

FCRR Technical Report #6

Brief Report of a Study to investigate the relationship between several brief measures of reading fluency and performance on the Florida Comprehensive Assessment Test-Reading in 4th, 6th, 8th, and 10th grades.

Joseph Torgesen
Steve Nettles
Pat Howard
Randy Winterbottom
Florida Center for Reading Research

With the apparent success of progress monitoring measures for reading growth in grades K-3, there is broad interest in the State of Florida in extending this assessment technology into the upper grades (4-12) for students who continue to perform below grade level on the reading portion of the Florida Comprehensive Assessment Test (FCAT). The desire is to have measures in place that will provide a standardized, state-wide metric for evaluating the progress of individual students who have an academic improvement plan (students who achieved Level 1 or Level 2 performance on the FCAT in the previous year) toward the goal of achieving grade level (Level 3) performance on the FCAT. The Just Read, Florida! Office in the State Department of Education asked the Florida Center for Reading Research to develop research information about the feasibility and accuracy of various alternatives for progress monitoring instruments in late elementary, middle, and high school.

The biggest challenge to this initiative is that the nature of the reading, language, and cognitive factors that account most importantly for individual differences in performance on the FCAT changes dramatically from grade three to grade ten (Schatschneider, et al., 2004). Reading fluency is the dominant factor in explaining individual differences in performance on the FCAT in grade three, while differences among students in verbal knowledge/reasoning is clearly the most important factor in the tenth grade. This finding reflects the fact that, during the beginning stages of learning to read, students accelerate enormously in the rate at which they can read words in text accurately and fluently. At the end of third grade, students who attain Level 1 performance on the FCAT are still very weak in their ability to read words in text accurately and fluently (54 words per minute on FCAT passages, vs. 102 words per minute for students at Level 3 and 148 wpm for students at Level 5). Young students who read at only 54 words per minute are still experiencing so many problems simply identifying the words in text that they are much less able than those who read fluently to focus on the meaning of the FCAT passages. By 10th grade, students at Level 1 read FCAT passages at an average rate of 130 words per minute, while students at Level 3 attained an average rate of 175 words per minute. These remaining fluency differences among students who perform at different levels on the FCAT still account for significant variance in performance on the FCAT (32%), but they account for less variance than in third grade (55%) for at least two reasons.

First, although students who read at 130 words per minute get through the text less quickly than those who read at 175 words per minute, a 130 word per minute rate

implies much more automaticity and ease of identifying words than the 54 words per minute rate of third graders. Thus, the effort and attention involved in identifying individual words for 10th grade Level 1 students is much less than for Level 1 students in third grade, and so the 10th graders with the slower reading rate are more able to focus on the meaning of what they are reading. One might say that, by tenth grade, even the Level 1 students have reached a “threshold” of reading fluency where they are less distracted by the effort involved in identifying individual words in text than are the Level 1 readers at third grade. Second, fluency differences among students account for less of the total variance in performance on the FCAT in 10th than in 3rd grade because the 10th grade FCAT places much heavier demands on broad knowledge and thinking skills than does the 3rd grade FCAT.

Between 3rd and 10th grades, the demands of the FCAT for “higher order” thinking skills accelerates dramatically. While only 30 percent of questions on the FCAT in third grade require complex thinking ability, 70% of the questions on the 10th grade FCAT are written to require these kinds of skills. Because of this change in the type of questions appearing on the FCAT, and because of the increasingly large differences among students in the kinds of thinking and reasoning skills the FCAT measures, differences among students in their knowledge and reasoning ability become the dominant factor in explaining individual differences in performance on the FCAT by grade 10.

This change in the nature of the reading, language, and cognitive abilities required for proficient performance on the FCAT is consistent with the generally changing goals of reading instruction from early reading to later reading development. While early reading instruction is focused to a great extent on helping students acquire access to text by becoming accurate and fluent readers, the goals of later reading instruction are to help students acquire the increasingly complex knowledge structures and reasoning skills required to comprehend complex text. As Perfetti (1985) has pointed out, when students move from 3rd grade to the higher grades, reading can be increasingly defined as “thinking guided by print.” Although it is important for students to continue to grow in their ability to read increasingly complex text fluently and accurately, it is even more important that they expand their knowledge base, strategic reading skills, and general reasoning abilities to accommodate the increasingly complex text they encounter at each succeeding grade level.

What we would like to have for use in Florida are measures that are sensitive to the kinds of reading growth in low performing 4th through 12th grade students that predict improved performance on the FCAT. These students will be receiving various types of special interventions for reading, and their teachers need to know whether the interventions they are providing are sufficiently powerful to improve the student’s performance on skills related to performance on the FCAT. One possibility would simply be to construct FCAT-like tests for students to take at various intervals during the year. Performance on this type of “progress monitoring” assessment should be highly predictive of improved performance on the real FCAT test when students take it in the spring. The major problem with this strategy is that the reading skills of many Level 1

students are so low that, although they may actually improve their reading skills (reading accuracy and fluency, or low level comprehension skills) from an assessment in September to one in December, their skills are still at such a low overall level that they will not enable improved performance on a grade level FCAT assessment. Another problem with this strategy is that the assessment simply restates that the student has a problem with performance on the FCAT, but it provides no information to teachers about the components of reading comprehension that are particularly in need of improvement for specific students. A final problem with this strategy is that, in order to have sufficient reliability, a test like this must have a significant number of questions, and the assessment time involved is substantial.(30 minutes or more). Of course, the time to take the test is not a large issue, as the test could be administered in groups. The expense required to develop multiple forms of the test, however, would be considerable.

Ideally, what we would like to have are separate measures of the vocabulary, strategic/reasoning processes, and reading fluency outcomes that are essential components of performance on the FCAT. It would also be desirable to have measures in these areas that are sensitive to student growth within a broad range of ability. Such measures are, in fact, available for many of the component skills required for proficient performance on the FCAT. The major problem with most of them, though, is that they involve individual assessments that take substantial time and require relatively extensive training before they can be administered and scored accurately. Assessing the complex array of cognitive and language skills that are critical to improved performance on the FCAT at higher grade levels is much more difficult than assessing the relatively discrete word-level skills that are part of progress-monitoring systems in grades K-3.

Since it is not feasible to perform a complete diagnostic/progress-monitoring assessment three or four times a year for struggling readers in grades 4-12, our goal shifts to identifying a metric that will be useful, but not fully comprehensive, for monitoring the reading growth of students with an Academic Improvement Plan in grades 4-12. Although we know that reading fluency, by itself, becomes increasingly less important in explaining individual differences in performance on the FCAT at higher grade levels, it is nevertheless true that many Level 1 students continue to have serious difficulties in this area (although the average for Level 1 students in our earlier study was 130 WPM, many students had rates far below that). For these students, one of the important goals of their individualized reading programs will be to improve their access to text by increasing their reading accuracy and fluency, in addition to improving their ability to think about the text they are reading. Thus, one possibility for assessing growth resulting from interventions to struggling readers would be to examine changes in their fluency and accuracy in reading FCAT like passages. Although reading fluency accounts for an increasingly smaller proportion of the variance in FCAT performance as students move through middle and into high school, any reasonably effective set of interventions should have an impact on reading fluency and accuracy, particularly for students who perform in the lower ranges on these measures. There is also evidence that reading fluency, itself, is influenced by the operation of “automatic comprehension processes” that also facilitate performance on a test like the FCAT (Jenkins, et al., 2003). Thus increases in reading fluency from strong interventions with middle and high schools students may reflect both

increases in efficient word identification and the further development of automatic comprehension processes that are developed from extensive practice reading text for meaning.

One of the limitations of assessing oral reading fluency is that, although the measures can be given very quickly, they must be administered individually. The possibility of developing a group administered progress monitoring assessment makes the work of Dr. Chris Espin at the University of Minnesota and the National Center for Student Progress Monitoring (<http://www.studentprogress.org/default.asp>) particularly attractive. Dr. Espin has been conducting research on progress monitoring measures in middle and high schools students for a number of years (Espin, Busch, & Shin, J.2001; Espin, Scierka, Skare, & Halverson,1999). One of the most recent findings from her research is that maze passages, in which students select which of three words best fits a blank space in the text, may be a more sensitive measure of reading growth in upper grade students than simple measures of oral reading fluency. The maze foils are not created to place high level demands on comprehension, but they do require that the student be monitoring the general meaning of the passages on a sentence or paragraph level. The score on this test is the number of mazes students can complete in 3 or 4 minutes. The alternate form reliability of these measures is sufficient for our needs (above .80), and the technique has face validity as a measure of both fluency and comprehension combined. Additionally, the measure can be given to groups of students.

In this study, we examined the relationship between performance on maze tests constructed from passages similar to those used on the FCAT test, and student's actual scores on the FCAT. We also administered three other brief assessments of reading skill that might be candidates for monitoring progress in reading for students receiving remedial instruction in grades 4-12. Our goal was to obtain initial evidence about the relationship between performance on these brief measures of reading skills and performance on the reading portion of the FCAT. Although it is also important that progress monitoring measures be sensitive to small increments in reading growth, the first criteria they must meet is a strong relationship with performance on the FCAT, since that test is used to determine whether students are meeting grade level standards in reading.

Method

Subjects were recruited for the study from Leon County School District in Tallahassee Florida, and from Dade County School District in Miami. 88 4th graders, 252 6th Graders, 161 8th graders, and 98 10th graders were tested.

Demographic Distribution across Grade Levels

Grade	Gender		Ethnicity					FCAT SSS Level				
	Male	Female	Caucasian	African American	Hispanic	Asian American	Multi-racial	1	2	3	4	5
4 th	35	52	31	16	33	3	4	9	11	28	33	5
6 th	112	140	99	94	49	4	6	31	56	72	67	25
8 th	74	110	57	64	55	3	5	45	61	46	28	3
10 th	47	58	13	31	57	3	1	22	30	27	13	13

The test administered were:

Espin Maze passages. The passages that Dr. Espin used in her own research were included as an anchor against which to compare performance on maze passages based on FCAT passages that are specifically developed for this study. The maze foils are not created to place high level demands on comprehension, but they do require that the student be monitoring the general meaning of the passages on a sentence or paragraph level. The score on this test was the number of mazes students completed in 3 minutes. The alternate form reliability of these measures is sufficient for our needs (above .80), and the technique has face validity as a measure of both fluency and comprehension combined. The passages were constructed from newspaper articles and they were administered only to the 8th and 10th grade students. Students completed three passages, and their score was the median score for the three passages.

FCAT-based maze passages. We used real FCAT passages to construct mazes for students in the 4th, 6th, 8th, and 10th grades. The passages were long enough so that students were able to complete them during the 3 minute reading time allowed for each passage. The students score was the median number of mazes completed correctly in three minutes, from reading three different passages. On both types of maze passages, scores were corrected for guessing by subtracting incorrect responses from correct responses.

Test of Silent Contextual Reading Fluency (TOSCRF). This is a newly developed test from Don Hammill at PRO-ED, inc. that allows an assessment of reading fluency in group administered format. It measures fluency by requiring students to place slashes between real words that are printed as strings of letters with no spaces between them. For example, the student was presented with a string for words such as: *thearticledidnotmentionthattheunithadaprimarmissionofofficersaftety*, and was required to identify the word segments by placing slashes like this: *the/article/did/not/mention/that/the/unit/had/a/primary/mission/of/officer/*. In order to correctly identify all word boundaries quickly, the student would have to have an ongoing sense of the gist of the meaning of the sequence of words, thus this test can be conceptualized as measuring both fluency and comprehension. The child's score on the

test was the number of correctly identified words in 90 seconds. The child was administered two forms of the test and the final score as the mean between the two forms.

Test of Sentences Reading Efficiency (TOWSRE). This test requires students to read sentences of increasing difficulty and indicate whether they make sense or not. It measures both silent reading fluency and a simple form of comprehension. To correct for guessing, incorrect responses are subtracted from correct responses. The child was administered two forms of the test, and the score was the mean between the two forms. The test is currently under development and standardization by Drs. Wagner and Torgesen, and will be published by PRO-ED, inc.

Oral reading fluency with FCAT passages. This test was included as the current “gold standard” for assessing reading fluency. We used FCAT passages, as they directly sample a student’s ability to read the kinds of words and sentences they are likely to encounter on the FCAT at their grade level. The student read three passages for one minute each and the score was the median correct words per minute across the three passages.

Results.

Results will be presented separately for each grade level. We will first present descriptive statistics for each test, and then will present correlations with the FCAT

4th Grade (N = 88)

Descriptive Statistics

Test	Minimum	Maximum	Mean	S.D.
FCAT SSS	207	430	319	44.2
Oral Reading Fluency	61	226	123	33.6
FCAT Maze	2	35	24	4.9
TOSRE	3	52	34	7.6
TOSCRF	3	134	88	23.3

Correlations Among Measures

	1	2	3	4
1. FCAT SSS				
2. Oral Reading Fluency	.56			
3. FCAT Maze	.54	.64		
4. TOSRE	.52	.57	.56	
5. TOSCRF	.48	.54	.64	.47

At fourth grade, there are no important differences in the strength of relationships between ORF and the mazes test and reading outcomes on the FCAT. The relationship between ORF and FCAT in this study is much lower than we obtained in two earlier studies with third graders (Buck & Torgesen, 2003; Schatschneider, et al., 2004). The correlations between FCAT SSS and ORF in these studies were .70 and .76, respectively. The earlier study had a large and more representative sample, and it also had a much smaller proportion of students in it that may have been English Language Learners. Thus, the relationships among all the progress monitoring measures and the FCAT may have been depressed in this 4th grade sample.

6th Grade (N = 228)

Descriptive Statistics

Test	Minimum	Maximum	Mean	S.D.
FCAT SSS	100	500	319	55.8
Oral Reading Fluency	55	231	154	32.1
FCAT Maze	3	56	25	9.2
TOSRE	17	59	34	8.6
TOSCRF	7	200	124	28.6

Correlations Among Measures

	1	2	3	4
1. FCAT SSS				
2. Oral Reading Fluency	.59			
3. FCAT Maze	.67	.71		
4. TOSRE	.58	.76	.72	
5. TOSCRF	.39	.53	.59	.50

At 6th grade, it looks as though the FCAT Mazes may have an advantage over ORF and the other measures, in terms of prediction of scores on the FCAT. It also looks as though the group administered TOSRE does as well as the ORF in predicting FCAT performance.

8th Grade (N = 161)

Descriptive Statistics

Test	Minimum	Maximum	Mean	S.D.
FCAT SSS	112	403	305	48.5
Oral Reading Fluency	41	241	144	35.3
FCAT Maze	0	58	29	11.1
Espin Maze	0	53	24	8.7
TOSRE	11	49	28	6.7
TOSCRF	4	191	129	26.2

Correlations Among Measures

	1	2	3	4	5
1. FCAT SSS					
2. Oral Reading Fluency	.62				
3. FCAT Maze	.63	.74			
4. Espin Maze	.59	.73	.79		
5. TOSRE	.58	.63	.59	.64	
6. TOSCRF	.22	.41	.38	.38	.29

At 8th graded, ORF and the Mazes test are very similarly related to performance on the FCAT, and they are not reliably better than the TOSRE

10th Grade (N = 98)

Descriptive Statistics

Test	Minimum	Maximum	Mean	S.D.
FCAT SSS	222	442	339	30.2
Oral Reading Fluency	94	222	154	29.3
FCAT Maze	5	49	26	8.8
Espin Maze	1	64	35	11.2
TOSRE	14	70	39	11.1
TOSCRF	2	211	138	35.8

Correlations Among Measures

	1	2	3	4	5
1. FCAT SSS					
2. Oral Reading Fluency	.55				
3. FCAT Maze	.32	.56			
4. Espin Maze	.47	.62	.47		
5. TOSRE	.56	.62	.57	.75	
6. TOSCRF	.36	.24	.00	.14	.30

At Tenth grade, the FCAT mazes were not as strongly related to the FCAT scores as were Oral Reading Fluency and the Test of Sentence Reading Efficiency. The Espin mazes were more strongly related to FCAT performance than were the mazes constructed from FCAT passages. This is a bit puzzling, since the relationship between FCAT mazes and the FCAT scores is so much weaker at 10th grade than at 6th and 8th. One potential explanation for this is that the range on the FCAT Mazes tests seems to be somewhat constricted compared to the 8th grade. It might also be the case that the high school students, in a group testing situation, took the FCAT Mazes test less seriously than did the middle school students.

Discussion

On the basis of these findings, we are encouraged to proceed with the development of the FCAT mazes test as a potential replacement for the ORF test for progress monitoring in middle school. Further study of its potential in high school needs to be undertaken in order to determine whether the low relationships in the present study were a characteristic of the specific sample used, or the specific passages used, or the engagement in “group testing” for students in high school.

References

Buck, J. & Torgesen, J. (2003). The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test. Technical Report #1, Florida Center for Reading Research

Espin, C.A., Busch, T.W., & Shin, J. (2001). Curriculum-based measurement in the content areas: validity of vocabulary-matching as an indicator of performance in social studies. Learning Disabilities Research & Practice 16 (3), 142-151

Espin, C.A., Scierka, B.J., Skare.S., & Halverson. N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary students. Reading and Writing Quarterly, 15, 5-27

Jenkins, J.R., Fuchs, L.S., van den Broek, P., Espin, C., & Deno, S.L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95, 719-729.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Schatschneider, C., Buck, J., Torgesen, J., Wagner, R., Hassler, L., Hecht, S., & Powell-Smith, K. (2004). A Multivariate Study of Individual Differences in Performance on the Reading Portion of the Florida Comprehensive Assessment Test:A Preliminary Report. Technical report #5 , Florida Center for Reading Research.