

FLORIDA DEPARTMENT OF EDUCATION



CHARLIE CRIST
Commissioner

BETTY COXE
DEPUTY COMMISSIONER
EDUCATIONAL PROGRAMS

CONTACT PERSON

NAME: Dr. Thomas Fisher
PHONE: (850) 488-8198
SUNCOM: 278-8198
EMAIL: fishert@mail.doe.state.fl.us
DPS: 03-015

MEMORANDUM

TO: District School Superintendents

FROM: Betty Coxe *BC*

DATE: August 14, 2002

SUBJECT: FCAT DEVELOPMENTAL SCORE SCALE

The purpose of this memorandum is to transmit a Technical Assistance Paper on the topic of the FCAT Developmental Score Scale. An executive summary of the paper follows.

The FCAT was originally developed to test reading in grades 4, 8, and 10, and mathematics in grades 5, 8, and 10. The test results were reported in terms of a succession of annual "status reports" that revealed the performance of Florida students who were in different cohorts. The data were reported for each subject using a score scale that ranged from 100-500, and each scale was separately computed for each grade level. Progress over time was reported as changes in the performance of each grade level group – i.e., this year's fourth grade students were compared to last year's fourth graders to see if the average score changed or, for example, if more students were earning "Level 3" scores.

In 2001, the FCAT program was expanded so that the tests now are being administered in all grades 3-10. This offers the opportunity to introduce a new score scale that will link adjacent grades together and permit progress to be tracked over time, based on what is commonly called a "developmental scale." Since this concept is often used with commercially produced, national norm-referenced tests, it may be familiar to you.

District Superintendents
August 14, 2002
Page Two

The developmental score scale was created separately for each subject area through a research design that placed items for a designated grade level on the adjacent grade level tests. For example, the grade six test contained some fifth and seventh grade items, although the student's FCAT score was calculated using only the items for the grade in which he/she was enrolled. Statistical analyses yielded a series of conversion equations that placed all grades on the same scale. The traditional 100-500 FCAT scale still exists, but now there also is a developmental scale for reading and mathematics that ranges from 0 to 3000.

The effect of this improvement is that student performance across the grades can be tracked across this scale. Third graders' scores will be on the lower end of the developmental scale while the scores of tenth graders will appear on the higher end. As a student moves from grade to grade, his/her performance can be monitored and compared to the performance of other students in Florida. Most importantly, the yearly progress of each student can be reported by the change in the developmental scale scores.

Because of the complexities of introducing the concept of a developmental scale and because of the difficulties in matching student records from 2001 to 2002 in a very short period of time, the 2002 FCAT results were not initially reported using the developmental scale. Results were reported in terms of changes in the FCAT Achievement Levels from 2001 to 2002. However, the electronic file districts received in May 2002 contained test scores for 2001 and 2002 and, through use of the information in the Technical Assistance Paper, gain scores can be calculated by each school and district, if desired.

Two new reporting steps are being implemented to utilize the power of the developmental scale. First, the Department created a special interactive Web site that will permit any user to type in an FCAT score on the 100-500 scale and have it automatically converted to the developmental scale. The student's scores can be compared to state-level values for greater understanding of growth patterns and trends. Second, we anticipate the release of a new report form in August 2002 that will provide each student's gain using the developmental scale. In May 2003, we expect to release all student results using both the traditional and the developmental scale scores. The Department is making preparations for new reports to show these gain scores.

Questions about this FCAT innovation in reporting can be directed to the Assessment and Evaluation Section at (850) 488-8198.

BC/tft

cc: District Assessment Coordinator
District Assistant Superintendents

Technical Assistance Paper

FCAT Developmental Scale

Preface

The purpose of this Technical Assistance Paper is to describe the new developmental score scale available for the FCAT administered in March 2002. This paper is intended for use by district assessment coordinators, district staff, and school personnel who need to be generally conversant with the topic and understand potential uses of the developmental scale.

Background

Classroom teachers regularly administer tests to students and typically report the results in terms of a raw score (17 out of 20 questions correct) or a percentage (85%). While these methods are acceptable at the classroom level, they are inadequate for use with large-scale assessment tests that must be equated from year to year and form to form and for which important, high stakes decisions are made.

For reporting purposes, large-scale assessment tests are statistically analyzed to determine the proper conversion from the raw score to a scale score. Educators are familiar with such scores as are reported with the SAT[®], which uses a scale of 200-800, or the ACT[®], with its scale of 1-36. Clearly, neither of these score scales describes the examinee's raw score, and one cannot ascertain the total number of raw score points by looking at the range of the scale scores.

Testing organizations select score scales that they believe will be understandable and easy to communicate. Over time, the score scale becomes readily "understandable" even though the examinees themselves may not completely understand how the numbers were calculated. The selection of the scale itself, however, is a decision with many options although there are some practical constraints. For example, one would not want a scale that permitted reporting negative scores for some students.

When FCAT was first conceived, it was designed to measure reading in grades 4, 8, and 10 and mathematics in grades 5, 8, and 10. Student proficiency was reported using the FCAT scale score, which ranges from 100-500. When the scale was first derived, the state average was 300, although after the test was implemented the average shifted from 300.

FCAT results have been reported as a series of status assessments. In other words, the fourth grade student scores in reading were reported annually, and progress was detected by changes in the average score. This method typically is used with statewide assessments and is also used with the National Assessment of Educational Progress (NAEP). It is immediately apparent, however, that measuring progress over time for

individual students and monitoring the achievement of students in multi-grade schools would be enhanced if the test results could be reported using a score scale that tracks the same student(s) as they move across the grade levels. Some commercially available nationally norm-referenced tests use such cross-grade scales. These scales are known either as “vertical” scales or “developmental” scales. The latter term will be used with the FCAT program.

Methodology

To build a developmental scale for the FCAT that spans all grade levels from 3 to 10, it was necessary to embed items from adjacent grade level tests into field test item positions. The embedded items from adjacent grade level tests are called “linking” items. This methodology assured that no student’s score would be affected by this research design. This was done in the 2001 administration of the FCAT, and the resultant data were analyzed in the summer of 2001.

Since there is no grade level tested lower than grade 3, there are no items from a lower level to place on the grade 3 test; therefore, only grade 4 items were embedded in the grade 3 test. At grade 4, it was possible to embed linking items from grade 3 as well as items from grade 5. The grade 10 test only contained linking items from the grade 9 test. By only using item positions ordinarily reserved for field test items, not all test forms contained linking items, and not all students actually had test forms that contained linking items. Student performance on the linking items was not used in calculating the final test results for the student, school, district, or state. The data were used strictly for psychometric analysis purposes.

The psychometric analysis began at grade 3 and worked up to grade 10. The objective was to link together the score scales from adjacent grade levels. For example, 4 linked to 3, 5 linked to 4, and so forth. The analysis used Item Response Theory (IRT) techniques like those which have been used by the FCAT program since 1998 to generate the traditional 100-500 score scale.

To create the final developmental scale, it was necessary to anchor it at two points, the lower end and the upper end, and decide what the final range would be. After analyzing the data in various ways, it was decided that the final scale should be anchored on grade 3 so that the average developmental score (in 2001) would be 1300, and for grade 10, the average score would be 2000. As a result, all students will have scores above 0 with the highest score being approximately 3000.

There are 8 grade levels and tests in both reading and mathematics, so there are 16 sets of equations that are needed to convert the FCAT scores from the traditional 100-500 scale to the developmental scale. These linear equations are of the form $Y = a(X) + b$, and the “a” and “b” values are shown below in Table I.

Table I				
Transformation Values: Traditional Scale Score to Developmental Scale Score				
	Reading		Mathematics	
Grade	Slope	Intercept	Slope	Intercept
3	6.072	-521.569	4.625	-87.499
4	5.860	-291.417	4.373	143.659
5	5.598	-86.090	4.719	97.196
6	5.547	-15.777	4.305	339.511
7	5.239	147.116	4.035	554.057
8	4.761	410.068	3.948	630.602
9	5.426	229.605	3.396	898.322
10	5.410	303.295	4.102	658.073

The conversion equations can be applied to the 2001 FCAT statewide mean scores on the traditional (100-500) scale (for grades 3 and 10) to determine the equivalent values on the developmental (0-3000) scale. These values are shown below in Table II. Since these are statewide mean values, it is possible for an individual student to earn a minimum score lower than 300 or a maximum score higher than 2000.

Table II		
Mean Scores Across Scales		
<i>READING</i>		
Grade	Traditional FCAT Scale Score	Vertical Scale Score
3	300.0	1300
10	313.6	2000
<i>MATHEMATICS</i>		
Grade	Traditional FCAT Scale Score	Vertical Scale Score
3	300.0	1300
10	327.1	2000

The Department has prepared tables that contain the conversions between the traditional FCAT 100-500 scale and the new developmental (0-3000) scale. These will be made available to districts electronically for those who do not care to calculate the conversions. By using the Department's tables, all districts can make conversions consistent with those prepared by the Department and its scoring contractors. If a district or school prepares its own conversions, there will be differences in rounding

conventions, and the end result may be different from what is provided in the tables distributed by the Department.

Applications

Figures 1 and 2 attached to this paper depict the “interim” reading and mathematics developmental scales for standard curriculum students in 2001. (The final version is shown in Figures 3 and 4 using the 0 to 3000 scale.) A number of observations can be made from these charts.

- 1. Understanding Trends.** It is clear that actual mean scores in 2001 were not linear across the grade levels, and there were distinct places where the average performance slightly decreased (e.g., grade 9 reading) or was almost level (e.g., grade 8 to 9 math). While the developmental scales could have been modified in some way to smooth out these differences, this was not done because it would hide important educational phenomena. It is important, for example, to understand that 6th graders’ math performance may decrease compared to 5th graders’ performance, perhaps because 5th graders have been tested in math for several years while 6th graders have not. While the reasons for the dip in performance are not entirely clear, it is important for educators to monitor test results over time and review instructional programs to gain a better understanding.
- 2. Coordinating Achievement Levels.** The work on the developmental scale was completed in the fall of 2001 as the Department was working on the specification of the cut scores for the FCAT Achievement Levels at all of the new grade/subject areas. As part of this process, it was important to understand the relationship between the existing cut-scores in grades 4, 5, 8, and 10, and those that were being recommended by the educator panels. The Department used the transformations to place all cut-scores for all grade levels on the developmental scale. By plotting the values across the grades, it was possible to see where the proposed new cut-scores were out of alignment with the existing cut-scores. Adjustments were made, and the final recommendations to the State Board of Education were based on this “smoothing” that was made possible only because of the availability of the developmental scale. Figures 3 and 4 provide a display of the achievement levels from the 2001 test administration.
- 3. Monitoring Individual Student Performance.** Imagine a student who attends a public school in Florida in grades 3, 4, 5, 6, and 7. This student will take the FCAT in reading and mathematics each spring, and his/her scores will be reported using the 100-500 score scale. The student’s performance can be compared to the five Achievement Levels. With a developmental score scale, it becomes possible to see more precisely how the student progressed, even if that student has maintained the same Achievement Level.

By using the developmental scale, parents and teachers can chart their students’ growth as they move across the grade levels. The 100 – 500 scores reported each year would be converted to the developmental scale, and the converted scores would show

that the student was making adequate progress. In fact, the student's progress can be tracked in terms of the mean score for all students in the state or in terms of the cut-off scores for the Achievement Levels. Such charts, or data plots, will make it possible to answer the question, "Is this student making one year of progress for each year in school?" and, "If this student keeps making the same amount of progress, is he/she going to be ready for the grade ten graduation testing requirement?" Clearly, this is powerful information to possess.

The use of the developmental scores make it possible to monitor the growth of students in different schools, in different instructional programs, in different scheduling modes, in classes taught by different teachers, etc. Such analyses can shed new light on our approaches to instruction and the provision of services to students.

Reporting FCAT Scores

In May 2002, the Department's scoring contractor, NCS Pearson, prepared individual score reports similar in design to those used in 2001. However, there were a few changes in order to begin reporting student improvement from 2001 to 2002. The student reports showed whether the student gained, stayed the same, or lost ground in terms of the FCAT Achievement Levels. The student's Achievement Levels for the two years was shown, and there were one or more statements on the report to interpret the change in performance. The 2002 student FCAT reports do not include developmental scale scores.

The electronic file of student results that the districts received in May 2002 contained two years of test scores. For students whose 2002 test results could be matched to their 2001 test results, traditional scale scores, achievement levels, and developmental scale scores for both years were provided. Districts are able to use their FCAT data files along with the conversion tables provided by the Department and convert all student FCAT scores to the developmental scale. The converted scores can then be used for a variety of research and evaluation purposes in addition to reporting to parents.

Final Comment

The FCAT program has been successfully implemented across all grades 3 – 10 in reading and mathematics. With the implementation of the new developmental scale, it is possible to monitor student progress across the grade levels. The measurement of "gains" is possible, and their use by districts and school permit educators to improve their understanding of educational processes and programs. Parents also benefit by being able to better understand their childrens' performance as they progress from grade to grade.

Scores on the developmental scale, however, must be interpreted with caution. For example, students achieving at the same developmental scale score on different grade level tests cannot be interpreted as equivalent. This is because the lower grade student may not have mastered all of the specific content assessed on the higher-grade level

test. Test content variations across grade levels must be taken into consideration when interpreting developmental scale scores.

Figure 1

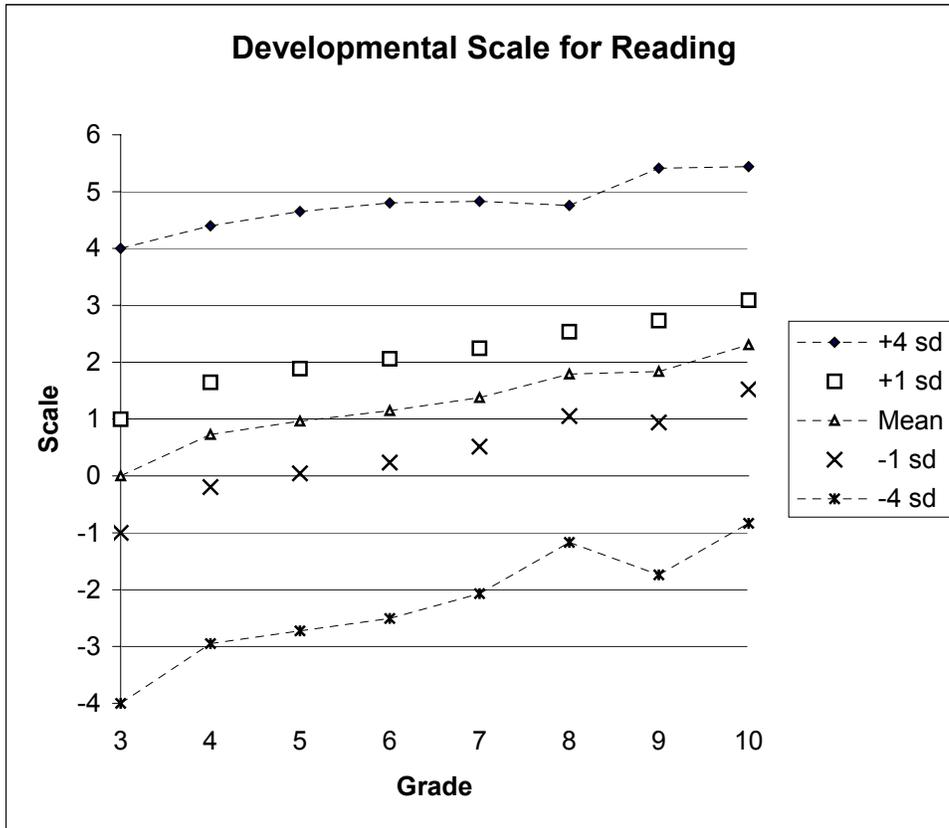


Figure 2

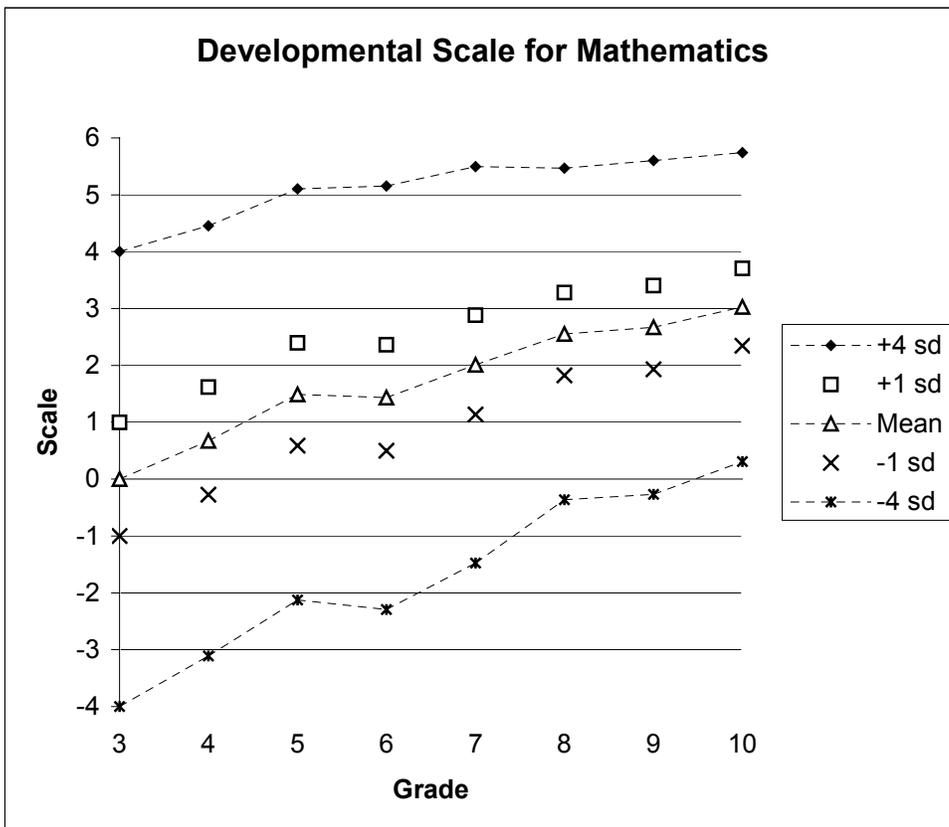


Figure 3

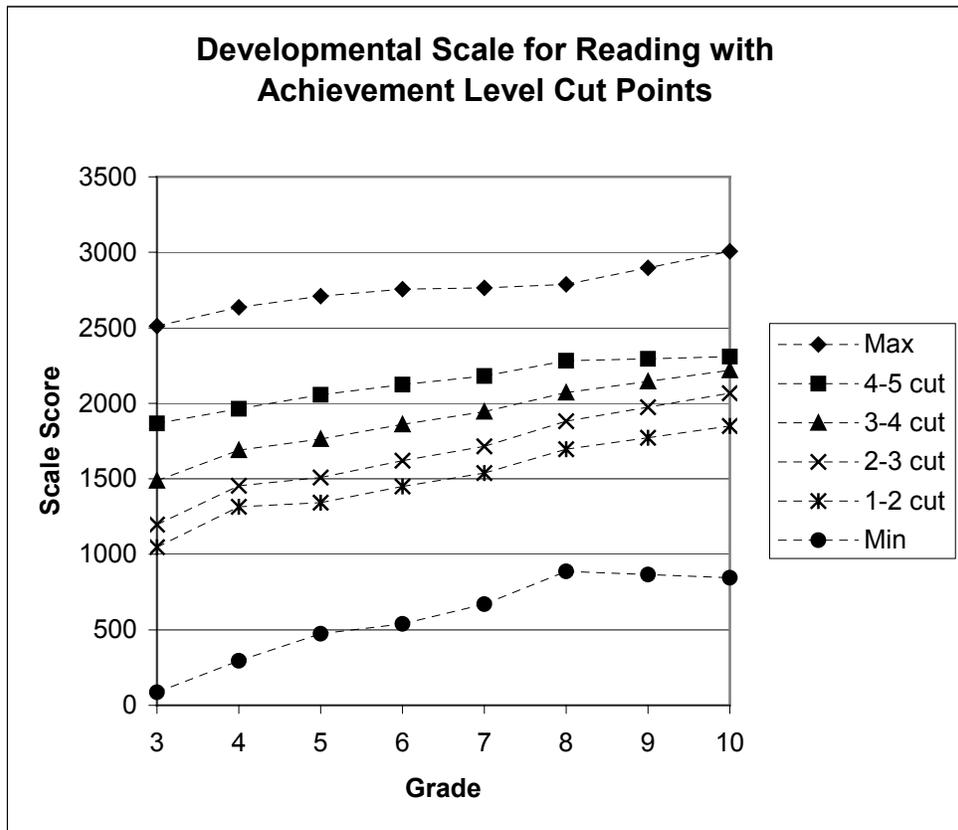


Figure 4

