of the STAR Reading program in general, is that it has helped to improve student computer skills, test-taking strategies, and has helped to motivate students. Students have virtually instant feedback on their progress/performance on the test, and have come to fully understand how their scores were obtained.

The program has also affected teachers. The district in which the study was completed, is now mandating all teachers assess reading with the STAR Reading software for report cards. The STAR Reading scores are combined with classroom observations and individual reading and comprehension, then used to assign grades.

Future Projects

The STAR Reading program serves as a very useful tool in the process of assessing student reading. In future projects, it would be beneficial to study the correlation of Accelerated Reader and the state standardized test of reading comprehension. Is there any correlation between these two reading comprehension assessments?

Definition of Terms

The following terms are defined as they apply to the project.

STAR Reading (STAR)

The Renaissance Learning computer software program that assesses student-reading skills.

STAR 9

The California standardized test for public education used during the time of the research.

Normal Curve Equivalency (NCE)

The nationally standardized means of reporting and comparing scores from different tests.

Computerized Reading Assessment

A computer which is configured with software that enables a student to take a reading test.

CHAPTER TWO

REVIEW OF THE LITERATURE

Introduction

Computers have become a vital part of daily life for many people, and in education, computers are used across the curriculum for a number of reasons. Computer use by students begins at a very early age when they learn the basic computer skills needed to write, research, and present their work. Within the past several years, computers have begun to be used for assessment. Industry related tests like the Microsoft Certification Exam (MCE) and the test for registered nurses, have been formatted and offered on the computer. Computerized assessment also offers certain benefits over standard paper and pencil assessment. Research into the use of computers for assessment can be divided into three basic categories.

First is the overall computer use by students or children. A review of the literature indicates that the students' interest in computers plays an important roll in the use of computer technology in the classroom. The level of interest will make or break a computer-rich classroom and the instruction or integration that goes with it.

Second, is the use of computer software within the school setting. Computer software is just as important, and tends to spark student interest in computers in the classroom. Third, there have been reports of gender-bias effecting student computer use at schools. Girls and boys do tend to view computers differently, but would this prevent one group or the other from achieving their goal when it comes to computer use?

Finally, the use of the computer for assessment is researched. Is computerized assessment a valid form of alternative assessment, and does this form of assessment justify the cost associated with it? These are relatively new issues to the field of education, and key to the success and implementation of computers in the school.

Computer Use by Students and Children

It really is no secret that one of students' favorite things to do on the computer is play games. It doesn't matter what kind of game, but if it looks fun, and there is some sort of action, the students will generally like it. Educational software has made some great advances in the recent years, like many other areas of instructional technology. Schools have seen, used, and purchased too

many educational software titles to list, and as one looks back the question arises, "What exactly were they trying to do?"

Software developers have attempted to maintain a high level of interest, while at the same time developing skills, which is the goal of most educational software. But, any software title may have a quite different effect on any number of students. What may improve critical thinking for one, may improve verbal reasoning for another.

Although there is no plan to stop using, purchasing, or teaching educational software, there is an ongoing concern with regard to its validity and design for widespread use. (www.ncrel.org/tplan/cbtl/execsum.htm) Teachers have raised the question several times, "What are my students getting out of this software?" Knowing what teachers do now about the advantages of "good" vs. "bad" software, there is a new need for critical evaluation of any educational software used in the classroom.

In a recent study Henderson, Klemes, and Eshet (2000) evaluated the usefulness of a science-based software within a thematic unit. The software called Message in a Fossil (MIF) is a micro world simulation, in which students can explore an archaeological site to uncover any number of

fossils. After uncovering these ancient objects, students make inferences and create dioramas to reconstruct the habitats of the organisms. The software claims such cognitive outcomes as higher level thinking skills, classification, and basic recall.

The research was conducted in a second grade class in Texas, during the fourth quarter of their school year. Three pairs of students were selected for the study, based on their academic ability in the classroom. The teacher was responsible for selecting the pairs, which were also selected on the basis of student familiarity. The students would participate in the cross curricular unit on the average of forty-five minutes per day, with average computer time of twenty minutes per day.

Data was collected from a pre and post written questionnaire and a pre and post hands on questionnaire. At the end of the six-week unit, there was an overall improvement of 24 percent in the number of correct answers between the pre and posttest. The test questions ranged from low to high cognitive skill. The results of this study did indicate improvement in several thinking skills and strategies, but also indicated there were several areas in which the study could have been improved.

From the descriptions of the simulation software being used, one could gather that it was designed for something higher than the second grade level. And after reviewing the results of the questionnaires and reading samples of student dialogue, that was obviously intended for a higher grade.

Another factor is the fact that this unit was not strictly comprised of the software simulation; it was an integrated thematic unit. So was this a fair assessment of software's ability to improve cognitive outcomes? While it did manage to get the students to achieve better scores on the post questionnaire, the thematic and integrated nature of the unit had more to do with that, than the software alone. Integrated technology has become an every increasing part of today's educational setting. Teachers use programs such as the thematic units to incorporate computer technology into their classroom (Tan, 2000). These are commonly referred to as Computer Assisted Instruction or CAI.

Higher academic achievement is the goal of educators, and it is their job to find whatever materials work to achieve that end. In the case of this simulation, the students were motivated by the lure of the computer and its

software, as well as the integration of the unit into their daily school activities.

Computer Software

When it comes to educational software, there are almost as many motivators as there are students. Terrell and Rendulic (1996) found that by using computer-managed instruction, student motivation would increase, therefore increasing the achievement of fifth grade students.

This study compared the spelling scores of two fifth grade classes. The question was how to motivate students enough to see an academic improvement. Many students at earlier ages are motivated extrinsically, and often, but not always, make a shift toward intrinsic motivation. Computer-managed instructional (CMI) feedback was used with one class, and traditional reporting was used with the other. The control class used traditional reporting methods at three reporting times. The experimental class was given a computer printout type of report, which included visual representation of grades as well as textual messages.

The results of this study found that the class that was given the CMI feedback did in fact score higher on the

Children's Academic Intrinsic Motivation Inventory (CAIMI). While weekly graphic feedback alone did not appear to increase student motivation of achievement, but during the second nine-week grading period, textual information was added to the bottom of each graphic. This did result in a "significant increase" in achievement.

There were several factors in this study that led to results that were less than expected. First of all, the classes were greatly different in their academic ability before the study began. The control class was functioning at a higher level in spelling from the beginning of the year. It stated that the school administration adjusted the class to include these higher functioning students. Therefore significant adjustments had to be made to the spelling scores before the study began. Secondly, this study claims to motivate students with the use of this CMI software. The students had really no input into the use of technology, but student motivation itself was the focus. Most students do feel the stress evaluation or test anxiety. These feeling can have a dramatic impact on their achievement on any given assessment (Hancock, 2001).

What exactly do children learn from using computers? This is an excellent question and one that is asked

frequently. Student computer literacy can have a significant impact on their ability to learn or function in a computer reliant environment, such as a lab (Tuner, Sweany, & Husman, 2000). Mayer, Schustack, and Blanton (1999) examine the cognitive outcomes of students learning to use technology in an informal, collaborative setting.

This study compares three after school computer clubs, in which the participation is voluntary, and the children are self-paced. 30 to 50 off the shelf educational programs were made available, and students were compared based on their regular participation. The study took place during the course of one academic year, and incorporated 10 to 20 student visits. The setting of each of the three sites included similar socioeconomic status, grade levels, opportunities for expression, and collaborative nature.

Four kinds of cognitive changes were examined: computer literacy knowledge, comprehension skills, game-playing skills, and academic skills. In the first, it was found that past experience in a computer environment had a positive effect on learning basic computer literacy skills. It was also shown that students who had figured out how to use a wide variety of educational computer games improved language comprehension skills. It was surmised that

students who learn to use computer games, might better learn how to comprehend instructions, and therefore perform better on any assessment.

When teachers think about computer use in education they mainly think about educational games that help build skills. How could computer game playing skills help students? There are two major factors involved in playing new games. The first is the ability to comprehend instructions, and the second is the ability to devise a plan or strategy. There is evidence that students who are exposed to a wide variety of educational computing experiences develop skills that can transfer to new situations.

At one of the sites where this study was conducted, there was evidence that computer experience has a positive effect on academic skills. Students within this group outscored their peers on both math and language sections of the state assessment. These results are consistent with the idea that participation in educational computing results in learning that goes beyond simple retention and procedures.

Chappell (1997) conducted a two-part study, to investigate factors that may affect girls' attitudes toward

educational software. A Computer Game Attitude Scale (CGAS) was created and used with forty-eight seventh grade girls who played Geometric Golfer in the first part. This game was chosen because the identity of the caddy could be switched from male to female. The second part of the study involved fifty-two sixth grade girls from a wide variety of backgrounds and ethnicities. A game called Treasure Math Storm was used for this part because it allowed the separation of "aggressive distracters".

Three hypotheses were tested within the two parts. 1: Girls' attitudes toward computer software are negatively affected by the presence of competition in the program. 2: Girls' attitudes toward computer software are negatively affected by the under-representation of female characters in the program. 3: Girls' attitudes toward computer software are negatively affected by the presence of aggressive distracters in the program. Part one of the study, tested hypothesis one and two, while part two tested hypothesis three.

The results from this study did not support the first hypothesis. It does not indicate a negative attitude toward software due to the presence of competition in the program. The second hypothesis was not supported by the

results from part one as well. There was no indication that girls' attitude was affected negatively due to the presence of all male characters in the program. The second part of the study, which tested the affect of aggressive distracters, did not produce results that would indicate any negative attitudes from girls either.

Bradshaw and Clegg (1995) found there to be significant evident that supports the notion for boys to identify with male characters and girls to identify with female characters.

In their study, first grade boys and girls were shown a gender-neutral character (PODD or PLAYPARK) and interviewed as to their impression of the compute characters. The result showed that the "male as norm" gender assignment is a dominant strategy used by children. It also revealed that by simply removing the sex-stereotypical features the character often retains the initial gender meaning. Initial identification of the characters were reported as 80% of the boys indicated a male character and 70% of the girls indicated a male character, while 18% of the boys thought it was a genderless and 22% of the girls thought it was genderless.

In a two part series by Jolicoeur and Berger (1988) they create a plan for implementing effective educational software into the classroom. In their eight-step plan, they identify the goals of the process and list the steps that will be taken toward meeting those goals.

1. Specify the overall goals of the implementation procedures.
2. Select appropriate software.
3. Develop software support materials.
4. Randomly assign students to comparable groups.
5. Schedule and implement computer time for students.
6. Test student skills at regular intervals.
7. Evaluate the success of the software implementation procedures.
8. Evaluate the results of the issues examined.

An interesting point brought up in this article is that when software is used, is it effective in teaching what it was designed to teach? If not, why continue to use it? It is therefore critical for schools to evaluate the effectiveness of the software programs they use.

This article continues to explain that by controlling the different features of any particular software, other

issues can be examined. Five other issues were evaluated in this study: 1. overall software effectiveness, 2. effectiveness of individual software programs, 3. the effectiveness of tutorials versus games, 4. knowledge retention patterns based on the type of learning processes, and 5. gender effects.

Schools often leave out one or more of these steps for one reason or another. At many school sites, teachers view the computer lab as a one hour block of free time for the students and planning time for the teacher. There is very little thought that goes into how to best use their computer lab time, and even less assessment of what students have accomplished. This rather old, (1988) but still useful implementation plan can be successfully used at any site.

The second part of the study by Jolicoeur and Berger (1988) examined the effectiveness of eight software programs at teaching fifth grade student new fraction concepts and spelling words. Five questions were asked before this study began, as previously described.

The results indicated that there was significant student learning in both the fractions and the spelling software. There did appear to be a difference in the

effectiveness of the fraction tutorial versus the game, but the spelling tutorial was just as effective as the spelling game. The study found that while fraction knowledge did not decrease in the weeks following the use of software, the spelling knowledge did decrease.

A large part of what makes a school's computer technology program successful is the educational software made available to the students. Therefore it is critical for teachers, administrators and parents to know that the software is beneficial to the students.

There are now hundreds if not thousands of educational software designers making their products available to schools throughout the world. How do we know if they really have the best interests of the students in mind when designing their product? Truett (1984) conducted a series of field tests on a variety of educational software. The goal was to determine if designers are making the effort to put students first when creating their software. Despite the age of this study, the results would be similar now.

The study consisted of a survey sent to 406 software producers/publishers within the country. Of these, only 56 were returned. 60% of the returned surveys indicated that their software was available for preview prior to purchase.

Nearly 70% indicated that they field-tested their software. There were 15% of the publishers who actually had their products field tested by more than 10 teachers, and 17% were tested by more than 100 students.

According to this study there appears to be a rather small group of software producers who are truly concerned with the successful application of their products. So how, without preview privileges are teachers to know if software is what they want. First of all there are very few publishers who sell directly to schools. This completely removes them from being accountable for their product. Secondly, resellers are not generally willing to allow a preview of any software title. Finally, once the software is purchased and used, the reseller makes it nearly impossible for a school to return a title if they are not happy with it.

Assessment Tools

With a recent shift toward the constructivist view of education, in which each child is responsible for themselves and the information they get, the whole idea if technology based learning and assessment has undergone a transformation. This change toward constructivism has

definitely sparked a change in the view of educational software, from "What can I learn from the software?" to "What can I do with the software?"

Squires (1999) looks at this issue in his article on educational software and constructivist learning. Within the article, he expresses his feelings that learning should be "authentic" (Squires, 1999, p.49), from both a cognitive and contextual perspective. He feels that educational software will be used best when the designers create software for "subversive use" (Squires, 1999, p.51). When this happens, the users can re-interpret the intentions of the software to suit their needs. The article states that both teachers and learners can assume the subversive role, but more appropriately, it should be worked into the design of the software itself.

Creation of materials is discussed as well as the use of the World Wide Web, each offering aspects that promote constructivism into the design. When students are put in charge of their learning, they will tend to be more responsible. Through the creation of authentic assessment, students are more able to show what they know about any given subject, and not just what their teacher wants to see. Authentic assessment and the idea of constructivism

in education will change the way teachers and students view things.

When school districts consider new technology, a major component of their decision is cost. What will this cost the district to implement? In many cases, the cost associated with computerized assessment of any kind, is far less individual review of the assessment (Clauser, Harik, & Clyman, 2000). Teachers tend to be overburdened with the many different forms of assessment in the classroom, and computerized assessment has been shown to save time and money.

In a study of teachers' assessment practices, (Bol, Stephenson, & O'Connell, 1998) it was shown that teachers who use alternative means of assessment, such as computerized assessment, report feeling confident of their validity. These forms of assessment go beyond the regular classroom observation, to assess student knowledge in a way no other assessment could. The teachers' feelings of confidence went along with their experience in the field they taught. The more experience a teacher had, the better they felt about the alternative assessment they were using. These feeling of confidence most likely stemmed from their

many years in giving the "traditional" assessment, and they gave.

Summary

The use of technology in the education system began years ago with the television, VCR, even the scantron answer sheet. Now education has broadened the use of instructional technology with the use of the computer. The computer is no longer used as a "drill and kill" practice. It is no longer used solely for typing reports. The use of the computer has is evident from kindergartener through high school. Students now need to know basic computer skills in the elementary grades, in order to be successful in the upper grades. The computer software has helped to motivate, excite, and teach students in any subject area, yet flaws remain in the selection, and application of certain software.

Finally, the computer has emerged as an assessment tool all in its own. The use of highly adaptable databases and sophisticated measures within the software itself that adjust for different student levels, have emerged from software producers. These types of computer adaptive assessment tools are becoming very popular among teachers

pressed to teach so much in such a little amount of time. Computerized assessment has offered the ease of use and the accuracy of assessment needed by teachers, schools, and districts.

CHAPTER THREE

METHODOLOGY

Introduction

The idea of standardized testing has been around for many years, as a means of assessing the knowledge, or the institute of education itself. Although there have been numerous changes to the format of the standardized testing, the concept has remained unchanged. With the advent of computers in the educational, whether for instruction, review, or assessment, the notion of computerized assessment is rather new (Gifford, 2001). By combining computerized assessment with the need for accurate reading assessment, teachers are better able to meet the instructional needs of the students and at the same time prepare them for the state standardized testing. This quantitative research study relies on the validity and reliability of the state standardized test (a national norm referenced test), as well as the validity and reliability of the STAR Reading test. Non-probability, convenience sampling was used acquire data from both these tests, which was compared using the Pearson product-moment correlation coefficient.

Population Served

The test subjects for this project consisted of thirty-two fourth grade students from an upper-middle class community in Southern California. Student population, which closely resembled the surrounding community, was as follows: fifteen Caucasian boys, two Hispanic boys, one Native American boy, thirteen Caucasian girls, and one Hispanic girl. Students ranged in age from eight to ten years old. One student in the class had an active Individualized Education Plan (IEP), and was serviced daily through Special Education services.

Student make-up of the class was consistent with that of the school, which was a year-round multi-track, K-5 elementary school. There were twenty-six full time credentialed teachers, one principal, and one assistant principal. A small support staff of four classified personnel was in place to assist in the classrooms. The school also housed one Special Day Class, along with two Resource rooms for "pullout" services.

The school was ten years old at the time of the study, and had just been through a major technology upgrade. A computer lab of thirty Macintosh Apple IIgs computers had been replaced with Gateway PCs. A campus-wide LAN had been

installed, which allowed each classroom access to the computer lab, or any other computer or printer on the campus. A file server was installed in the computer lab, and a high quality laser printer was set up for teacher use.

Along with the new hardware, there was a large selection of computer software that was purchased at the same time. Some of these were education games, word processing and spreadsheets, graphics applications, and keyboarding tutorials. Teachers were polled by the administration, as to which computer software was desired, and which was needed, before the purchases were made.

Data Collection

At this point, it was fairly obvious that the assessment of reading, and reading processes in particular, was of great importance to teachers. The school district had only a sampling of specific reading assessments available to the classroom teachers. These included the Johns Basic Reading Inventory, the Degrees of Reading Proficiency (DRP), and the Names Test, a test of basic phonemic awareness (Baker, 2001). Teachers were free to choose any one or more to meet their assessment needs.

But these district forms of assessment were often inadequate, and still rather subjective. What was needed was a "standardized" means of assessing student reading achievement across the grade levels.

Renaissance Learning, founded in 1986, created a product called STAR Reading. STAR Reading is a computer-adaptive diagnostic reading test (Renaissance Learning, 2002). It combines the need for standardized reading assessment with the current computer technology now in place in most public schools. STAR Reading is one component of Renaissance Learning's large variety of computer-adaptive test.

STAR Reading had been purchased for the schools where the study was to be completed, but it had not been installed or even reviewed buy the teachers. Upon learning of it, the new principal was eager to get it in use.

The STAR Reading software was installed on the school's Windows NT server and various teacher computers around the campus. These "workstations" could access the STAR Reading program across the existing Local Area Network (LAN). A secure data location was set up on the server,

where the STAR Reading program could record and track student test results, as well as modify their next test. Figure 1 shows the STAR Reading new student screen, in which characteristics for students can be added. These characteristics can be used when printing reports within the STAR Reading program.

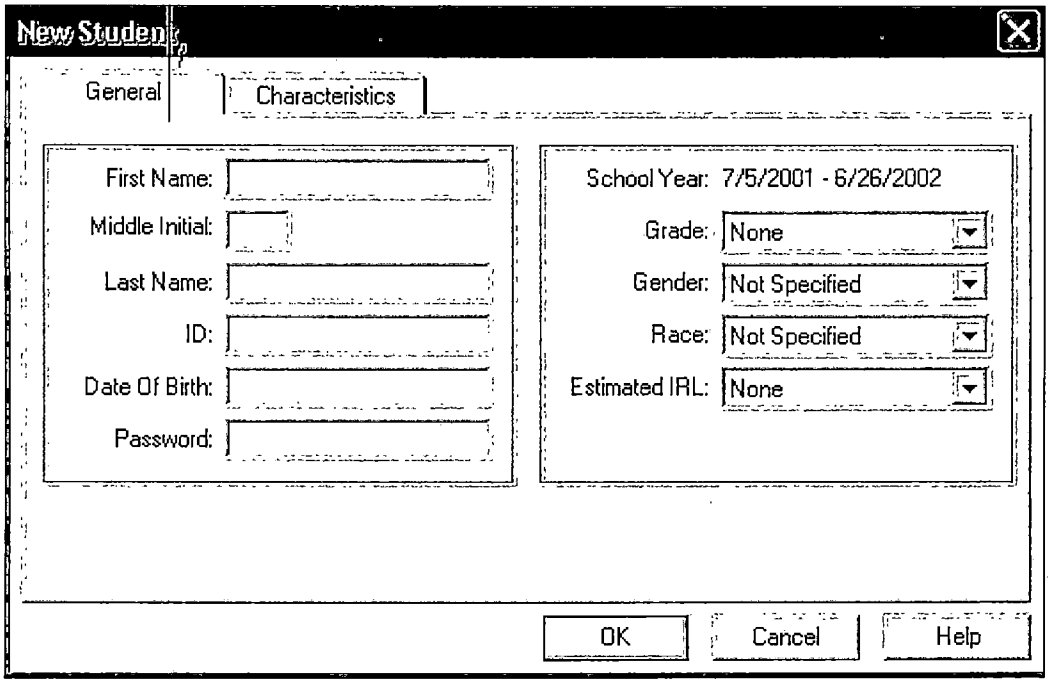

Figure 1. STAR Reading New Student Screen

To begin using the program, the experimental group had to be set up in the program, as a class, with an assigned teacher. The student names, birth date, sex, ethnicity, student number, and grade were entered into the database

used by the program. The first time the students took a STAR Reading test, the program would store their results in the database, for future reference.

Once the students are added to the program, they can need to be enrolled in a particular class. The class enrollment marks their grade and placement during the year, which can also be used for STAR reports. Figure 2 shows a snapshot of the STAR Reading class setup. Here five students have been enrolled in the class called "Test group".

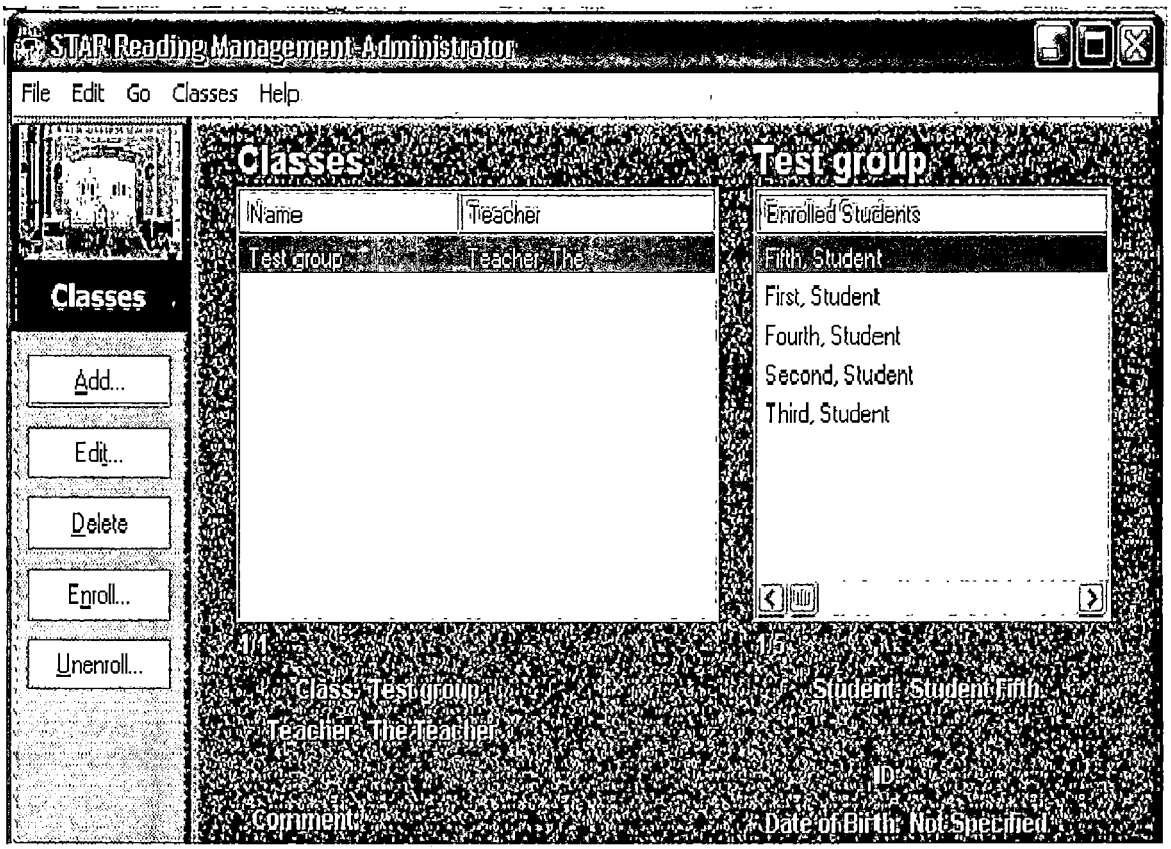

Figure 2. Snapshot of the STAR Reading Class Setup

The STAR Reading software incorporates computer-adaptive technology, which allows for an individualized test each time a student logs into the program. The software draws the information from the student's previous test results, which are stored in the database. It then creates a new, unique test for the student. Each time a student takes the STAR Reading test, the program adjusts to the student's last test results. If a student performed

well on all twenty-five test questions, the next test would include more difficult questions.

The student test itself consists of three practice questions followed by twenty-five test questions. The questions are in a multiple-choice format, with a choice of four answers. Each question has a time limit, and will move on to the next question if not answered within the allowed time. Questions in the student test begin rather short and simple, and develop into lengthy, complex passages.

Before students could begin using the STAR Reading program they had to become familiar with the basic computer operation. The class was instructed for several weeks on functions such as turning on and off the computer, opening and closing programs on the desktop, using drop-down menus, and minimizing or maximizing a window.

Specific skills were targeted for fifty-minute lessons twice a week. Students, who became proficient early on, were allowed to explore more complex Windows commands and other educational software available on their computer.

As the project developed and the school year progressed, several problems were noted with the STAR Reading software. These ranged from rather simple to fix,

individual computer lock-ups, to data corruption within the program itself. One remedy for preventing these potential problems included restarting all the computers in the computer lab prior to testing the whole class. Because the program was dependant on date stamped student data, if one computer in the computer lab had an incorrect date, all computers who began the STAR after that computer would indicate an error. As the need arose, students were taught to check and correct any incorrect computer dates.

Computer lab management and even student computer management around the campus was crucial. All computers, which students were using to take the STAR Reading test, were updated weekly. The STAR data was backed up weekly on the school's server and on CD. Before the year was over, the students in the experimental group were well versed in the smallest details of the program, and because of the time spent on the computers, these students were considered "advanced" computer users around the campus.

Students were tested at the beginning of the year and at the end of each trimester using the STAR Reading program. This data was stored, and their test record was used to mark their progress in reading. STAR Reading allows for reports to be printed for reference. These

Data Analysis

At the end of the school year, a "growth" report was generated. This report showed the student growth from the first test of the year to the last test of the year. As part of this report, the scores were also given in Normal Curve Equivalent (NCE) points. The NCE is a score report used when comparing two or more different tests. This scoring method is used nationally to compare students on state standardized tests. Patterns of growth were observed, with close attention being paid to the "Total Reading" NCE score.

Table 1 is a sample of the data collected from the reading assessments. It shows the Total Reading NCE scores for the 2001 and 2002 state standardized test along with any growth made from 2001 to 2002. It also shows the August and June NCE score for the STAR Reading computerized assessment along with any growth made from August to June.

Table 1. Student Score Sample

2001 STAR 9	2002 STAR 9	Total Reading	Star Reading		
Total Reading NCE	Total Reading NCE	STAR 9 NCE growth	Aug-01 NCE	June-02 NCE	Star Reading NCE growth
54.3	50.5	-3.8	51.6	48.4	-3.2
84.6	84.6	0	72.8	66.3	-6.5
75.8	82.7	6.9	62.9	73.7	10.8

NCE scores were aggregated, and the STAR Reading NCE growth score was compared to the state standardized test (STAR 9) NCE growth scores. The correlation coefficient of was calculated using the Pearson product-moment correlation coefficient.

Summary

A thorough examination of the test results, revealed interesting details about the STAR Reading and the state standardized test. Through a carefully designed quantitative research study using reliable methods, student test data was collected over a one year period, and compared. NCE growth over the one year period was recorded, aggregated, and the correlation coefficient was calculated. The details that this study revealed, provide

an insight into results from the two different reading assessments, and the relationship between them.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

The research data gathered from this project was, and will continue to be used to develop methods to better meet the needs of students. Through a review and discussion of the findings, this project can be used to modify classroom teaching and the process of assessing student reading. This data was collected through district and state mandated assessment; although it is not directly responsible for student reading grades, it does provide a valuable opportunity to improve teaching methods and a possible foresight into future reading test scores.

Presentation of the Findings

Students' Total Reading NCE from the California standardized test were compared to the Total Reading NCE from the 2002 test, and the growth, or difference between the two scores recorded. This growth ranged from -24.7 NCE points to 8.5 NCE points. Twenty students did not gain in Total Reading NCE points from the 2001 test to the 2002 test, while three students scored exactly the same and nine

gained NCE points slightly. As the data shows, there was an overall decline in the Total Reading NCE points from the 2001 to the 2002 test.

Students' STAR Reading NCE scores from the first month of the 2002 year were compared to the STAR Reading NCE scores from the last month of the school year. Here, growth ranged from -18.7 NCE points to 18.6 points. Again, twenty students did not gain in NCE points from the beginning of the year to the end. There were no students whose score was the same, but twelve students gained NCE points on the STAR Reading.

Finally, the two sets of NCE growth scores, (the state standardized test growth from year to year, and the STAR Reading test growth from the first month to last month) were compared. Nineteen students showed growth from the state test to the STAR Reading test, while thirteen did not show growth.

The correlation between the aggregated student scores on the state standardized test for reading, and the aggregated scores for students on the STAR Reading test, was examined. The correlation coefficient was calculated as 0.10 for these two sets of NCE growth. This correlation was positive but weak, which does not show any significant

correlation between the two sets of scores. The 0.10 correlation coefficient is positive, and does show a very small correlation between the state test Total Reading NCE and the STAR Reading NCE.

Two hypotheses were made regarding the use of the STAR Reading software. First was that the implementation of regular classroom reading assessment through the use of the computerized software would prepare students for the state standardized test. This hypothesis was proved inaccurate by the results of the study. While there was a slight increase in students who scored higher on the STAR Reading, there was a decrease in student scores on the state standardized test.

The second hypothesis stated that scores from the computerized reading assessment will correlate highly with scores from the state standardized test. This hypothesis was also proved inaccurate by the results of the study which indicated a positive, yet very weak correlation between the two forms of assessment.

Discussion of the Findings

Results from this project did not match the hypothesized results made prior to data collection and analysis. Renaissance Learning makes large claims for the success of their STAR Reading computerized assessment software, and STAR Reading has come into widespread use throughout the nation. While claims are made that refer to the use of their reading comprehension program (Accelerated Reader), there are no claims that the STAR Reading program will improve actual reading ability, nor will it improve student standardized test scores.

The goal of this project was not to raise standardized test scores as a result of using the STAR Reading program. The goal was to determine if a relationship exists between the data collected from the STAR Reading and data collected through the state standardized reading test. If a correlation exists between the two test scores, then, through careful and targeted classroom reading instruction, accurate student assessment could be made through the use of the STAR Reading program.

Results from the project did not indicate that the STAR Reading assessment was an inaccurate means to assess student reading. The fact remains that reading assessment

is difficult to objectively assess, and through the use of this program, it can become an avenue for standardized reading assessment across the grade levels.

Results from this study could have important ramifications for schools that rely heavily of the STAR Reading test to assess their students' reading abilities. The STAR Reading test does offer schools a means of "standardizing" reading assessment within the school, but it was found to give no definitive proof of student performance on state standardized reading tests, and does not correlate with the state standardized test. Will schools continue to rely on the STAR Reading test?

Summary

The findings presented in this study reveal a very low correlation between the two reading assessment tools used. A low Pearson product-moment correlation coefficient indicated that NCE growth in the state standardized reading test would not necessarily point to growth in the STAR Reading assessment. The inverse can also be stated. Although the test scores are not connected, the use of computerized reading assessment can still be used to improve the reading instruction in the classroom. The

computer adaptive STAR Reading assessment remains an easy to administer reading test as well as a test that provides results that are uncomplicated.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

Introduction

The STAR Reading assessment and the state standardized reading test are both well designed forms of assessment each with its own strengths and weaknesses. Results from each test provide similar information, yet at the same time different. These test scores allege student achievement on that given day. As any educator knows, achievement cannot be measured on one given day. The computer adaptive STAR Reading assessment is a tool similar to the state standardized reading assessment, yet the STAR can be given numerous times throughout the year, therefore would have an advantage over any state standardized assessment.

Conclusion

The data gathered as a result of this study was not conclusive. The correlation coefficient of 0.10 did not indicate a strong correlation between the state test Total Reading scores and the STAR Reading test score. As with all standardized style tests, there are certain criticisms. A lower student score from any given day could be dismissed

with the notion of the student having a "bad day". Since it is a "snapshot" into the students' total knowledge, any number of factors could affect their score on that day.

Reading comprehension and classroom observations of student reading continue to be a needed component of a successful overall assessment plan. Students in this study were tested using computerized reading assessment and conventional classroom reading assessment. While the data indicated little correlation, individual progress was observed over the span of the school year.

In conclusion, the STAR Reading computer-adaptive reading assessment program showed inconclusive evidence of its success in predicting student performance on state standardized tests. It did prove useful in its effectiveness as a reading assessment alternative.

Recommendations

When assessing student reading, it is important to understand the individuality of each student. At the same time, it is important to be completely objective in the assessment. Computerized reading assessment is an excellent alternative to "traditional" means of assessing student reading.

For teachers interested in implementing the STAR Reading program at their school, it is an invaluable addition to any site. Site administrators and teachers can access information on the STAR Reading and Accelerated Reader software from the Renaissance Learning website. Free trial versions of all the software can be ordered, and evaluated at each school site.

Before attempting to begin any computerized reading assessment, or any other computerized assessment for that matter, it is important to have a thorough understanding of your computer network at the school site. Because this software runs over a network, it is reliant on a strong, consistent, well-maintained computer network. Several people at the site should be trained on the overall structure of the computer network, and have a general understanding of how information travels across a network. These few people should also know how to access the network server across the network, in order to do routine back-ups of the software data.

Another recommendation would be to have teachers whom would be involved in the computerized assessment, go through basic computer training on the software. All

teachers would need to know the basic commands and operations of the specific software they would be using.

Finally, for schools that choose to implement such a computerized reading assessment program, the students would need basic computer knowledge as well. Within the STAR Reading program, there is a certain amount of computer knowledge needed by the test taker. These basic skills include clicking with the mouse, and using shortcuts or the Start menu.

Summary

The conclusions and recommendations drawn from this study were based on the data reviewed, as well as personal experience with the reading assessment tools discussed. Further study would provide educators with results and information needed to continue to improve the process of reading education.

APPENDIX A:

INSTITUTIONAL REVIEW BOARD APPROVAL

CALIFORNIA STATE UNIVERSITY
SAN BERNARDINO

5500 University Parkway, San Bernardino, CA 92407-2397

March 05, 2004

CSUSB INSTITUTIONAL REVIEW BOARD
Full Board Review
IRB# 03068
Status
APPROVED

Mr. Brian Michael Bartlett
c/o: Prof. Eun-Ok Baek
Department of Science, Math, & Technology
California State University
5500 University Parkway
San Bernardino, California 92407

Dear Mr. Bartlett:

Your application to use human subjects, titled, "Computerized Reading Assessment" has been reviewed and approved by the Institutional Review Board (IRB). Your informed consent document is attached. This consent document has been stamped and signed by the IRB chairperson. All subsequent copies used must be this officially approved version. A change in your informed consent requires resubmission of your protocol as amended.

You are required to notify the IRB if any substantive changes are made in your research prospectus/protocol, if any unanticipated adverse events are experienced by subjects during your research, and when your project has ended. If your project lasts longer than one year, you (the investigator/researcher) are required to notify the IRB by email or correspondence of *Notice of Project Ending* or *Request for Continuation* at the end of each year. Failure to notify the IRB of the above may result in disciplinary action. You are required to keep copies of the informed consent forms and data for at least three years.

If you have any questions regarding the IRB decision, please contact Michael Gillespie, IRB Secretary. Mr. Gillespie can be reached by phone at (909) 880-5027, by fax at (909) 880-7028, or by email at mgillesp@csusb.edu. Please include your application identification number (above) in all correspondence.

Best of luck with your research.

Sincerely,

Joseph Lovett, Chair
Institutional Review Board

JL/mg

cc: Prof. Eun-Ok Baek, Department of Science, Math, & Technology

APPENDIX B:

PARENT INFORMED CONSENT FORM

STUDY OF COMPUTERIZED READING ASSESSMENT
PARENT INFORMED CONSENT

The study in which your child is being asked to participate is designed to investigate the benefit of the STAR Reading computerized reading assessment software. This study is being conducted by Mr. Brian Bartlett under the supervision of Eun-Ok Baek, PhD, professor of the College of Education, Department of Science, Math, and Technology. This study has been approved by the Institutional Review Board, California State University, San Bernardino.

In this study your child took the STAR Reading test throughout the year as a means of district reading assessment policies. The test takes about 10 to 15 minutes and was given at the beginning of the year as well as the end of each trimester marking period. Scores for these tests were recorded. Copies of all tests and information about the software are available from Mr. Jeff Litel, Principal of Ridgeview Elementary School for parent review. In addition to STAR Reading data, the "Total Reading NCE" scores from the 2001 state standardized test, and the 2002 state standardized test were recorded. Once these pieces of data were collected, student names were removed and the list randomly arranged to protect your child's privacy.

Your child's participation in this study is totally voluntary, however, your child's STAR Reading scores will still be used for grading purposes.

There are no foreseeable risks involved in this study. Students were not be asked to do anything outside of the normal Reading curriculum, teaching practices, or district/state mandated assessment. The intended benefit is improved reading skills.

If you have any questions or concerns about this study or would like to receive the results of the study, please feel free to contact Brian Bartlett at 909-797-8382 or Professor Eun-Ok Baek, PhD at (909) 880-5454.

By placing a check mark in the box below, I acknowledge that I have been informed of, and that I understand, the nature and purpose of this study, and I freely give consent to my minor child to participate.

Place a check mark here ☐ Today's Date: _____

Student Name: _____ Signature: _____

Parent/Guardian Name: _____ Signature: _____

CALIFORNIA STATE UNIVERSITY, SAN BERNARDINO
INSTITUTIONAL REVIEW BOARD COMMITTEE
APPROVED 02/20/04 VOID AFTER 02/20/05
IRB# 03068 CHAIR _____

APPENDIX C:

STUDENT INFORMED CONSENT FORM

STUDY OF COMPUTERIZED READING ASSESSMENT
STUDENT INFORMED CONSENT

The study in which you are being asked to participate involves the use of the computer and the STAR Reading software. It will help to tell the teachers of your reading strengths and areas where you may need additional help. The study will also give the teachers an indication of how you might do on the standardized test in reading. This study is part of a class at California State University, San Bernardino.

In this study you will take the STAR Reading test several times throughout the year. The test should take about 10 to 15 minutes and will be given at the beginning of the year as well as the end of each trimester marking period. Scores for these tests will be recorded. In addition to STAR Reading data, the scores from the 2001 standardized test, and the 2002 standardized test will be recorded. Once these scores are collected, your names will be removed and the list will randomly arranged to protect your privacy.

Your participation in this study is totally voluntary, however, your STAR Reading scores will still be used for grading purposes. When the study is complete, it will be described more for you.

There are no risks involved in this study. You will not be asked to do anything outside of the normal Reading curriculum, teaching practices, or district/state mandated assessment. The intended benefit is improved reading skills.

If you have any questions or concerns about this study or would like to receive the results of the study, please feel free to contact Brian Bartlett at 909-797-8382.

By placing a check mark in the box below, I indicate that I have been told about, and that I understand, this study, and I am willing to participate.

Place a check mark here ☐ Today's Date: _____

Student Name: _____ Signature: _____

CALIFORNIA STATE UNIVERSITY, SAN BERNARDINO
INSTITUTIONAL REVIEW BOARD COMMITTEE
APPROVED 02/20/04 VOID AFTER 02/20/05
IRB# 030608 CHAIR Joseph Lovett

REFERENCES

Baker, E. A. (2001). The Nature of Literacy in a Technology-rich Fourth-grade Classroom. Reading Research and Instruction, 40(3), 159-184.

Bol, L., Stephenson, P. L., O'Connell, A. A., & Nunnery, J. A. (1998). Influence of Experience, Grade Level, and Subject Area on Teachers' Assessment Prectices. Journal of Educational Research, 91(6), 223-330.

Bradshaw, J., & Clegg, S., (1995). An Investigation Into Gender Bias In Educational Software Used in English Primary Schools. Gender & Education, 7(2), 167-175.

Buly, M. R., & Valencia, S. W. (2002). Below the Bar: Profiles of Students who Fail State Reading Assessments. Education Evaluation and Policy Analysis, 24(3), 219-239.

Chappell, K. K., (1997). Investigating the Impact of Elements in Educational Mathematics Software On Girls' Attitudes. Journal of Educational Computing Research, 17(2), 119-133.

Clauser, B., Harik, P., & Clyman, S. G., (2000). The Generalizability of Scores for a Performance Assessment Scores with a Computer-Automated Scoring System. Journal of Educational Measurment, Fall, 2000, 37(3), 245-261.

Gifford, A. P. (2001). Is There Only One Way to Evaluate Students? Reading Improvement, 17(2), 35-37.

Hancock, D. R. (2001). Effects of Test Anxiety and Evaluative Threat on Students' Achievement and Motivation. Journal of Educational Research, 94(5), 284-290.

Henderson, L., Klemes, J., & Eshet, Y., (2000). Just Playing a Game? Educational Software Simulation Software and Cognitive Outcomes. Journal of Educational Computing Research, 22(1), 105-129.

Jolicoeur, K., & Berger, D.E., (1988). Implementing Educational Software and Evaluating Its Academic Effectiveness: Part I. Educational Technology, September, 13-19.

Jolicoeur, K., & Berger, D.E., (1988). Implementing Educational Software and Evaluating Its Academic Effectiveness: Part II. Educational Technology, October, 13-19.

Mayer, R. E., Schustack, M. W., & Blanton, W. E., (1999). What Do Children Learn From Using Computers in an Informal Collaborative Setting? Educational Technology, March-April, 27-31.

Molebash, P. & Fisher, D. (2000). Teaching and Learning Literacy with Technology. Reading Improvement, 15(1), 63-129-75.

Renaissance Learning (2002). http://www.renlearn.com

Squires, D. (1999). Educational Software for Constructivist Learning Environments: Subversive Use and Volatile Design. Educational Technology, May-June, 48-53.

Tan, S. C. (2000). The Effects of Incorporating Concept Mapping Into Computer Assisted Instruction. Journal of Educational Computing Research, 23(2), 113-131.

Terrell, S., & Rendulic, P., (1996). Using Computer-Managed Instructional Software to Increase Motivation and Achievement in Elementary School. Journal of Research on Computing in Education, 28(3), 403-415.

Truett, C. (1984). Field Testing Educational Software. Educational Technology, May, 7-12.

Turner, G. M., Sweany, N. W., & Husman, J. (2000). Development of the Computer Interface Literacy Measure. Journal of Educational Computing Research, 22(1), 37-54.

Vollands, S. R., Topping, K. J., & Evans, R. M. (1999). Computerized Self-Assessment of Reading Comprehension with the Accelerated Reader: Action Research. Reading and Writing Quarterly, 15, 197-211.

Woodrow, J. E. J., Mayer-Smith, J. A., & Pedretti, E. G. (2000). Assessing Technology Enhanced Instruction: A Case Study in Secondary Science. Journal of Educational Computing Research, 23(1), 15-42.

www.agsnet.com/glos/assessment.asp#N