

August 13, 2015

To: Parties

From: Bill Hawley

Thoughts on the Mendoza Request for an R&R on Teacher and Principal Evaluation

### Introduction

On July 30, Mendoza plaintiffs requested an R&R on issues related to the teacher and principal evaluation plans. This memo identifies some issues that I hope can be resolved without going to the Court. I am asking the District and the plaintiffs to consider my observations and proposals and indicate whether they agree or disagree. This is not a draft of the R&R, it is an attempt at resolution.

Assessing Academic Performance of Students for Purposes of Evaluating Teachers and Principals.

I think that the plans are less clear than they need to be relating to this issue but in my discussions with the District I was assured that all teachers will have the equivalent of a pre-and posttest measure of student performance. Tests have been developed by the District that will apply to grades 3-12 covering the material being taught. K-2 students' performance will be assessed using periodic DIBELS tests.

This clarification should obviate the need for an R&R dealing with academic growth measures.

### Teacher Evaluators

I have argued that teaching practices measured by the observational instrument should be assessed by persons other than or in addition to principals and

assistant principals. The Mendoza plaintiffs agree. Rather than ask the Court to resolve this matter at this time, I propose that a pilot study be conducted that will allow comparison of assessments of teaching practice by principals and assistant principals on the one hand and trained evaluators on the other. The Superintendent has indicated his openness to exploring this option. It should be possible to design such study within the next 3 to 4 weeks. An R&R or stipulation could include a provision that keeps the matter open the District decide not to undertake pilot once design is developed. The results of the pilot will not affect scores received by teachers from administrators' assessments this year.

### Cut Scores

The state requires that the District establish criteria for determining levels of teacher effectiveness. These "cut scores" in TUSD are suspect because only a handful teachers are judged to be ineffective. (This may be as much a problem with evaluation as it is with the cut scores themselves).

I am told that the cut scores were established by looking at research on the percentage teachers judged to be ineffective, namely 4-6%. If this conclusion is reports of how principals and assistant principals rate teachers, then we have to reckon with research that shows that principals and assistant principals rate teachers much higher than do expert evaluators. Even so, the USP cut scores came nowhere near identifying 4-6% teachers as ineffective. It may be that TUSD teachers are significantly more effective than teachers in other districts but given the difficulty that the District says it has recruiting and keeping good teachers, it seems plausible that the TUSD cut scores do not effectively differentiate teachers on the basis of professional proficiency.

Establishing cut scores is not easy and requires an analysis of various dimensions teacher performance. It does not seem feasible for the Court to establish what the cuts for should be. I propose that this matter be "resolved" by having the District commit to describing and justifying the bases on which it establishes cut scores that differentiate levels of teacher proficiency. Indeed, the District acknowledges that rethinking the cut scores is necessary.

### Alignment of Instruments for Measuring Teacher and Principal Effectiveness\*

Measurement of teacher effectiveness is inherently “high-inference”. In such cases, it is important to have multiple measures of the same phenomena. This means that teacher and student surveys, as well as observational measures of teacher and principal behaviors, should embody similar concepts. I fail to see how there could be any disagreement with this proposition. Indeed, while on the one hand the District says that it such an alignment effort would be burdensome and redundant, staff who developed the instruments say that they undertook such an alignment. I suggest that this issue could be resolved by the District developing a chart showing how important aspects of teaching and leadership are reflected in these instruments. This is not a difficult task and has the value of making clear to principals and teachers behaviors that are important for them to know about and be able to do.

### Linking Evaluations to Improvement

The district is sending me a report that addresses the specifics of the processes related to how evaluations are used to improve teacher performance. So, I will send you a suggestion on this matter as soon as I can.

### Training Evaluators

The efficacy of a measurement tool depends on the capabilities of the person who does the measurement. This is particularly true when one is assessing behaviors not easily defined. The Mendoza plaintiffs want the District to specify how it will prepare those who assess teachers and principals to undertake this evaluation. The response of the District is that the training takes many forms in many venues and that one could look at the professional development plans to determine what the District proposes to do. I have looked at the professional development plans and while there are numerous references to the training of educators with respect to effective teaching practices, this is not the same as the training of evaluators. The observational instruments being used in TUSD are

\*The District does not address this objection by the Mendoza plaintiffs in its response but does justify its position in an email dated June 9.

complex, extensive in the range of behaviors being assessed, and presumably discrete items are repeated in somewhat different terms throughout the instruments. I have supervised studies that involve the observation of educators. In preparation for those studies, we spend considerable time training the observers (evaluators) to ensure inter-rater reliability. If the District is conducting such training, which its staff says that it is, it would seem useful to spell out just how and when that will happen. This would allow the Implementation Committee to monitor this activity. There is no need for the Court to order such a description of how evaluators will be prepared for this difficult task assuming the District is willing to do so. Surely it has a plan; why not make it more transparent.

#### Assessing the Capabilities of Teachers and Administrators to Use Data on Student Outcomes.

The Mendoza plaintiffs claim rightly that the USP is specific about the measurement of teachers' and administrators' capacity to utilize data to improve student performance. Moreover this is a high priority of the Superintendent. In response to the Mendoza this concern, the District says that the structure of the evaluation as prescribed by the state is an impediment and that this skill is covered by student surveys. This assertion apparently misunderstands what is involved. Assessing the capabilities of educators to utilize data on student behavior and achievement, among other things, is an appropriate component of the observational instruments. Indeed, those instruments do include relevant rubrics. Why the District would not identify them is beyond me. In any case, there is no need to take this issue to the Court.

#### The Weight of Teacher and Student Surveys and Principal Evaluation

The Mendoza plaintiffs point out that only 10 of the 100 points on the principal evaluation score are derived from the combination of the teacher and principal surveys. No doubt this reflects the wishes of principals. (At one point in the process they proposed giving one percent weight to the perceptions of students). I wonder how teachers feel about having student surveys account for 10% of their evaluation but only 4% of evaluation of principals. And I wonder too, whether teachers believe that their judgements about principal behaviors and school

conditions should be given a weight that District staff identifies correctly as negligible. In short, the principal evaluation plan says that the views of teachers and students don't count.

The instrument to be used to account for more than half of a principal's evaluation score includes numerous items that can be better if not only assessed by teachers and students. This is true for several domains in the principal instrument, especially with respect to those items referred to as "School Behaviors". Many of these items require the evaluator to determine what teachers and students believe and experience. How might they do that? Interviewing a few teachers or students in each school is hardly fair to the principals themselves much less a valid way of determining reality. A great deal of emphasis in virtually all school improvement efforts, and certainly in the USP, is placed on the importance of creating (1) school cultures that are inclusive, respectful, supportive, and reflect high expectations and (2) fostering teacher collaboration, supporting teachers' professional growth, retaining effective teachers, and developing a sense of physical and psychological safety that enhances teaching and student learning. What better way to measure whether principals have accomplished these things than by asking teachers and students.

State guidelines place a constraint on the points that can be assigned to teachers and principal and teacher and student surveys but there's no reason not to use all of those 17 points. So, I propose that teacher surveys account for 11 points and student surveys account for six. Or 12 for teachers and five for students. I am prepared to ask the Court for such a determination and I believe the case for this is strong.

### Concluding Comment

I may be too optimistic but I believe that all of the issues discussed in this memo can be resolved without going to the Court except, perhaps, the issue of survey weights for principal evaluation.