

Oregon's Matrix for evaluating teachers and administrators is a model that is unacceptable because it fuses goals and ratings with plans. The model's most serious flaw is that ratings and plans are not strictly referenced to student performance.

Oregon's Matrix, The Emperor's Latest Clothes
by
Siegfried Engelmann

Oregon's recently adopted matrix is an illusion on the order of the emperor's new clothes, which were visible only to those who were "perceptive." The Oregon Matrix purports to present a scheme for districts and schools to create summative evaluations of teachers and administrators. A summative evaluation is usually conceived of as something like an overall score, 88% or "well above average."

A reasonable way to fashion a "summative" evaluation would be to obtain relevant data on student performance, and then judge the teachers and administrators almost exclusively on the basis of this data. If the student performance is poor, the teachers, administrators, and district are provisionally rated D or F.

The most relevant data the district could use would be how students perform *on all parts of the instructional programs the district adopted*. Note that this is virtually never done. Yet the adopted materials are the central tools designed to cause students to learn relevant content. Poor performance on the adopted material should therefore have substantial implications for teachers and administrators.

Student failure to learn the content of the adopted programs would imply changes in how students are grouped for instruction and how they are taught. No sensible person would put a student who performed on the third-grade level in material appropriate for students who read at the 7th grade level. Yet, this is done routinely in inner-city schools, where “average” 7th graders perform on the third-grade level.

If student performance is close to or above the target score no changes are implied; however in schools that have more than about 25% of the classes in trouble, the principal and failing teachers would be placed on a plan of assistance. The assistance would be provided by people who have demonstrated skill in accelerating the performance of failed students and teachers.

Compare this overall orientation with Oregon’s Matrix model for the summative evaluation of teachers and school administrators. The following is the state’s introduction to the Matrix. It may leave you with the impression that if you knew a little more or studied it a little bit more thoroughly, you would understand it. Not so. This plan lacks logical cohesion.

Oregon’s Matrix Model for Educator Summative Evaluations is governed by Oregon’s Framework for Teacher and Administrator Evaluation and Support Systems: The matrix model is based on five requirements specified in the Framework.

1. **Standards of Professional Practice.** The state adopted Model Core Teaching Standards and Educational Leadership/Administrator Standards define what teachers and administrators should know and be able to do to ensure that every student is ready for college, careers, and engaged citizenship in today’s world.
2. **Differentiated (4) Performance Levels.** Districts select a rubric to evaluate teacher and administrator performance on the standards of professional practice measured on four performance levels. Each level is defined as follows: Level 1=does not meet standards; Level 2 = progress toward meeting standards; Level 3=meets standards; Level 4=exceeds standards.
3. **Multiple Measures.** Multiple sources of data are used to measure teacher and administrator performance on the Standards of Professional Practice, including evidence from: professional practice, professional responsibilities, and student learning and growth.

4. Evaluation and Professional Growth Cycle. Teachers and administrators are evaluated on a regular cycle of continuous improvement, which includes self-reflection, goal setting, observations, formative assessment, and summative evaluation. The Oregon Matrix Model is used for the summative evaluation. The matrix model combines measures for professional practice (PP) and professional responsibilities (PR) and student learning and growth (SLG). The Y-axis represents the performance level for PP/PR, and the X-axis represents the performance level for SLG. The educator's Professional Growth Plan and overall summative performance level are determined by the intersection of the Y- and X-axes.
5. Aligned Professional Learning. Relevant professional learning opportunities to improve professional practice and impact on student learning are aligned to the teacher's or administrator's evaluation and his/her need for professional growth.

In the Oregon Matrix, Professional Practice (PP) and Professional Responsibilities (PR) intersects with Student Learning and Growth (SLG) culminating in a Professional Growth Plan and summative performance level. When there is a discrepancy between the PP/PR level and SLG level, further inquiry is triggered to explore and understand the reasons for the discrepancy.

Here is the Oregon Matrix:

Y-AXIS: Combined Rating on Professional Practice and Professional Responsibilities (PP/PR)	LEVEL 4 (Highest)	COLLEGIAL PLAN With focus on SLG Goals *SLG INQUIRY due to LOW level of fidelity between measures 3	FACILITATIVE or COLLEGIAL PLAN With focus on SLG Goals Determined post inquiry *SLG INQUIRY due to only SOME level of fidelity between measures 3 or 4	FACILITATIVE PLAN Educator leads development of Professional Growth Plan GOOD level of fidelity between measures 4	FACILITATIVE PLAN Educator leads development of Professional Growth Plan HIGHEST level of fidelity between measures 4
	LEVEL 3	COLLEGIAL or CONSULTING PLAN With focus on SLG Goals Determined post inquiry *SLG INQUIRY due to SOME level of fidelity between measures 2 or 3	COLLEGIAL PLAN With focus on SLG Goals GOOD level of fidelity between measures 3	COLLEGIAL PLAN Educator and evaluator collaboratively develop Professional Growth Plan HIGHEST level of fidelity between measures 3	COLLEGIAL PLAN Educator and evaluator collaboratively develop Professional Growth Plan GOOD level of fidelity between measures 3
	LEVEL 2	CONSULTING PLAN With focus on SLG Goals Evaluator consults with the educator and guides development of Professional Growth Plan GOOD level of fidelity between measures 2	CONSULTING PLAN With focus on SLG Goals Evaluator consults with the educator and guides development of Professional Growth Plan HIGHEST level of fidelity between measures 2	CONSULTING PLAN Evaluator consults with the educator and guides development of Professional Growth Plan GOOD level of fidelity between measures 2	COLLEGIAL or CONSULTING PLAN Determined post inquiry *PP/PR INQUIRY due to only SOME level of fidelity between measures 2 or 3
	LEVEL 1 (Lowest)	DIRECTED PLAN With focus on SLG Goals Evaluator determines Professional Growth Plan HIGHEST level of fidelity between measures 1	DIRECTED PLAN With focus on SLG Goals Evaluator determines Professional Growth Plan GOOD level of fidelity between measures 1	CONSULTING or DIRECTED PLAN Determined post inquiry *PP/PR INQUIRY due to only SOME level of fidelity between measures 1 or 2	CONSULTING PLAN Evaluator consults with the educator and guides development of Professional Growth Plan *PP/PR INQUIRY due to only LOW level of fidelity between measures 2
		LEVEL 1 (Lowest)	LEVEL 2	LEVEL 3	LEVEL 4 (Highest)
	X-AXIS: Rating on Student Learning and Growth				

*Ratings in these areas require an inquiry process in order to determine a summative performance level and Professional Growth Plan

The lower left cell is the lowest in both student performance (the axis that goes to the right) and in “professional” ranking (the axis that goes up). The “preferred” route of improvement goes from the lower-left box on the diagonal to

the upper-right box, which shows the highest student and highest professional rankings. The authors of this matrix assume that the levels for professional practices and those for student learning should be parallel. They indicate, “ When there is a discrepancy between the PP/PR level and SLG level, further inquiry is triggered to explore and understand the reasons for the discrepancy.”

This is a bizarre orientation. The measures of student performance and professional responsibilities are independent of each other, which means that there can be low-grade professionals in a school that has a very high student population. At the other extreme, the student population may be extremely low and the rate of turnover is high, but the school may have very good teachers and principal.

The comment in each cell on the diagonal route from lower left to upper right indicates, “Highest level of fidelity between measures.” All this means is that the numbers for both axes are matched: both levels are 1 or are 2 or are 3 or are 4. Note: that is all this comment means. It carries no implications that this is a superior combination, or even that it is a desirable combination. It is nothing more than a property of any matrix.

The other wording in the cells tells who is developing plans for future growth. For the bottom left cell, the “evaluator” makes the plan. For the intermediate cells “evaluator” and “educator” make the plan. For the upper-right cell, the “educator” makes the plan. In an effective system, the only ones who would be making plans would be those who have a record of making successful plans.

Possibly, the most deceptive practice refers to continuous improvement:

Teachers and administrators are evaluated on a regular cycle of continuous improvement which includes self-reflection, goal setting, observations, formative assessment, and summative evaluation.

There is no evidence of such evaluation leading to improvement and no evidence that self-reflection and goal setting lead to improved student or teacher performance.

Also it would be very interesting to see reliability data on how consistently teachers, administrators, educators and evaluators make the same judgments

about the specific cell that is identified for a specific plan and specific student data. I would make a substantial wager that reliability data would be so low that the model would have to be scrapped.

Let's apply two tests to the matrix.

Test 1. Can you identify a concrete example of what is being described? If you can't, what is being described could be some form of illusion.

Test 2. How directly is the matrix linked to the problem it is supposed to solve? If there is a long sausage chain between what is being proposed and the problem it is supposed to address, the rhetoric serves only as a disguise.

Here are four cells from the middle of the matrix:

<p>COLLEGIAL PLAN With focus on SLG Goals</p> <p>GOOD level of fidelity between measures</p> <p style="text-align: right;">3</p>	<p>COLLEGIAL PLAN Educator and evaluator collaboratively develop Professional Growth Plan</p> <p>HIGHEST level of fidelity between measures</p> <p style="text-align: right;">3</p>
<p>CONSULTING PLAN With focus on SLG Goals</p> <p>Evaluator consults with the educator and guides development of Professional Growth Plan</p> <p>HIGHEST level of fidelity between measures</p> <p style="text-align: right;">2</p>	<p>CONSULTING PLAN Evaluator consults with the educator and guides development of Professional Growth Plan</p> <p>GOOD level of fidelity between measures</p> <p style="text-align: right;">2</p>

Test 1. Can you identify a concrete example of what is being described by each cell?

I can't. If somebody presented me with a plan for one of the four cells above, I doubt very much that I would be able to identify the "correct" cell.

Test 2: How directly is what is being described linked to the problem it is supposed to solve?

The matrix shows the “summative evaluation” scale. The summative evaluation is achieved simply by assigning a school administrator or teacher to one of the 16 cells in the matrix. The cells, however, focus on other issues. They refer to “professional growth plans” and identify who is responsible for developing these plans. The plans—labeled directed, consulting, collegial, and facilitative—seem more important than what should be the central focus of the matrix, to provide a summative score that has clear implications for how the school or district must change. The matrix doesn’t provide direct links between problems of low summative scores and their solutions. All we know is that there will be a plan that has no particular parameters or relation to the problem, and this plan is deemed acceptable according to unstated criteria.

An article on the matrix put out by Oregon Education Association declares, “The matrix model is a far cry from percentage models adopted by other states that put heavy emphasis . . . on student growth. The Feds have favored a percentage formula....” Indeed, that’s where the emphasis should be. To make the Oregon criteria even more elastic, the OEA article indicates that personnel decisions don’t “have to hinge exclusively on the final summative performance level.”

Another document, *The Oregon Framework (SB290/ESEA Waiver): Lessons learned from Pilot Districts and support for Statewide Implementation*, presents a specific formula for weighting the components that are evaluated:

Component	Description of Component	Rubric Rating
(A) Professional Practice Standards 1-8 60 %	Evidence of teacher's proficiency re: Learner Development, Learning Differences & Environments Content Knowledge, Application of Content, Assessment, Planning for Instruction, Instructional Strategies 60 % x rating (3) =	1.8
(B) Professional Responsibilities Standards 9-10 20 %	Evidence of teacher's progress toward their own professional goals, contribution to school improvement goals/plan and collegial learning. 20 % x rating (4) =	.8
(C) Student Learning & Growth 20 %	Evidence of teachers' impact on student learning and growth through goal setting, planning, assessment, and instructional strategies 20 % x rating (3) =	.6
(D) Summative Rating	Sum of A + B + C =	3.2

These weightings are self-contradictory. The rubric rating for student learning is less than 20% of the total score. In contrast, the score for Professional practice standards and responsibility account for 80% of the total score. (Note that both professional responsibilities and student learning are 20% in the left column, but in the right column professional responsibilities are .8 of the summative rating and student learning is only .6.)

The contradiction is that it is hard to imagine how a person being evaluated could have a high score unless there was substantial evidence that student performance was good. Student performance would therefore have to be the most heavily weighted component.

Consider the reaction there would be if auto repair shops adopted a scheme parallel to this rubric. The issue of whether the car is properly repaired is weighted 1/5 the total score, while the professional skills of those who work on cars are 4/5 of the total. If more than half of the jobs are total failures, the mechanics could still receive very high scores.

This plan is obviously unacceptable for mechanics and should be as obviously unacceptable for teachers and administrators who have the job of fixing the kids so they learn what they are scheduled to learn. If the kids aren't learning, the teachers aren't teaching and the supervisors are not supervising.

As it stands now, the teacher's knowledge of learner development, learning differences, planning for instruction and instructional strategies are judged without serious reference to student performance. That orientation is not only unacceptable; it flies in the face of the central purpose of schools' and Oregon's recent commitment to student achievement.

The ambitious excerpt below is from Oregon Framework for Teacher and Administrator Evaluation and Support Systems, September 2013:

The ultimate goal of strengthening teacher and leader evaluation systems in Oregon is to ensure equitable outcomes where all students, regardless of background, are ready for college, careers, and engaged citizenship by ensuring the following outcomes:

- • Improved student learning at all schools and for all students
- • Effective teachers in every classroom
- • Effective leaders in every school and district
- • Reducing achievement gaps between the highest and lowest performing student groups, while increasing achievement and success for every student
- • Continuous professional growth for teachers and leaders throughout their careers

You don't reduce achievement gaps and increase achievement for every student without having an unwavering primary focus on student achievement. The Oregon Matrix does not have this focus. Possibly more to the point, the Oregon Framework observed that "*for final approval, Oregon's state guidelines must include how student learning and growth will be included as a significant factor in educator summative evaluations.*" Student growth is a factor in the Matrix score, but arguably not a significant one.

A final demerit: The job facing those who fashion a legislated plan is made very difficult by the unbridled requirements that appear in documents that "govern" or greatly influence the new plan. These documents are laced with undisciplined wording, assertions, and requirements, such as using multiple

measures, a four-level scale, and the other details that must be included in the model the state develops. Despite these distractions, the documents specify foundational requirements that have been egregiously ignored by those who fashioned Oregon's Matrix.

Summary

The Oregon Matrix is an insult to the actual process a district necessarily must go through to clean up the unbelievably inadequate instruction that occurs, particularly in schools that have the greatest need for reform. The Matrix provides a vague and elastic standard that is largely independent of facts about the actual performance of students or even honest acceptance of the notion that the performance of the students is the most relevant variable needed to evaluate the performance of teachers and administrators. We don't need teachers who are incapable of teaching the population of students that are in schools, or administrators who spend less than 10% of their time in classrooms.

We are sympathetic with teachers and administrators, but we would not take our car to a shop that wasn't concerned with whether repairs were successful or take our children to a surgeon who didn't consider success of surgical outcomes very important. Therefore, we should emphatically reject models and rhetoric that don't consider student performance the primary reference point for evaluating personnel. We don't need a system that uses smoke, mirrors, and emperor robes, but one in which those who teach and administrator teachers embrace the central goal of greatly improving student performance.