Classroom Grades and Their Relationship to FCAT Scores

Introduction

The teacher's task of assigning letter grades to students and the public's interpretation of them can be perplexing. Does the student's grade represent the level of achievement, the gain in achievement, or some combination of the two? Is the student's effort included in the grade, or are high achievers given good marks regardless of effort? Are pupils marked according to their own potential learning ability or in relation to their classmates' achievement? There are no clear answers to these questions. Despite attempts at precise directives, practices tend to vary by school, teacher, and context.

Scores on an FCAT standardized test seem to be a different matter. Such tests are designed to be criterion-referenced, intending to determine whether a student has mastered the material taught in a specific grade or course. The apparent objectivity of the test and its connection to a specific curriculum gives the impression of a pure measure of student achievement. But tests, no matter how carefully designed, are inevitably imperfect and measure only a sample of student behavior over a very short period of time.

What exactly do classroom grades represent in Miami-Dade schools and what is their relationship to FCAT scores? How is it that some students can get very high classroom grades and still score at below-proficiency levels on the FCAT? What should we expect the relationship to be between classroom grades and FCAT scores? These and other issues regarding grades and test scores are explored in this paper.

The M-DCPS Definition of Grades

The Student Progression Plan tries to make clear the definition of grading in our district. According to the official district stance, grades are to be pure reflections of academic achievement and are not to include references to effort. Of particular interest in this definition and elsewhere in the Student Progression Plan are references to comparing the student's performance to norms representing "the typical student in the same program or course." What is left unspecified is the identification of these typical students. Are the norms supposed to relate to other students in the same class, other students in the teacher's previous experience, other students in the district, or other students in some even more encompassing population?

> **M-DCPS Student Progression Plan 2007-2008**
> "Academic grades are to reflect the student's academic progress based on the competencies/benchmarks for the grade level/course in which the student is enrolled. The grade must not be based upon the student's effort and/or conduct. The grade must provide for both students and parents a clear indication of each student's academic performance as compared with norms which would be appropriate for the grade or subject."

In another part of the Student Progression Plan, letter grades are associated with specific percentages: "A" (90-100%), "B" (80-89%), "C" (70-79%), "D" (60-69%), and "F" (0-59%). It is unclear whether these percentages refer to the amount of material mastered, or the quantity of competencies achieved, or even possibly the number of test

items answered correctly. Since most of the classroom tests are designed by individual teachers and vary in difficulty and scope, there is considerable room for variation in applying these standards. Moreover, it is difficult to see how these percentage references in the Student Progression Plan fit together with the suggestion that grades reflect student performance norms.

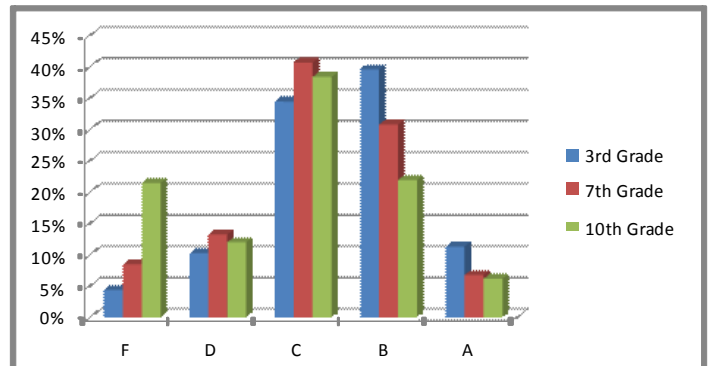## Norm-Referenced and Criterion-Referenced Grading

Just as we talk about norm- and criterion-referenced tests, so we can make the same distinction among classroom grading systems. In norm-referenced grading, students are compared to one another and given a grade relative to their standing in the group. This is the "grading on a curve" approach wherein, characteristically, the percentage of students receiving each type of grade (A, B, C, etc.) is predetermined. Criterion-referenced grading, on the other hand, is intended to be interpretable in terms of the students' accomplishments on a clearly defined set of tasks. Here the grades describe what a student can do, without reference to others' performances. The prescriptions for grading in the M-DCPS Student Progression Plan seem to be a mixture of the two grading methods, with emphasis on percentage ranges in one part and reference to norm comparisons in another.

In practice, the differences between norm- and criterion-referenced grading are mostly a matter of emphasis. Both require a clearly defined curriculum, a representative set of test items, and good test construction qualities. One stresses the discrimination among pupils, the other highlights description of performance. Customarily, in the process of setting performance standards for criterion-referenced grading, the psychometrician refers to norm distributions appropriate to the grade level and age of the students. There is a tendency to expect both criterion and norm interpretations from any grading system, and yet, neither puts specific restrictions on the distribution of grades.

The Distribution of Grades in Our District

In the traditional scheme for norm-referenced grading, the midpoint of the distribution is given a grade of C and a predetermined percentage of students receive each grade, A through F. In practice, few educators are willing to follow this orthodox approach with complete consistency.

In the graph to the right, the actual percentage of each type of grade in M-DCPS is shown for three different grade levels. The grades depicted here are from the final course grade in the main reading/language arts course for each of the depicted grade levels. Distributions for other grade levels and other content areas show the same tendencies. We can see that the distribution of grades tends to be somewhat positively skewed for the lower grade levels and negatively skewed for the higher grade levels. In any case, the great majority are in the C and B grade ranges. Although these distributions may be within common expectations overall, the pattern for individual schools can differ remarkably.
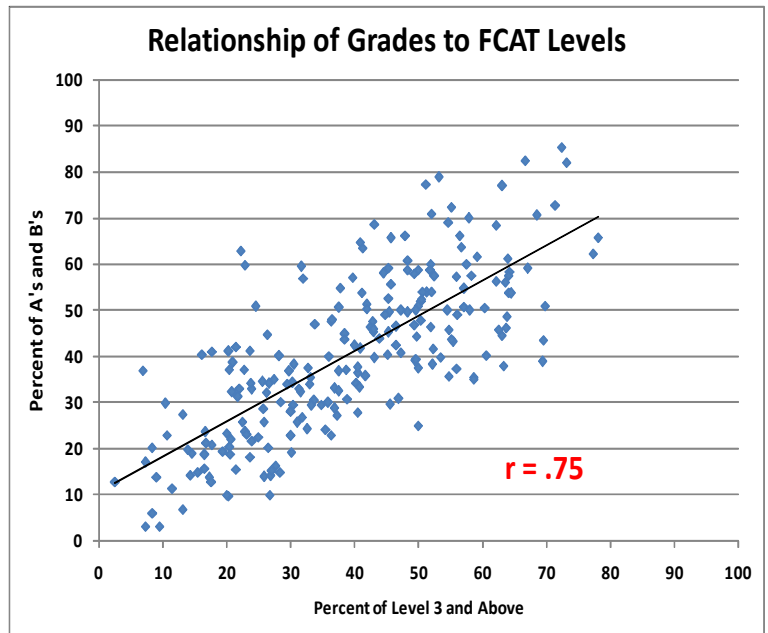
Differential Grading Distributions

In this graph we can see very different grade distributions between two selected schools in their 3rd grades. The High-Performing school is here defined by having a much greater proportion of students scoring in the higher levels of the FCAT. It is easy to see that the higher performing school, in an absolute sense, also has a greater proportion of high classroom grades.

Of course, this picture only shows two schools specifically selected to illustrate the potential differences in possible grade
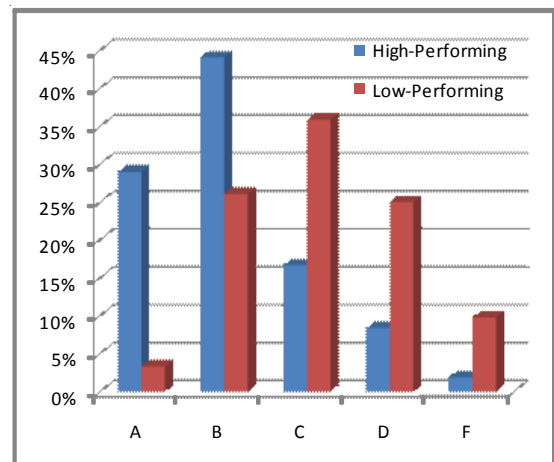
distributions. The graph to the right summarizes the relationship between high grades and high test scores for all our elementary school third grades on the FCAT Reading Test. Each dot on the graph represents the percent of students scoring Level 3 and Above on the FCAT and the percent of students with either an A or a B in the Language Arts. The closer the dots are to the summary line, the more consistent is the relationship. Although some schools deviate considerably, there is a strong general trend of matching the percents in high grades and test scores. This kind of pattern provides support for the hypothesis that classroom grades are assigned, at least in part, on a criterion-referenced basis.

Relationship of Grades to FCAT Levels

r = .75

Percent of A's and B's (y-axis)
*Percent of Level 3 and Above* (x-axis)

These kinds of consistent adjustments in grading distributions suggest that the teachers in our schools have some kind of external anchor for judging academic performance. This kind of external reference may be found in a consistent set of criterion for mastery of a content area based on a standardized curriculum, or it may be the general understanding of performance levels found in a standardized test.

Displacement of Distributions Between Schools

When we compare the relationship between FCAT scores and classroom grades among different schools, we can see a shifting of the complete distributions. The graph below depicts the average FCAT score at each classroom grade for the 10th grade in four selected high schools. Two of the high schools have overall high performance and the other two high schools have overall low performance as indicated by the percent of students scoring Level 3 and above on the FCAT.

In this graph we can see that the average FCAT score for the A students in the low performing schools is equivalent to the average FCAT score for the C students in the high performing schools. Furthermore, students graded F in the high performing schools are, on average scoring higher than students graded C in the low performing schools. If we judge absolute academic performance by FCAT scores, it is apparent that a particular classroom grade in one school can mean something quite different in another school.

Relationship of FCAT Scores to Final Grades

Overall High Performing High Schools

Overall Low Performing High Schools
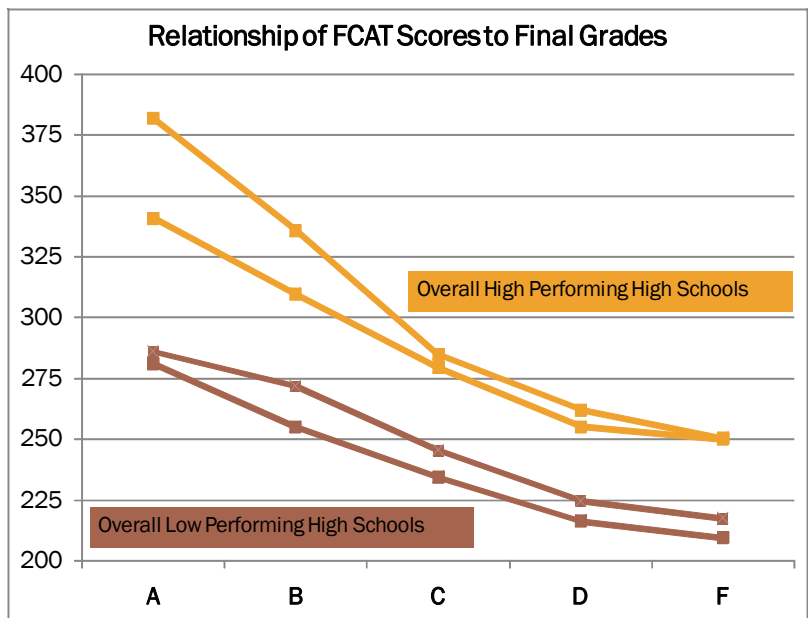
Once again, the schools in the above graph were selected specifically to clearly illustrate the kinds of displacements in grades that might occur. In a broader perspective, the following graph shows the relative displacement of grade distributions to 10th grade FCAT Reading scores for all the regular high schools in the district.

Apart from a few glaring exceptions, there is a very consistent pattern of grade displacement. As the average FCAT score corresponding to the A students decreases, the average FCAT scores for the C and F students decrease by a proportional amount. We can also see quite a few schools in which the C students have scores

higher than the A students of other schools, and many cases in which the F students have scores higher the C students of other schools. This kind of pattern provides support for the hypothesis that classroom grades are assigned, at least in part, on a norm-referenced basis.

Displacement of Grades by Average FCAT
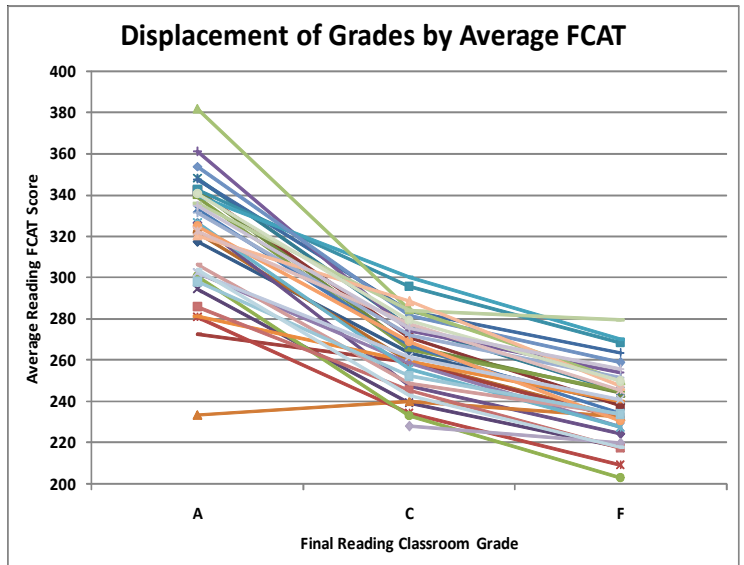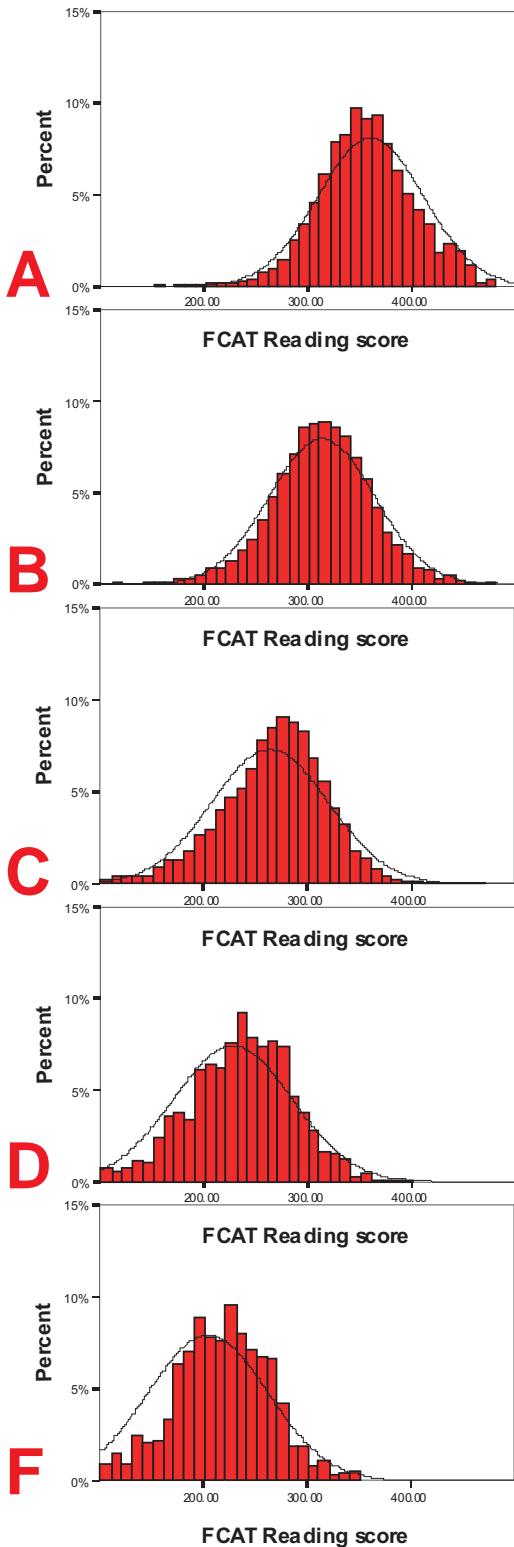
FCAT Scores Associated to Grades

The graphs at the left depict the distributions of 3rd Grade FCAT Reading scores for students receiving each of the final classroom grades A through F for language arts courses. The graphs are aligned so that the appearance of horizontal shifting of FCAT scores is readily apparent.

In general, the pattern of FCAT distributions by classroom grade is what one might expect -- the higher the classroom grade, the higher the distribution of FCAT scores. This demonstrates that there is a basic correlation between grades and scores.

However, there is considerable overlap between FCAT score distributions between grades. Clearly there are many B students scoring as high or higher on the FCAT than many A students. Overlap of scores of this type may be expected to occur between any consecutive set of classroom grades. However, as is evident in the set of graphs, the overlap extends beyond consecutive grades. One can see, for instance, that there are many C students scoring higher on the FCAT than do some A students. Of course, the degree of overlap observed in these graphs is over the entire third grade -- the overlap of FCAT scores across classroom grades is much less when we restrict our observations to individual schools.

4

Discussion

The rather complicated data patterns discussed in the preceding section lead to a rather simple interpretation of the relationship between classroom grades and standardized test scores. Within any classroom, the relationship between grades and test scores is as one might suspect: higher grades are associated with higher test scores. If one were to rank the students in any one classroom on the basis of their standardized test scores, the order would generally agree with that derived from a ranking by classroom grades.

The relationship between grades and test scores (again, within any one classroom) is not perfect. This should come as little surprise, since classroom grades are qualitative ratings for academic performance over the entire year, as opposed to quantitative scores of a sample of academic behaviors over a few days. In a discussion on a related issue, the Stanford test publishers, themselves, warn against assumptions of a strong relationship between test scores and classroom grades. In addressing the potential misuse of tests involving the possible assigning students grades on the basis of their achievement test scores, the publishers state,

> "The assignment of grades is a personal process that is subject to the standards of the individual teacher. Standardized achievement tests simply are not designed to provide comprehensive coverage of what is taught at a particular grade level or in a particular course. As such, they provide only a sampling of the student's knowledge and skills."(*Guide for Organizational Planning,* Stanford Achievement Test Series, The Psychological Corporation, Harcourt Brace Jovanovich, Inc.)

The observed relationship between classroom grades and standardized test scores is relative (i.e., the rankings agree) rather than absolute (i.e., a specific grade corresponds to a specific test score). Naturally, this relative relationship exists strictly within grade levels and within course levels. It is apparent that this relative relationship is further confined to individual schools. Consequently, an A in a third grade regular reading program in *this* school may not mean the same (in terms of absolute academic achievement) as an A in a third grade regular reading program in *that* school. This school specificity may indicate that, to some extent, grades are norm-referenced relative to the typical performance at the individual school.

On the other hand, grade distributions do not maintain the same shape from school to school. Schools with higher overall standardized test performance tend to have greater percentages of higher grades. Apparently, there are forces (e.g., regulated curricula, publisher-designed tests, etc.) that tend to equate grades across different schools. These forces are evidently stronger in early grade levels but do not completely overcome the school specificity of the grade-test score relationship.

We may expect to see classroom grades directly associated with FCAT test scores, and there is some evidence of this kind of absolute, criterion-referenced grading system. We may also expect to see the proportions of the different kinds of grades be similar from one school to another, and there is some evidence of this kind of relative, norm-referenced grading system. What may not be immediately obvious is that these expectations are in conflict with each other to a certain extent. The range of FCAT test scores varies among schools. If grades were strictly absolutely related to test scores, then, contrary to relative grading, we would see some schools with almost no A's and B's, and others with almost no D's and F's. In fact, we do see this in extreme cases. If grades were strictly relatively distributed in schools, then, contrary to absolute grading, we would see some C students scoring higher than some A students, and some F students scoring higher than some C students. In fact, we also see this in extreme cases. The kind of grading system we observe in our district seems to be a measured compromise between these absolute and relative trends. Grades relate directly to FCAT scores, but the grading scale tends to be shifted somewhat from school to school. This is a compromise between norm-referenced and criterion-referenced grading systems. It is the type of compromise relationship between grades and test scores we would be likely to observe at any time with any standardized test.

When teachers assign grades to students, they act as both judges of and advocates for the students. Teachers can act in these roles in ways that standardized tests cannot. However, these roles of judge and advocate often diverge. For grades to serve their ultimate purpose of facilitating learning, a compromise in the teacher's roles and a compromise in the relationship of grades to standardized tests are probably both necessary and desirable.