Florida Journal of Educational Research
Fall 1999, Vol. 39, No 1, 79-94

Preparing Students for High-Stakes Testing in Florida:

An Interactive Website

Valerie Wright

St. Leo University

Ann E. Barron

Jeffrey D. Kromrey

University of South Florida

The Florida Comprehensive Assessment Test (8th Grade Reading): A Staff Development Tool is a large, instructional website that is designed to help teachers prepare students for a high-stakes test. It was designed and developed by graduate students and staff of the Florida Center for Instructional Technology at the University of South Florida. This article provides information on the analysis, design, and development of the interactive website. Results of an evaluation study with 8th grade students (n=1430) and teachers (n=18) indicate that the students found the program to be easy to use and useful in preparing for the FCAT.

Introduction

There is a growing trend towards implementing high stakes tests -- standardized tests that are used to decide whether a student is promoted or graduates. While Florida has always conducted annual tests with regard to achievement (e.g., SAT-9, CTBS), the prospect of being retained or not graduating was never adopted as a consequence. In 1996, Florida's Department of Education (DOE) adopted the Sunshine State Standards, a set of benchmarks for student

knowledge and performance at grades PreK-2, 3-5, 6-8, and 9-12 (see http://www.firn.edu). In order to measure the progress of students' achievement relative to these benchmarks, the Florida DOE also created the Florida Comprehensive Assessment Test (FCAT) as the primary vehicle for educational accountability in Florida. Initially, students took the FCAT reading tests in grades 4, 8 and 10, and the FCAT mathematics test in grades 5, 8 and 10. In 1999, the Florida Legislature expanded the mandate to include grades 3-10 for both reading and mathematics. In 2003, tests are being added for science (Florida Department of Education, 1999).

The FCAT is a standard-referenced assessment. The results provide a snapshot of where students stand, not in relation to each other, but in relation to their ability to reach educational standards for public school students. Since 1999, FCAT results have been included in the School Accountability Report, which is used to identify Critically Low-Performing Schools, as well as High-Performing Schools. In addition, FCAT results are used to identify students in need of remediation, to obtain feedback on curriculum and teaching strategies, and to gauge student progress (for promotion and graduation purposes).

Interactive Website for 8th Grade Reading

Recognizing the need to prepare teachers and students for high-stakes testing, the Florida Center for Instructional Technology at the University of South Florida, along with Pinellas School District, secured a grant to create an online, instructional program that focused on FCAT strategies. This article focuses on the design, development, and evaluation of an instructional program for FCAT: 8th Grade Reading.

The FCAT reading test for 8th grade contains multiple choice and student performance tasks. Performance tasks require the students to write either a short or extended response to a question using information from a reading passage. The 8th grade reading test consists of the following:

- 6-7 reading passages
- 40-45 multiple-choice questions
- 6-8 short response performance tasks
- 2-3 extended response performance tasks

All multiple-choice questions have 4 answer choices. The short and expended response questions are assessed using scoring rubrics. Short responses are scored using a 2-point scoring rubric -- a complete and correct answer is worth 2 points, and a partial answer is worth 1 point. The extended responses are scored using a 4-point scoring rubric -- a complete and correct answer is worth 4 points, and a partial answer is worth 1, 2, or 3 points (Florida Department of Education, 1997).

Analysis

The initial target audience for the program was teachers of 8th grade language arts classes. However, a series of interviews with teachers soon revealed that the program should be designed so that 8th grade students could also benefit from the instruction, practice, and interactions. Although Florida teachers and students were familiar with standardized tests that contained primarily multiple-choice questions, the presence of performance items on the FCAT introduced a new domain. With this format, teachers were forced to learn the parameters of the

performance tasks, become proficient in the use of the rubrics, and prepare their students to write effective answers in the short and extended response format.

The needs assessment also indicated that the major emphasis of the program should be on strategies and techniques for writing and scoring the short and extended response questions. Teachers and language arts supervisors who were serving as subject matter experts felt strongly that if students and teachers had the opportunity to view and score short and extended responses (using the DOE rubric), it would help them in test preparation. In addition, the teachers requested practice tests (using FCAT structure) that would provide immediate feedback for the students.

Design

Extensive research was conducted to determine an appropriate "look and feel" for the program. This assessment included interviews with 8th grade students, teachers, and principals, as well as an analysis of reading material popular with 8th grade students. Based on the feedback, bright colors were selected for the palette, and a "mascot" of a 3D character was purchased for the program (see Figure 1).

The interface was designed in a clear, intuitive manner. The browser toolbars were hidden, and fixed links (Site Map, Exit, Previous Menu, Back, and Next) appear in the upper-right corner of all appropriate screens. Throughout the program, section headings appear in the upper-left corner of the screen, and prompts appear below the navigation options. Although the program is created using HTML frames, scrolling is constrained whenever possible (with the exception of the reading passages).

The content for the passages and questions in the program were obtained from the Florida Department of Education and the Pinellas County Supervisor of

Language Arts. All articles used in the program were tested to ensure the appropriate reading level, and all test items were piloted in local schools.

The program is divided into five main areas, and each area focuses on specific content/interactions and employs an appropriate instructional strategy (see Figure 1). The five areas include:

1) Introduction
2) Scoring Activities
3) Self-test Using Scoring Rubric
4) FCAT Practice Tests
5) Resources

Figure 1. Main Menu

Introduction

The *Introduction* is designed in a linear, informational format. It provides background information about the FCAT – its origin, purpose, construction, and

evaluation. The section also presents an overview of the Sunshine State Standards (with a link to the corresponding website). Information about the program's navigation, delivery, and procedure for printing are also included.

Scoring Activities

In the *Scoring Activities*, users can view actual student answers to short response questions, along with the corresponding rubric scores. They also have an opportunity to practice using the rubric to score student answers. The screen in Figure 2 presents a scoring exercise. In this example, the students read the story (Wayward Bears), read the student response (in the rectangular area of the screen), and score the response as a 2, 1, or 0 (based on the scoring rubric). Immediate feedback is provided (below the rubric score), and users are encouraged to try again if their score is incorrect.

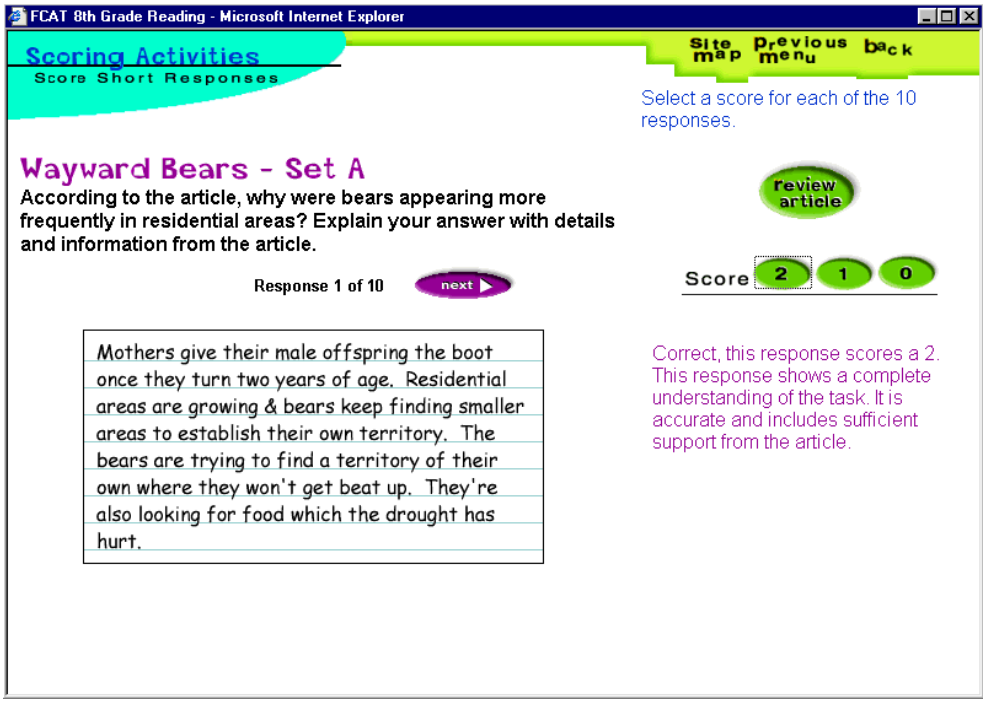

Figure 2. Rubric Scoring Activity.

Self-test Using Scoring Rubric

The *Self-test Using the Scoring Rubric* section is designed to assess the users' ability to score short and extended responses using the FCAT scoring rubric. The strategy employed is similar to that of the previous section (Scoring Activities), with the addition of a running score and a final assessment score.

FCAT Practice Tests

The *FCAT Practice Tests* section seeks to duplicate the structure of the FCAT, with the added benefit of immediate feedback. This section contains both multiple choice and short response questions. The multiple-choice questions are followed by immediate feedback (see Figure 3). Note that the reading passage is visible in a left frame – during the administration of the actual FCAT, students are encouraged to refer to the article as often as they like. The short response items in this section provide a field for student to type their input, along with an opportunity to view an "expert" answer for comparison.

Resources

The *Resources* section provides extensive information and resources related to FCAT. The information has been divided into two main categories – related websites and teaching strategies. Both sections are presented using an explore strategy.

Development

The program was developed with HTML and JavaScript (using Macromedia's DreamWeaver). Multiple frames were used to create each screen, and scrollbars appear when necessary. Based on the user's computer hardware, the program will automatically adjust to a fixed screen size -- either 640 X 480 or 800

X 600. No plug-ins are required, and the screens and graphics are optimized to download quickly for delivery over 28.8 modems.

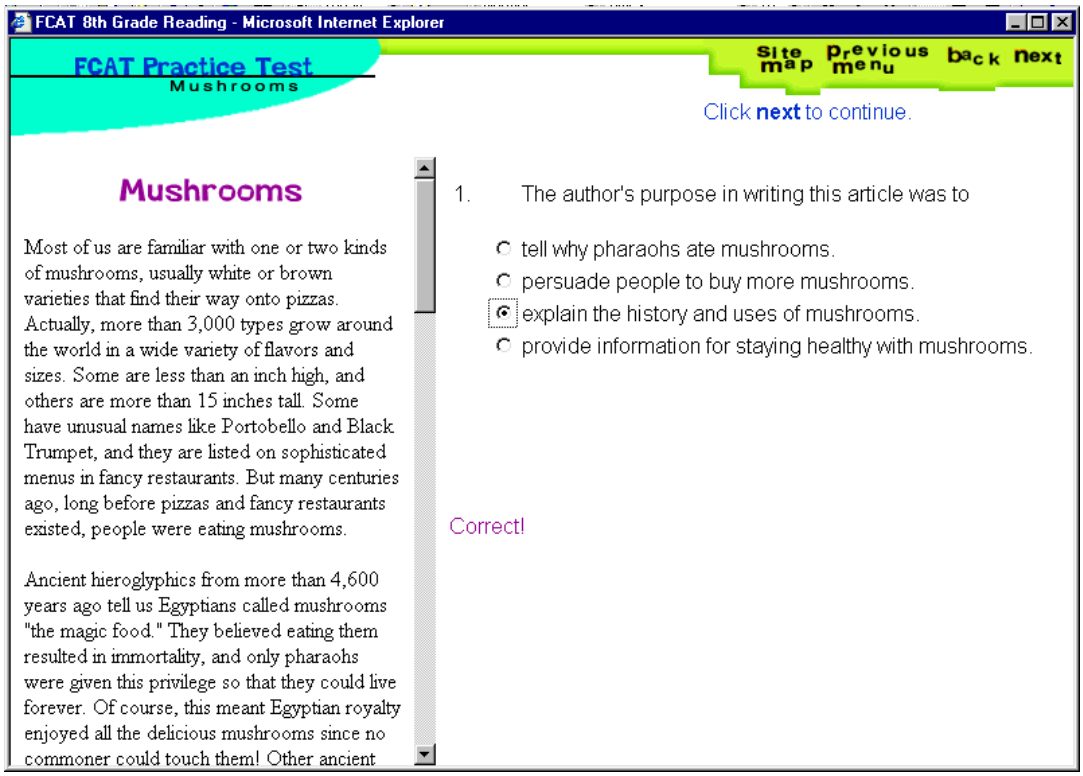

Figure 3. Multiple Choice question in Practice Test.

The major challenge in the development of the program was ensuring that it would operate effectively on both Mac and PC computers, with either Netscape or Internet Explorer browsers (4.0 and above). Although style sheets were employed to help control the size of the fonts, different versions were required for different software/hardware configurations. Therefore, commands were written (in JavaScript) to ascertain the user's hardware and browser, and to display the appropriate version of the program. For example, Netscape 4.0 offers a Print Button, but Internet Explorer 4.0 does not. Extensive testing was required to

ensure that the program would work correctly in a variety of delivery configurations.

To reach the largest possible audience, the program was developed for delivery via either CD-ROM or the web. To make sure that it would run either locally or remotely, client-based interactions and tracking were employed. Scores appear on the screen for the student to view, but records are not tracked to the server.

Evaluation

After beta-testing the program, a formal software evaluation study was conducted. The objective of the study was to obtain information about both teacher and student experiences with the program when it was implemented as a classroom/computer lab activity. The evaluation included measurement of teacher and student perceptions of the software and its usefulness for FCAT preparation.

Sample

Seven public middle schools in a west central Florida school district participated in the study. With the assistance of the school district's Supervisor of Reading/Language Arts, a letter was sent to the principal at each of the 23 middle schools in the district asking for volunteers for the research study. Incentives for participation included on-site training, workbooks for each student, and class sets of novels for their reading classes. Seven principals volunteered their schools to be part of the study (a total of 18 teachers and 1,430 students).

Instrumentation

Two instruments were used for data collection: a teacher's log and a student software evaluation checklist. Each teacher (n = 18) maintained a log during the

course of the study. The log included a record of activities completed by each class, the time spent on each activity, and any difficulties encountered with the software. The log also provided pertinent directions for the teacher, such as a delineation of the required activities, step-by-step directions for each activity, and a site map of the program for navigation.

The student software evaluation checklist was developed by the researchers in order to ascertain the views of the students about the program. The checklist was divided into four main sections, 1) ease of use, 2) program appearance, 3) usefulness, and 4) overall rating. In addition, space was provided for the students to describe what they would change about the program if they could. The software evaluation form was pilot tested prior to the study using 30 students in 8th grade from another school district.

Procedure

Teachers at each school received on-site training from a technology specialist who was involved in the development of the program. Each training session lasted approximately one hour in length. The teachers were given step-by-step directions on program navigation and the required student activities. The teachers were asked to have their students complete five activities in the program during a span of one month prior to the state administered FCAT. The activities consisted of articles for the students to read, rubric scoring practice, and FCAT-like practice tests.

For each rubric scoring activity, the students read the article in their classroom and answered the short response question in the workbook. Subsequently, in the computer lab, the students located the article and question in the program, viewed examples of students' responses corresponding to each rubric

score point, practiced scoring student sample responses, and recorded the number of correct scores in their workbooks. Finally, returning to their classrooms, the students self-scored their answers to the short response question and (optionally) revised their responses.

For the activities involving the practice tests, the students read the article and answered the short response question in their classrooms. Then, they moved to the computer lab where they located the practice test in the program, responded to the multiple choice items, recorded their scores in their workbooks, and compared their short response to the "expert" answer in the program. When all activities were complete, the students filled out the software evaluation checklists.

Results

After five weeks, the teacher logs and student software evaluation checklists were collected. In addition, the teachers were invited to add their comments about the program and recommendations for future use. The teacher logs were analyzed for the following, 1) amount of time students devoted to the software, 2) specific student activities completed, and 3) problems or concerns encountered.

Teachers' Logs

During the course of this study, the amount of time students spent working with the software ranged from 1.5 to 9.0 hours, with an average of 3.5 hours. The teachers were required to complete five activities with their students, and five supplemental activities were included that the students could complete if the teachers so desired. Fifteen of the 18 teachers completed the required activities, and 14 of the 18 completed some or all of the additional activities with their students.

The majority of problems/concerns expressed in the teacher logs dealt with difficulties experienced within the schools' computer labs (e.g., the computers would freeze-up during an activity, or the system was very slow). Four teachers, or twenty-two percent (mostly ESE teachers), expressed concern about the difficulty of the reading material. Additionally, four teachers (22%) expressed confusion about classroom organization while using the software (i.e., having students work at their own pace vs. teacher directed lessons). The majority of teachers (94%, n = 17), reported that they liked the program and believed that it benefited their students.

Student Software Evaluation Checklist

The software evaluation checklist was distributed to every student who participated in the study (n=1430). For student ratings of Ease of Use, the response options were *very easy*, *easy*, *difficult* or *very difficult*. In this section, 94% of the students found the directions provided in the software to be *very easy* or *easy*.

The response options for rating the two sections on Program Appearance and Usefulness were *excellent*, *good*, *fair*, or *poor*. For Program Appearance, *excellent* or *good* ratings were provided by 78% of the students in rating the navigational buttons, by 69% in rating the program graphics, and by 74% in rating the colors used in the program. For Usefulness, 78% of the students reported that the program helped them understand the FCAT scoring procedures, 67% reported that it increased their confidence with the FCAT, and 74% reported that the program prepared them for the FCAT. In addition, 78% found the practice tests useful. Overall, 70% of the students reported that they enjoyed the program (see Table 1).

At the end of the checklist, the students were asked, "If you could change one thing about this software, what would it be?" Thirteen percent of the students

(n = 186) reported that they would change nothing about the program, and 31% (n = 440) offered no comment about changes. The most frequently requested changes (23%, n = 335) were related to the content of the program (e.g., more articles that were easier to read, more questions for written responses, more explanation of the scoring rubrics, and the provision of a "Hint" button). Notably, few students (8%, n = 118) recommended changes in the fonts or graphics used in the program. Finally, four percent of the students (n = 61) commented that the school computers were slow and tended to freeze during the activities (a concern that was also expressed by some of the teachers). Many of the students (13%, n=186) responded that they would not change anything about the program (e.g., "I wouldn't change a thing, this software is excellent"). Some students even responded that it had helped prepare them for the FCAT (e.g., "Your product is very useful and it helped me understand how to write a good essay" and "After all this hard work, I felt like I was ready to take the FCAT").

Teacher Recommendations

At the end of the study, 12 teachers (66%) provided open-ended comments about the program in the back of their logs. Many of the comments made were very positive (e.g., "I think students benefited from the program" and "I think the lessons did help their writing"). Three teachers (17%), however, expressed some concerns with accountability, such as students hurrying through the program without attending to the readings, and guessing at the answers.

The teachers also made several recommendations. One common request (22%, n=4) was to provide a printout of students' scores upon completion of each activity. The teachers felt that these data would provide a simple mechanism to monitor progress of the students. Another frequent request (22%, n=4) was to

Table 1.

Results of Student Software Evaluation Checklist

	N	Very Difficult or Difficult	Very Easy or Easy
How easy was it to:			
1. Follow the directions?	1426	6%	94%
2. Use the software without help from your teacher?	1420	12%	88%
3. Move around the program?	1417	12%	88%
4. Find the required information on the screen?	1415	9%	91%
5. Read the text on the screen?	1414	6%	94%
	N	**Poor or Fair**	**Good or Excellent**
How would you rate the appearance of the following:			
6. Buttons used to move to different sections?	1427	21%	79%
7. Graphics (pictures) used throughout the software	1425	31%	69%
8. Colors used in the software?	1418	26%	74%
How well did the software:			
9. Maintain your attention?	1418	41%	59%
10. Make learning fun?	1411	53%	47%
11. Interest you in studying by computer?	1413	35%	65%
12. Help you understand the FCAT scoring procedures?	1410	22%	78%
13. Increase your confidence about the FCAT?	1410	33%	67%
14. Help in your preparation for the FCAT?	1413	26%	74%
How useful was/were the:			
15. Feedback provided by the software?	1424	30%	70%
16. Self-scoring activities?	1417	26%	74%
17. Examples of student writing?	1411	21%	79%
18. Practice tests?	1413	23%	77%
	N	**No**	**Yes**
19. If the program were on the web would you use it?	1412	57%	43%
20. Overall, did you enjoy the program?	1404	30%	70%

provide electronic scoring of the short and extended responses that the students typed into text boxes in the Practice Tests section. Finally, some of the teachers suggested that it would be helpful in the future to have a direction sheet for the students to use as they navigate through the program at their own pace.

Conclusion

The analysis, design, development, and evaluation of the *FCAT 8th Grade Reading: A Staff Development Tool* required a team of instructional designers, programmers, graphic artists, and researchers. The final product consists of over 1050 HTML files and 490 graphics, and it provides several hours of courseware. In January 2001, the program was been distributed (on CD-ROM) to all middle and high schools in Florida (n=2463). The distribution included both public and private schools. In addition, the program is available on the web at http://fcit.coedu.usf.edu/fcat8r. Analysis of the server logs reveal that the site receives approximately 1200 hits each month (with a noticeable surge prior to the administration of the FCAT).

In 2001, a similar program for FCAT 10th Grade Reading was produced. The implementation and evaluation of the 8th grade program provided valuable insight into design changes to enhance this and future products. For example, the menu options are labeled to more clearly indicate which parts of the program are designed for teachers and which parts are appropriate for students. In addition, running scores are being provided for all activities in the program, and print options for scores are included.

Although there is controversy about the appropriateness of high-stakes tests such as the FCAT (Dunne, 2000; AERA Position Statement, 2000), the tests remain an operative component of contemporary education and teachers are being

held accountable for their students' performance. The information presented in this article demonstrates that interactive, instructional websites can offer efficient and effective means to help teachers prepare their students.

Note: This research was supported, in part, by the University of South Florida and the Technology Literacy Challenge Fund for 1999-2000, under Grant No. # 520-12101-OCT22. The opinions expressed are those of the authors and do not reflect the views of the Florida Department of Education or the University of South Florida.

References

AERA position statement concerning high-stakes testing in preK-12 education. (2000, July). Washington DC: American Educational Research Association. Retrieved June 8, 2001, from the World Wide Web: http://www.aera.net/about/policy/stakes.htm.

Dunne, D. W. (2000, June 30). Some teachers, students, parents say no to tests! *Education World.* Retrieved June 8, 2001, from the World Wide Web: http://www.education-world.com/a_issues/issues094.shtml.

Florida Department of Education. (1997). FCAT: Reading test item and performance task specifications. (Available from the Assessment and Evaluation Services, 325 West Gaines Street, Room 414, Tallahassee, FL 32399).

High stakes testing spreads. Washington DC: National Center for Policy Analysis. Retrieved June 8, 2001, from the World Wide Web: http://www.ncpa.org/pi/edu/pd081099g.html.