Miami-Dade County Public Schools
giving our students the world

RESEARCH BRIEF
Research Services

Vol. 0509
June 2006

Dr. Terry Froman, Research Services
Vilma Rubiera, Assessment and Data Analysis

Florida's Pay-for-Performance Program:
The Potential Impact of the Value Table

Starting next year, a new pay-for-performance program will tie raises and bonuses for Florida's teachers directly to the standardized test scores of their students. The program, approved in February 2006 by the Florida Board of Education, is mandatory and intended to ensure compliance with a 2002 Florida law requiring performance pay for teachers. The centerpiece for the effort, known as E-Comp, requires all school districts in Florida to identify the top 10 percent of teachers in each content area and award them a 5 percent salary supplement. For a teacher earning the average Florida salary of $41,578, that amounts to just over $2,000.[1]

This marks the first time that a state has so closely linked teachers' wages to students' exam results. Schools in Houston, Denver, Minnesota and elsewhere have similarly tried to link teacher pay to performance, but those efforts have been either less focused on test scores or narrower in scope.

Identifying the Top 10 Percent

Some controversy still surrounds how the top 10 percent of teachers will be identified. Teachers of math and reading will be ranked exclusively according to how much their students improve in FCAT achievement levels over the previous year. Teachers of other subject areas must also be ranked objectively according to measures designed by the district. The State encourages teachers to be evaluated on test scores and other objective assessments, even for subjects such as art and music.

Although only the top 10 percent in each field will receive the 5 percent salary supplement, all teachers will be affected by the new pay policy because, according to the new rules, their annual evaluations will rely primarily on improved achievement by students. Furthermore, since it is unclear whether the state will provide separate funding for the bonuses, the other 90 percent of teachers not receiving extra money may question whether the bonuses are at their expense.

The Point System

The State has devised a system in which teachers will earn a set number of points for advancing their students from one specific level of proficiency to another. "One of the biggest questions, aside from how teachers of subjects such as special education and art might be measured objectively, is whether the point system will fairly evaluate teachers in schools where students are impoverished or lack English skills."[1]

The State's point system attempts to address these concerns in two ways. First, it awards teachers points not for test scores, but for improvements in test scores, so a previously low-scoring student will not necessarily be a disadvantage to a teacher's evaluation. Second, having analyzed historical test scores, the system purports to be "neutral" by assigning a different amount of points for different changes in FCAT proficiency level. To a certain extent, this is intended to accurately reflect the varying degree of difficulty of lifting students from one level to another.

Florida's Value Neutral Tables

The State has released sample value tables illustrating the differential points for changes in achievement levels and how these points would be combined to yield a final score for a teacher.[3] For example, the Sample Elementary School Reading Value Table is reproduced below in Table 1.

Table 1: Sample Elementary School Reading -- Value Table

		Year 2 Level			
Year 1 Level	**1**	**2**	**3**	**4**	**5**
1	0	190	300	415	500
2	0	75	175	210	250
3	0	0	120	155	175
4	0	0	0	135	180
5	0	0	0	70	140

For a given teacher, each student would be awarded the points associated with the particular change in levels between the previous year and the current year. Thus, a student that moved from level 1 to level 2 would be worth 190 points to the teacher, and a student moving from level 3 to level 4 would be worth 155 points.

Although the value tables vary from Elementary to Middle to Senior High, and between subject areas, certain attributes of the table illustrate policy decisions that reflect the differing values for differing educational outcomes as determined by the State. For example, the value for any student in Level 1 in the second year is zero. Students that maintain the same performance level at higher levels are awarded more points than students that hold their own at lower levels. Even students who drop a level from Level 5 to Level 4 receive some points.

The final score for teachers would be the average of value table points for all of their students in the particular subject area.

Problems in the Development of a Value Table

In a media brief released by the State, the idea of a value table was attributed to the National Center for the Improvement of Education Assessment (NCIEA). A paper by Richard Hill of NCIEA, lays out the basic account of the development of value tables.[2] The goals were to create a system that (a) would be easy to understand, (b) would be easy to compute, and (c) would be flexible with regard to subject and grade levels. With respect to these goals, their efforts appear to be quite successful. However, the Hill paper leaves unresolved certain admitted reservations regarding some statistical aspects of value tables.

In addressing whether value tables are a valid and reliable way to determine teacher effectiveness, the Florida DOE states that, "compared to other methods of assessing value, the value table has a high correlation with analysis of covariance and hierarchical linear models."[3] However, the Hill paper states that

> "...initial analyses show a relatively low correlation among school growth scores [i.e., analysis of covariance and hierarchical linear models] depending on the Value Table chosen. We also know some Value Tables that appear on their

> surface to be appropriate turn out to be poorly correlated with other school-level statistics that should be indicating school effectiveness. This suggests that the process for establishing the Value Table to be used in a state school-level accountability system needs to be better understood than it [is] now. For example, we have developed a procedure similar to standard setting that allows policy-makers to articulate the values they wish to see reflected in their accountability system. We do not yet know how to create a Value Table that accurately reflects those values." (Hill, 2006, p.6).

Doran and Cohen (2006) showed that the process of constructing vertical scales introduces additional variance components that lead to underestimation of standard errors.[4] Consequently, gains may appear to fluctuate over time due to measurement noise, not instructional quality. This problem is exacerbated by trying to apply value tables to teacher assessment instead of the school-level assessment for which it was designed.

On the other hand, "looking at the actual results of how students change levels from one year to the next almost certainly overestimates the amount of true regression going on."[5]

> "Although we have almost certainly overstated the amount of regression in our calculations...,and thereby understated the difficulty teachers with lower performing students will have achieving the same scores from this Value Table as teachers with higher performing students, policy makers were satisfied with that. One of the considerations that led to their acceptance of this Value Table is that there is a perception within the state that lower performing students already are receiving less effective instruction, and the fact that their scores tended to be lower was probably an appropriate reflection of that fact." (Hill, 2006, p.5)

Contrasting Priorities

The value tables do not regard all gains with equal value. In fact, the whole idea of a value table invites the policy makers to explicitly state which educational outcomes they value most highly and to what degree. There is nothing inherently wrong with this subjective valuing process; it can be viewed as setting different incentives for different accomplishments. However, the lack of alignment with other value systems imposed on teachers can weaken the effectiveness of all performance incentives.

For example, under NCLB, all gains below Level 3 are considered inconsequential, gains from below Level 3 to Level 3 and above are highly and equally valued, and gains above Level 3 are once again considered inconsequential. Under the Florida School Grades initiative, both gains in level and maintenance of level are valued equally beyond Level 3. Furthermore, some students maintaining at Levels 1 and 2 are valued as gains if they are making normal progress. Clearly the rewards to the District under NCLB and the rewards to the school under the A+ Plan conflict to a certain degree with the rewards to the teacher under the performance pay system.

Expected Points by Starting Level

One way to get some idea of the potential impact of the value tables is to apply the point system to actual historical data. For illustration purposes, only the elementary reading value table will be considered. Expectations will vary depending on the specifics of the value table, but the general trends will hold across tables. Table 2 presents the percentage of students making each kind of level change in the FCAT Reading between 2004 and 2005 for grades 4 and 5 combined.

In the final column of Table 2 are the average number of points each student would contribute to the teacher assessment given the starting achievement level of each student. Although the average value for students starting in Levels 2 through 5 are roughly equivalent, the average points for

Level 1 students is less. This trend is consistent across different value tables and subject areas. To a degree, the fewer average points for Level 1 students is due to the large percentage of these students who stay in Level 1 and, thus, contribute no points. Apparently the disadvantage to teachers of Level 1 students is intended to be partially offset by the much higher potential point value for these students when they advance in level.

Table 2: Percentage of Students by Starting level

	Year 2 Level					Average
Year 1 Level	**1**	**2**	**3**	**4**	**5**	**Points**
1	61%	22%	15%	1%	0%	**93**
2	22%	33%	40%	5%	0%	**105**
3	6%	16%	54%	23%	1%	**102**
4	1%	2%	28%	56%	14%	**101**
5	0%	0%	5%	51%	44%	**97**

Students Maintaining or Dropping in Achievement Level

One way to consider the potential impact of the value tables is to focus only on those students who are **not** gaining in achievement level. Of course, students starting out in Level 1 cannot drop in level, and those starting out in Level 5 cannot gain. But, among those students not gaining in level, the rewards to the teacher differ considerably among the various starting levels. Table 3 shows the percentage of students who maintain or drop an achievement level broken down by the starting achievement level.

Table 3: Students Maintaining or Dropping an Achievement Level

Year 1 Level	Percent Maintaining or Dropping a Level	Average Points per Student
1	61%	**0**
2	55%	**45**
3	76%	**85**
4	87%	**86**
5	100%	**97**

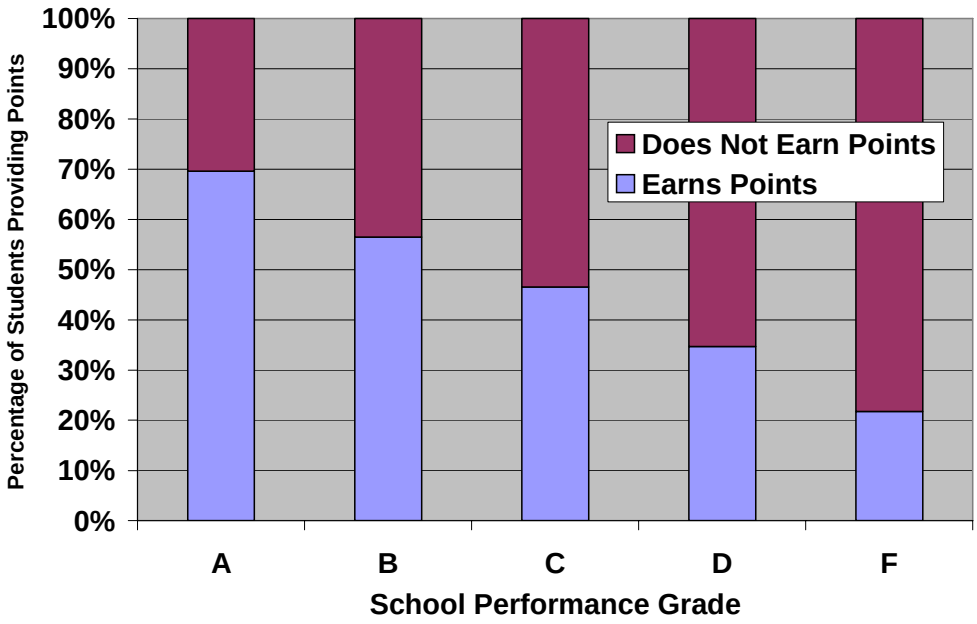
One can see from this table that, *among those students who are doing no more than maintaining their achievement level*, the higher the starting level, the greater the average points attributed to the teacher. Once again, this trend is apparently intended to be counterbalanced by the potential for much greater points if the teacher has students who have **gained** in achievement level.

Value Points and School Grades

If the value tables were truly neutral, one would think that there would be no particular advantage to the teacher to be in an “A” school versus and a school with a lower performance grade. Graph 1 depicts the percentage of students contributing points to the teacher assessment based on the school grade. The data is for 4th and 5th graders combined throughout the district for the changes in achievement level between 2004 and 2005 on the FCAT Reading test.

It certainly appears from this graph that teachers in the higher graded schools would have advantages in acquiring assessment points and, therefore, would be more likely to receive a salary supplement. Of course, the points per student are different depending on the specific levels from which changes are made. However, if the points are neutral across the different starting year achievement levels, the advantages would persist.

Graph 1: Based on FCAT Reading SSS 2004 to 2005



Value Points and School Performance Status

Proponents of the value tables would argue that the differences among school grades seen in Graph 1 are as they should be. After all, the grades themselves are partly determined by gains. Furthermore, it is not beyond reason that the better teachers (i.e., those receiving more value points) are more likely to be found in schools with higher grades.

To correct for the confounding of school grades and gains, it is possible to rank the schools in terms of their absolute performance status, not taking into consideration the way gains are typically measured in the determination of school performance grades. After ranking the elementary schools on the basis of the percent of students scoring 3 or higher on the FCAT Reading Test, three schools could be randomly chosen from the high and low end of the ranking. Tables 4 and 5 present the results for such an analysis using actual FCAT data from 2004-2005.

Table 4: Typical Teacher's Value Score at Low-Performing School

Average Free/Reduced Lunch=96%, Limited English Proficiency=37%, Mobility Index=37

Year 1 Achievement	Year 2 Achievement					
	1	**2**	**3**	**4**	**5**	**Total**
1	17% x 0 = 0	5% x 190 = 9.9	4% x 300 = 11.3	0% x 415 = 0	0% x 500 = 0	21.2
2	6% x 0 = 0	9% x 75 = 6.7	8% x 175 = 14.9	1% x 210 = 1.5	0% x 250 = 0	23.1
3	3% x 0 = 0	9% x 0 = 0	18% x 120 = 21.8	6% x 155 = 9.1	0% x 175 = 0	30.9
4	0% x 0 = 0	1% x 0 = 0	4% x 0 = 0	7% x 135 = 9.9	0% x 180 = 0	9.9
5	0% x 0 = 0	0% x 0 = 0	0% x 0 = 0	1% x 70 = 0.5	0.5% x 140 = 0.7	1.2
Total Value Score						**86.2**

Table 5: Typical Teacher's Value Score at High-Performing School

Average Free/Reduced Lunch=40%, Limited English Proficiency=16%, Mobility Index=20

Year 1 Achievement	Year 2 Achievement					
	1	**2**	**3**	**4**	**5**	**Total**
1	6% x 0 = 0	3% x 190 = 4.9	1% x 300 = 3.9	0% x 415 = 0.8	0% x 500 = 0	9.5
2	2% x 0 = 0	3% x 75 = 2.1	5% x 175 = 9.3	1% x 210 = 3.1	0% x 250 = 0	14.5
3	1% x 0 = 0	4% x 0 = 0	22% x 120 = 26.2	11% x 155 = 17.1	2% x 175 = 2.9	46.2
4	0% x 0 = 0	0% x 0 = 0	6% x 0 = 0	18% x 135 = 24.0	6% x 180 = 10.6	34.6
5	0% x 0 = 0	0% x 0 = 0	1% x 0 = 0	4% x 70 = 3.0	4% x 140 = 5.4	8.3
Total Value Score						**113.1**

The value points per student aggregated across grades 4 and 5 for the two groupings of schools are presented at the bottom of the last column of the tables. It is apparent that higher performing schools will have a greater expected number of value points per teacher (113.1 points) than would teachers from lower performing schools (86.2 points). This is consistent with the opinion of Hill, that the approach "understated the difficulty teachers with lower performing students will have achieving the same scores from this Value Table as teachers with higher performing students..."[2]

Conclusions

The State has said that "value-neutral" means that the table is designed to be neutral when it comes to acknowledging the improvements of students. An appropriate value is placed on each student given the likelihood for that student to make improvement. No matter where a student starts -- whether they start at the lowest achievement level or the highest -- a teacher will be given points based on how much his/her students improve.[3]

However, from the initial analyses presented in this paper, equity issues regarding the awarding of salary supplements to Florida teachers arose since teachers of students in the higher performance levels will have a distinct advantage in collecting value points and, therefore, be more likely to receive salary supplements. In addition, there is a growing body of research questioning the validity of measuring teacher effectiveness solely on the basis of student test performance. In the words of one such study, psychometricians tend to agree that scales spanning wide grade/developmental ranges also span wide content ranges, and that scores cannot be considered comparable across different grade levels. Martineau (2006) concluded:

> "This study demonstrates mathematically that the use of such 'construct-shifting' vertical scales in longitudinal, value-added models introduces remarkable distortions in the value-added estimates of the majority of educators. These distortions include (a) identification of effective teachers/ schools as ineffective (and vice versa) simply because their students' achievement is outside the developmental range measured well by 'appropriate' grade-level tests, and (b) the attribution of prior teacher/school effects to later teachers/schools. Therefore, theories, models, policies, rewards, and sanctions based upon such value-added estimates are likely to be invalid because of distorted conclusions about educator effectiveness in eliciting student growth." (p. 35).

Endnotes

(1) Peter Whoriskey, **Florida to Link Teacher Pay to Students' Test Scores**, The Washington Post, March 22, 2006.

(2) Richard Hill, **Using Value Tables for a School-Level Accountability System**. A paper presented at a symposium during the 2006 NCME Annual Conference that outlines the process of using Value Tables for a statewide accountability system, April 2006.

(3) Cathy Schroeder, Supporting materials for the Florida **Department of Education briefing regarding Pay for Performance Plan Value Neutral Table,** http://www.fldoe.org/news/2006/2006_04_05.asp, April 5, 2006.

(4) Harold Doran, and Jon Cohen, *The Confounding Effect of Linking Bias on Gains Estimated from Value-Added Models,* **Longitudinal and Value-Added Models in Education: Theory and Applications**, Robert Lissitz, ed., Journal of Applied Measurement Press, 2006.

(5) Richard Hill, Brian Gong, Scott Marion and Charles De Pascale, **Using Value Tables to Explicitly Value Student Growth,** National Center for the Improvement of Education Assessment, April 2006.

(6) Joseph Martineau, *Distorting Value Added: The Use of Longitudinal, Vertically Scaled Student Achievement Data for Growth-Based, Value-Added Accountability*, **Journal of Educational and Behavioral Statistics,** Vol. 31, No1. 1, Spring 2006.