

THE DEBATE OVER TEACHER EVALUATION – ARE WE ASKING THE RIGHT QUESTION?

Timothy B. Baird Ed.D.

POSING THE RIGHT QUESTION

The debate over how to evaluate teachers is longstanding and complex. Mired in issues over forms, scores, types of data used, and countless other details this one focus area has recently become the single biggest element of many “school improvement programs.” In today’s political climate, Republicans and Democrats are both looking at “value added” teacher evaluation systems that are heavily based on student performance on standardized test scores and systems that reward high test scores with additional pay and low test scores with punitive actions leading to dismissal. The premise is that if teachers are evaluated based on test scores, they will work harder to make sure that students do better on the tests. Principals will work with teachers to improve the weak areas found in their test data and everyone will be better off. The real perceived payoff in this system, though, is that this will finally shine the spotlight on ineffective teachers and give school districts a tool to remove them. In the political arena, teacher evaluation is the Holy Grail that will lead to improved educational systems across the nation. This seemingly simple solution may not be the panacea that it is heralded to be.

This is not to imply that there are not bad teachers in the workforce and that something shouldn’t be done about them. Almost every school administrator would agree with this assessment. An argument could be made that every profession has its lowest achievers and the world would be better off if we removed the lowest performing 3% of our firefighters, police officers, mechanics, judges, and teachers from the work force. However, when critics have identified this as the single most important factor of school reform, they are overstating the problem. Removing bad teachers should be part of the conversation but really just a small part. The problem with focusing so much attention on such a small part of the complex issues surrounding improving our educational system seems counter-productive and designed to take our focus off of the larger problems our school systems face.

So if it’s not just about eliminating bad teachers, doesn’t a value added system still work because it forces teachers into working harder to achieve results? Not really. The vast majority of teachers are already working hard to help students achieve. To assume that most teachers are operating at less than 100% and they need to be threatened or rewarded to get the maximum out of them is insulting to most educators. To improve the quality of teachers, shouldn’t we be asking how do we help teachers develop their knowledge and skills of effective teaching rather than what is the best way to measure whether this is happening? Proponents of tougher evaluation systems linked to student data will make the argument that what gets measured gets taught. And that is the basic weakness of these evaluation system

reforms. They tend to narrow the curriculum to the most basic elements of the testing system. If the assessment tool is heavily multiple choice, then I would not expect to see much writing in our children's curriculum. This focus on the assessment tool and not the broader scope of the knowledge and skills to be taught is the natural outgrowth of this heavy reliance on statewide assessment data linked to teacher evaluation. The other major flaw with tying teacher evaluation to student testing data is that there is no evidence in the research on this practice that it works. Linda Darling Hammond in a recent interview takes on the practice of using value added for two major reasons:

She states, "First, test-score gains-even using very fancy value-added models – reflect much more than an individual teacher's effort, including students' health, home life, and school attendance, and schools' class sizes, curriculum materials, and administrative supports, as well as the influence of other teachers, tutors, and specialists. These factors differ widely in rich and poor schools.

Second, teachers' ratings are highly unstable: They differ substantially across classes, tests, and years. (Education Week, Hammond, March 5, 2012)

So if how do we measure effective teaching isn't the right question, then what is? How about, how do we help teachers develop their instructional skills to support student learning? Analysis of data across many different districts, show us that most teachers are deemed to be effective at what they do and they are continuously working on improving teaching skills. How about if the concept of teacher development was prioritized and separated from the task of evaluating teachers? Separating these two distinct systems that have always been linked together, may actually make both systems more effective.

Applying Motivational Research to Teacher Evaluation and Reward Systems

To address this issue, we must go back to the underlying assumptions of the Teach-Test-Reward or Punish thinking. To assume that this system works, one must first throw out years of research on effective motivation strategies. As highlighted in his book DRIVE, Daniel Pink exposes the facts behind extrinsic motivation failing to produce results when the task is anything beyond basic rote skills. Indeed, this same motivation research shows that excessive rewards or punishment actually diminishes results when the task at hand is a higher cognitive task such as teaching. So when applied to the real world complex skill of improving instruction, motivational strategies that rely heavily on gains such as merit pay or punishments such as dismissal may actually be working against the goal of improving teacher performance.

There are other inherent flaws in using extrinsic motivation to improve teacher performance. I take particular offense at the concept of merit pay. The assumption is that teachers are not giving 100% of their efforts toward improving student

outcomes. If we only rewarded them with a little more pay, they will try harder and get better results. In my anecdotal evidence gathering on this issue, I have found that most teachers usually commit high effort levels to help students achieve success. Teaching by its very nature is a purpose driven profession. Those seeking big paychecks usually did not look to a teaching career to meet this objective. To assume that a small bonus is all that's missing from the complicated equation of increasing student performance is an insult to teachers and completely disregards years of motivational research related to the issue.

So let's assume that teachers are intrinsically motivated to help their students achieve success. Why does our current system of staff development and teacher evaluation not work as effectively as it could? There are a number of reasons but I'm going to focus on two.

The first has to do with the disconnect of typical staff development within a district and school and the needs of any individual teacher. District staff development plans tend to be written from a one size fits all perspective. Although many districts and schools have done increasingly better jobs at differentiating staff development to individual needs, resources and scope still tend to limit this effort. Modern technologies are helping to make this differentiation more possible and we may see more and more refinement of connecting development resources to specific teacher needs in the coming years. But that alone still doesn't solve the problem.

The Battle Between Formative and Summative Feedback Systems

The real problem actually lies within the evaluation process itself. Our current system has the teacher evaluation cycle start with goal setting by the teacher. Throughout the course of the year (depending upon the skill of the evaluator) much work will be done around these goals. Teachers will be observed teaching a lesson and then will receive administrative feedback. In more progressive systems, student work will even be analyzed. Ultimately, this will lead to a summative evaluation where feedback will be given to the teacher on goal achievement and they will finally receive a rating identifying them as meeting standards or not. Herein lies the problem. Any system that leads to a summative evaluation inherently weakens the formative process leading to that evaluation. For formative feedback to be most effective, it must exist in a climate of openness and trust. Teachers working with the principal must have complete confidence that acknowledging a weak area that needs additional work will not lead to an unfavorable summative evaluation.

Evaluation practices are also weakened by first working on formative goals and then moving the focus to established standards of practice at the very end of the process. If teachers are being held to specific standards of practice, these standards should be evident, clear, and well articulated. Not mastering a self-imposed high-level goal such as improving critical thinking skills is not the same as not having acceptable

standards of basic classroom control in place. The mixed purpose of the current process often muddles this basic distinction between teacher development and teacher assessment of meeting district standards.

Teachers often have trouble with this same issue of formative feedback leading to summative assessment in their classroom. For many, the concept of “Every assignment counts,” is the foundation of the grading system. Certainly, every student needs to understand where they went astray on the assignment (formative feedback) but that doesn’t change the fact that in most classrooms today, the failing grade went into the grade book and will be averaged with the rest of the grades when the semester scores are computed. Whether the student can demonstrate mastery of that exact same concept at the end of the grading period is irrelevant in this grading universe. Every action matters, every day. In this system, formative feedback becomes smaller stand-alone pieces of the larger summative package. If this is the basic philosophy of most teachers related to purpose of formative feedback as a component of summative feedback, then is it any wonder that few teachers are comfortable sharing weak areas with their evaluator as part of a “formative process?”

Contrast this philosophy with a complete different worldview – that of the video gamer. Video games can be intricately elaborate in their structure and composition. Gamers can take weeks and months working through one game’s universe. Formative feedback in this worldview is helping a partner find specific clues or teaching them how to get past a particularly nasty adversary. Summative assessment in this universe is simply determining what level you are on at the end of the assessment period. Here, formative feedback may help you achieve a higher level of summative success but the amount and quality of your formative feedback does not count for or against you in the overall summative score. In fact, there is no recording required whatsoever of the formative feedback since it doesn’t matter in the final summation. In this system formative and summative feedback are kept as two distinct systems with both being used to fulfill different purposes. So what would this system look like in a teacher development / evaluation model?

The Teacher Development Model

The teacher standards that exist in the development model should be the same that we are looking for in the evaluation model. These standards should reflect the knowledge, skills, actions, and outcomes that the district expects of all teachers. The difference lies in the purpose of each tool. The teacher development model assumes that teachers are at their own unique level in these various standards. The role of this process is to work with the teacher or a group of teachers working on the same outcomes and give them formative feedback around mutually agreed upon goals in the different areas to help them move up the continuum. At the end of the development cycle, there is teacher self-reflection and discussion but it is not concluded with a summative evaluation. The end goal here is to improve teacher

awareness, understanding, instructional skills, and student outcomes. Here is one sample set of teacher standards that could be used for this model:

TEACHER STANDARDS OF PRACTICE

1. Pre-Teaching

- A. Teacher as life-long learner – Focus on continuous learning and development of the teacher’s knowledge and instructional skills.
- B. Teacher as shaper of learning environment, climate, and culture – Focus on developing knowledge and skills to inspire, develop, and ensure learning for all students.
- C. Organization of learning – Focus on providing learning in an organized and comprehensible structure and format for all learners.

2. Instructional Strategies

- A. The teacher is able to use the appropriate instructional strategy from a variety of strategies that is individualized to the specific learning context, environment, and individual learner needs.
- B. The teacher is able to interpret student feedback and change instructional strategies based upon that feedback as needed.

3. Assessment of Learner Mastery

- A. The learner demonstrates mastery of taught standards through a clearly articulated assessment process.

TEACHER DEVELOPMENT PROCESS

1. Pre-teaching - The teacher and coach can determine specific focus areas for reflection. Data gathering and feedback tools may include portfolios, staff development documentation, teacher observation, lesson plans, and other data as supplied.

2. Instructional Strategies – The teacher can receive specific feedback from the coach and other teachers on their instruction as well as visit other classrooms to observe effective strategies in action.

3. Assessment of Learner Mastery - The teacher and the coach will agree on what data should be used to reflect program success. This data should then be periodically reviewed to measure impact of curriculum and instructional strategies being used.

4. Teacher as Researcher – In addition to the strategies listed above, completing educational research should be conducted on a regular basis as one unique strand of teacher development that takes place outside of the context given above for individual development. During an action research cycle, the teacher should

identify effective teaching/learning practices from recognized educational research that could apply to their instructional setting. The goal of this type of action research is to learn new ideas and then develop an implementation plan for study and deployment. An alternative version of action research is to review key learning components of the current instructional plan and then develop a methodology to measure this activity for effectiveness. These two versions of “outside,” and “inside,” action research can be done individually but are usually more effective when done with a learning team.

WRAPPING UP THE TEACHER DEVELOPMENT CYCLE

The teacher feedback system concludes with the teacher conducting a self-reflection process and sharing this with his or her instructional coach. The teacher determines what next steps will look like and there are no summative assessments made. The only paperwork completed and kept simply shows that the teacher underwent a teacher development program during the school year. The district does not retain the goals, progress monitoring notes, nor the final reflection in this process.

The conclusion of the Teacher as Researcher Cycle ends with a review of findings shared with the staff of the school. This cycle will often result in changes to the instructional program of the teacher, a group of teachers, a school, or a district.

TEACHER EVALUATION PROCESS

At specific intervals, the district does need to evaluate whether individual teachers are meeting specific district standards at the required level. Although this process should include formative feedback throughout the process, the main goal of this cycle is to determine whether the teacher meets the district established standards. All newer teachers, teachers who have received negative evaluations, or any teacher that a principal has concerns about should be in an evaluation cycle until they have met standards or been removed from their position. In this cycle, standards for expected performance should be clearly defined and articulated with exemplar models of performance available for review. Again, I would use the same teacher standards as listed above only this time; I would be looking for expected levels of competence on each standard. Feedback on whether the expected level of competence was mastered would be given throughout the process but ultimately; the teacher would need to demonstrate mastery. Failure to do so would result in intensive coaching leading to dismissal if the required improvements were not attained.

The Five-Year Teacher Development / Evaluation Cycle

So if we put all the pieces together, what does this model look like in practice? Here is a five-year sample plan for an average teacher who has met expectations in the past:

Year One - Teacher Development Cycle

Year Two – Teacher as Researcher Cycle

Year Three – Teacher Development Cycle

Year Four – Evaluation Cycle

Year Five - Off Cycle

Of course, this plan would look different for a new teacher, a teacher coming off a negative evaluation cycle, or any teacher that the principal has concerns may not be meeting district standards in any area.

Conclusion

Teacher evaluation is the hot topic on the national circuit right now. A great deal of effort, time, money, and debate is being utilized to determine how to use the cumbersome tool of teacher evaluation systems to improve teacher performance. I believe that it's time to decouple these concepts of development and evaluation. This will allow districts to spend most their resources working with teachers in a truly formative process to improve teaching practices and student outcomes. This single change will target the appropriate supports to the vast majority of teachers who are continuous learners and want their students to be successful. This will also allow schools to finally use evaluation as a tool to truly measure teacher effectiveness and give districts a clear path for removing those teachers who aren't measuring up to standard.