STULL EVALUATIONS AND STUDENT PERFORMANCE

Stephen Newton, Ph.D.

Los Angeles Unified School District
Program Evaluation and Research Branch
Planning, Assessment and Research Division Publication No. 186

January 13, 2004

Introduction

The Stull evaluation provides a formal structure for administrators to evaluate individual teacher performance through directly observing and rating the teacher's classroom instruction. The results of the Stull evaluation can provide formative feedback for teachers about their strengths and weaknesses, and can play a key role in the process of dismissing teachers for teaching below the expected standard. Stull ratings thus can play a role in helping the District maintain quality control of instruction by identifying poor teachers that can then be provided coaching through the Peer Assistance and Review (PAR) program, or dismissed if they are unable to demonstrate improvement.

Given the significant consequences of the Stull evaluation for individual teachers and for their students, it is important to explore the relationship of Stull ratings with student achievement outcomes and with teacher characteristics. Specifically, this seeks to answer these questions about the Stull process in LAUSD:

- 1. What are the results of Stull teacher evaluations?
- 2. What is the relationship between Stull ratings and student achievement outcomes?

This analysis focuses on Stull evaluations of classroom teachers in elementary, middle, and senior high schools because other settings differ regarding types of data, student assessment, and student and teacher characteristics. Teachers in other settings (such as early education centers, special education centers, and adult schools) and non-classroom certificated staff are not included in any of the data presented here. Analysis of student test score outcomes requires focusing on students in elementary school because teachers at these grades typically are with students throughout the school day, and therefore instruct students for longer periods of time in some subjects (especially reading/language arts) and in more subject areas than teachers in middle school and senior high school.

What are the results of Stull teacher evaluations?

Permanent teachers receive Stull evaluations every other year, while non-permanent teachers receive Stull evaluations every year (LAUSD memo S-18). Teachers rated *Below Standard* Performance overall also receive a Stull evaluation every year until they are rated as *Meets Standard Performance* overall. Over the past four years, between 41% and 47% of teachers in LAUSD received Stull evaluations in any given year.

Teachers receiving Stull evaluations in the time period from 2000-2003 were rated on 25 items within five *Areas of Evaluation* (Achievement of Instructional Objectives, Preparation and Planning, Classroom Performance, General Professional Skills, and Punctuality and Attendance). Administrators conduct a formal observation of the teacher's classroom instruction and then complete the Stull evaluation form. In response to each of the 25 items, administrators mark *Yes* when performance meets expectations and *No* or *Needs Improvement* when it does not. Administrators also give teachers an overall rating of *Meets Standard Performance* or *Below Standard Performance*. The evaluation form is currently in revision during the 2003-04 school year, but the same form was used from 2000-2003.

One can see in Chart 1 and Chart 2 below that a small percentage of teachers are rated *Below Standard Performance* on the Stull *Overall Evaluation*. Approximately one-half a percent of elementary school teachers, one and a half percent of middle school teachers, and one percent of senior high teachers were identified each year as *Below Standard* between 2000 and 2003.

Bearing in mind that just under half of all teachers are evaluated in any given year, these percentages overstate the proportion of teachers rated *Below Standard Overall* compared with the total population of teachers in the district. When considering the total population of teachers, only about one in 400 elementary teachers, one in 130 middle school teachers, and one in 200 high school teachers are rated *Below Standard Overall* each year.

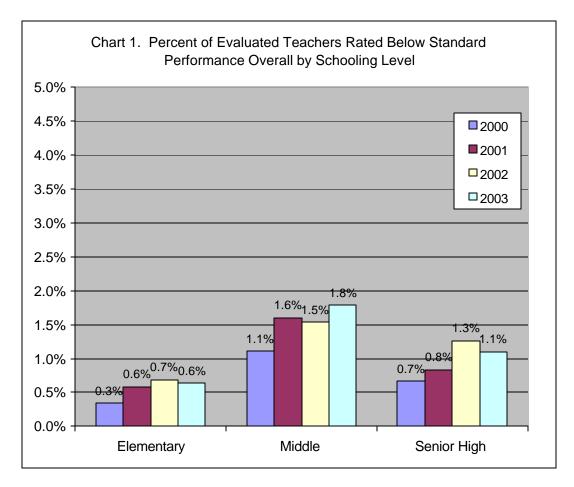
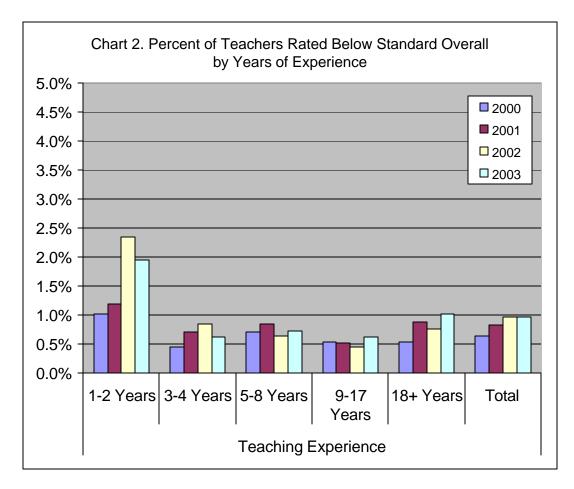


Chart 2 on the next page shows the percentage of teachers rated *Below Standard* according to level of experience. New teachers (1-2 years experience) were the most likely to be rated *Below Standard Performance*. This pattern was especially pronounced in the past two school years (2001-02 and 2002-03). These findings are consistent with results of another recent study in LAUSD that found superior results for experienced teachers compared with beginning teachers.¹

¹ "Students of novice teachers are less successful on standardized achievement tests than students of more experienced teachers by a wide margin." Cantrell, S. (September 2003). Pay and Performance: The Utility of Teacher Experience, Education, Credentials, and Attendance as Predictors of Student Achievement at Elementary Schools in LAUSD. Program Evaluation and Research Branch, Los Angeles Unified School District.

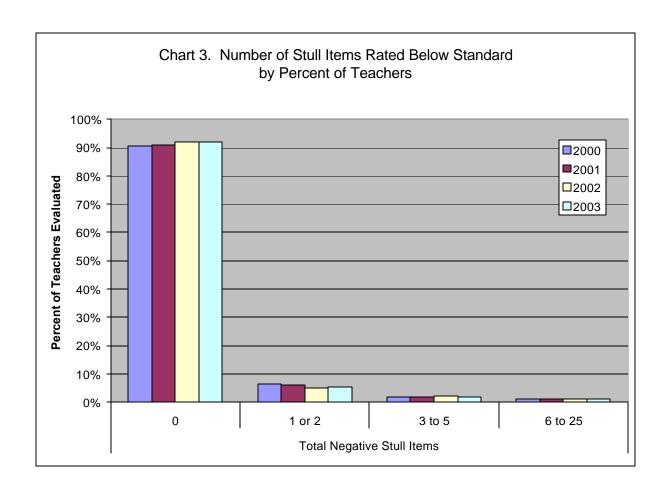


In addition to the Stull overall rating, teachers are rated on 25 items related to teaching practice and job behavior. Table 1 presents these items and summarizes the frequency with which the evaluated teachers were marked as "No" or "Needs Improvement" on each. Teachers were most frequently rated negatively on items 5a. Regularly arrives on time (mean = 2.8%), 3i Maintains appropriate discipline (2.3%), and 5c. Regularly in attendance for the prescribed class day (2.0%). Teachers were least frequently rated negatively on items 4a. Maintains appropriate professional relationships and communications with other staff members (0.2%), 3e. Provides independent practice (0.4%), and 3h. Assigns regular and appropriate homework (0.4%). Thus, the most common negatively rated items were in the areas of absenteeism and classroom discipline, and the least common were for how well teachers relate with other staff and the kinds of work they assign students. It is unclear, however, whether these frequencies reflect the extent to which teachers actually perform below standard. The frequently marked items can be documented with relative ease, while the infrequently marked items would be more difficult to document without a close familiarity with the teacher's practice.

Table 1. Percent of Teachers Rated Negatively for each Stull Item

STULL AREAS OF EVALUATION	Percent of Negative Responses				
	2000		-		Mean
1. ACHIEVEMENT OF INSTRUCTIONAL OBJECTIVES					
a. Adheres to curricular objectives	1.0%	0.8%	0.8%	0.8%	0.9%
b. Continuing indication of effective student progress and performance	1.2%	1.4%	1.2%	1.2%	1.3%
2. PREPARATION AND PLANNING					
a. Weekly and long-term planning evident and available	1.9%	1.4%	1.3%	1.5%	1.5%
b. Uses appropriate instructional materials	0.7%	0.8%	0.6%	0.6%	0.7%
c. Provides effective classroom environment	1.7%	2.0%	2.0%	2.0%	1.9%
3. CLASSROOM PERFORMANCE					
a. Has clear instructional objectives	1.5%	1.4%	1.2%	1.3%	1.4%
b. Motivation appropriate	1.0%	1.4%	1.1%	1.0%	1.1%
c. Directs lessons related to knowledge/skills	0.8%	0.8%	0.8%	0.8%	0.8%
d. Provides guided group practice	0.7%	0.8%	0.6%	0.7%	0.7%
e. Provides independent practice	0.5%	0.5%	0.3%	0.4%	0.4%
f. Provides for individual differences	1.3%	1.4%	1.1%	1.3%	1.3%
g. Evaluates student progress	0.6%	0.6%	0.5%	0.7%	0.6%
h. Assigns regular and appropriate homework					
i. Maintains appropriate discipline	2.3%	2.5%	2.1%	2.1%	2.3%
4. GENERAL PROFESSIONAL SKILLS					
 a. Maintains appropriate professional relationships and communications with: 					
Students	0.4%	0.5%	0.5%	0.4%	0.5%
Parents	0.5%	0.6%	0.8%	0.6%	0.6%
Other staff members	0.2%	0.1%	0.2%	0.1%	0.2%
b. Provides effective nonclassroom supervision	1.3%	1.2%	1.2%	1.4%	1.3%
c. Performs adjunct duties effectively	0.9%	1.2%	0.9%	0.9%	1.0%
d. Maintains professional appearance	0.6%	0.8%	0.6%	0.7%	0.7%
e. Maintains accurate and timely records					
5. PUNCTUALITY AND ATTENDANCE					
a. Regularly arrives on time	3.3%	2.8%	2.7%	2.2%	2.8%
b. Regularly starts classes on schedule	1.2%	1.0%	1.1%	0.9%	1.1%
c. Regularly in attendance for total prescribed contract day	2.0%	1.8%	2.2%	1.8%	2.0%

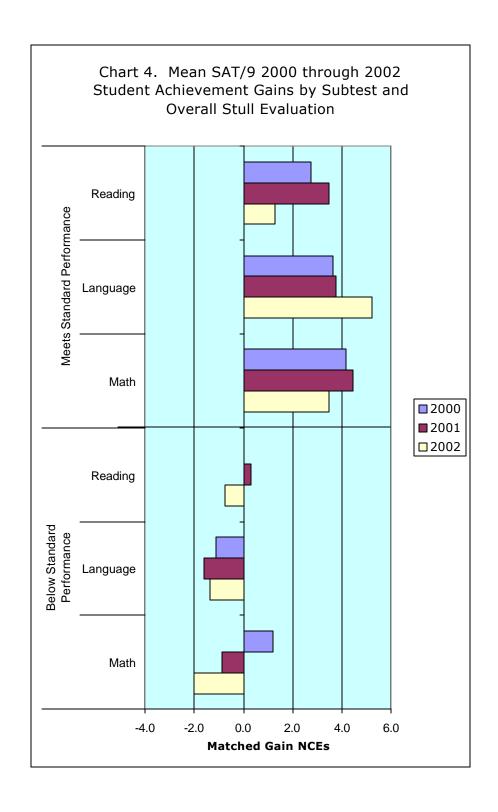
Examination of the results regarding the extent to which teachers receive negative ratings on areas of their teaching shows that not only are few teachers rated as *Below Standard Performance overall*, rarely do administrators identify any items negatively. As can be seen in Chart 3, during each of the past four years over 90% of teachers given Stull evaluations do not have any items rated negatively and only three percent of teachers were rated negatively on more than two items. Thus, for most teachers, the Stull ratings provide little or no formative information about areas for improvement.



What is the relationship between Stull ratings and student achievement outcomes?

Since improved student achievement is a paramount goal in the district, a teacher evaluation tool should be aligned such that teachers rated *Below Standard* would have lower classroom achievement gains than those rated as *Meets Standard*. In order to examine this issue, SAT/9 classroom level matched student gains in reading, math, and language from teachers rated as *Meets Standard Performance*. Because elementary school students are exposed to the same teachers most of the instructional day, student achievement gains are more attributable to teacher performance. For this reason, these comparisons are limited to elementary teachers only. The *Below Standard Performance* group for each year includes all teachers who were rated as *Below Standard Performance* in any of the four years of the study (2000-03). This larger sample of teachers rated *Below Standard* allowed for greater confidence in the robustness of the results, since a small sample can be unduly influenced by a few extreme cases.

Student gains of teachers rated *Meets Standard Performance* were much higher (see Chart 4) than those rated *Below Standard Performance* in each subject area in each of the three years. The same pattern was found when these analyses were replicated controlling for student prior performance and student and classroom characteristics. Thus, the analysis of student achievement gains corroborates the Stull ratings. Teachers with an overall evaluation of *Below Standard Performance* were less effective in raising student achievement.



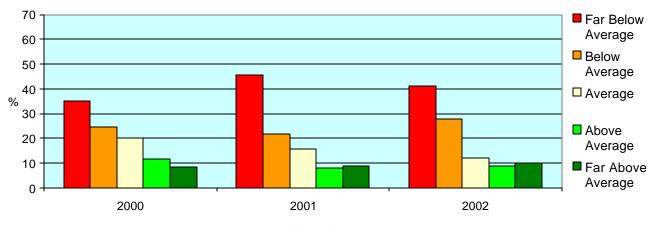
To more closely examine the relationship between overall Stull evaluation and student achievement, relative classroom performance was calculated by dividing SAT/9 mean classroom gains into five equal-sized groups (quintiles). Chart 5 shows the percentage of teachers rated *Below Standard* that fell within each group. Teachers rated as Below Standard performance most often exhibited the lowest relative classroom gains. It is noteworthy, however, that teachers rated *Below Standard* were found in all quintiles of student performance throughout all of the years. Low matched gains in math were the most consistently associated with *Below Standard* Stull ratings, while low matched gains in reading were the least associated with *Below Standard* ratings.

This analysis has assumed that Stull ratings and student achievement are both indicators of teacher quality, and has found that few teachers are rated *Below Standard* but that those teachers tend to have lower classroom gains. This suggests that for the most part, the Stull ratings accurately identify teachers that are performing poorly.

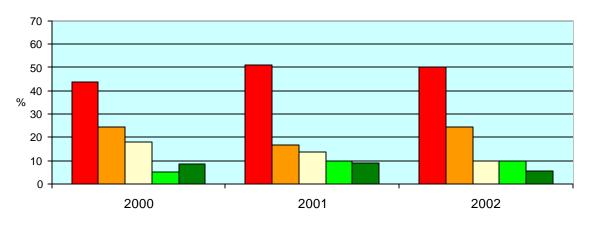
Conclusion

Through examining the relationship between teacher Stull evaluations and student achievement, some key findings have emerged. First, it does not appear that school administrators districtwide are systematically using the results of Stull evaluation forms to provide formative feedback to their teachers. This is suggested by the fact that very few teachers receive overall Stull evaluations of *Below Standard Performance* and very few teachers are rated negatively on individual Stull items. Second, overall Stull evaluation ratings of *Below Standard* tend to accurately identify ineffective teachers, however, the majority of ineffective teachers are probably misidentified as meeting standard performance.

Chart 5. Percentage of Teachers With Overall Stull Evaluation Rating Below Standard by SAT/9 Mean Classroom Gain (2000 to 2002)



Reading Gain Quintile



Language Gain Quintile

