

Teacher Development and Evaluation:  
A Study of Policy and Practice in Colorado

Al Ramirez,

Mike Lamphere,

Jim Smith,

Shelmon Brown,

Jennifer Pierceall-Herman

University of Colorado, Colorado Springs

A paper presented at the annual meeting of the

American Educational Research Association

Denver, Colorado May 2010

### Abstract

This study undertook an investigation of policy and practice related to teacher evaluation and development among Colorado school districts. A conceptual analysis was applied to a sample of teacher evaluation policies and process materials from 30 school districts. The study was based on policy objectives stipulated in state statute and best practice as reported in the literature. Results of the study show little relationship to teacher development and an emphasis in the policies and processes toward summative evaluations. Student outcomes received modest attention while the processes of teacher evaluation seemed de-contextualized to other school district systems.

*Keywords:* teacher evaluation, conceptual analysis, education policy

## Teacher Development and Evaluation: A Study of Policy and Practice in Colorado

### **Introduction**

This study explores the relationship between espoused policy and practice in regards to teacher evaluation and teacher development among Colorado school districts. At the core of the study are the beliefs that: 1) effective teachers are essential to student learning; 2) novice teachers are not fully formed professionals and need concentrated and sustained growth opportunities; 3) teachers develop professionally at different rates over the arc of their careers; 4) teachers have individual professional development needs; 5) teacher evaluation policies must fulfill goals related to accountability and teacher development; and, 6) good practice in teacher evaluation will build stronger school faculties. The study tests these beliefs through a process of conceptual analysis of state and school district policy and practice among a representative sample of school districts in Colorado.

Good instruction and school improvement are inexorably linked. The notion that instruction can be measured and improved is at the heart of many school reform initiatives. Effective teacher evaluations are viewed as a vehicle for driving school improvement and central to many school reform initiatives aimed at increasing student learning in low performing schools (Rowan, Correnti, Miller & Camburn, 2009). Yet, the idea of linking successful teaching, student academic achievement and teacher evaluation is hardly new. Scriven (1987) raised the issue of validity in teacher evaluation decades ago while challenging the assertion that research

based effective teaching techniques could be incorporated into teacher evaluation systems and used as a basis for judging performance. In essence, Scriven focused on the construct validity of such evaluation systems. We raise similar questions about the evaluation system in Colorado from the perspective of policy and practice.

### **Theoretical Framework**

Teachers matter and good teachers make a significant contribution to student learning. Researchers continue to discover this fact as they attempt to explain those factors that contribute to or detract from student learning in pre-collegiate education (Sanders and Rivers, 1996; Rockoff, 2004; Rivkin, Hanusheck & Kain, 2005). Throughout its history from the genesis and advancement of the public school systems in the various states in America issues of teacher development, supervision and evaluation have remained a continuing concern (Kaestle, 1983).

Today, we see many of the same concerns expressed by contemporary researchers. The Center for American Progress published two recent reports related to teacher effectiveness, evaluation and teacher development that underscore the continuing search for keys to good teaching and learning (Donaldson & Peske, 2010; Chait & Miller, 2010). A national study (Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D., 2009) concluded that most teacher evaluation systems lack meaning for participants in the process, have little to no impact, and tend toward ritual as opposed to substance.

In Colorado the state legislature has promulgated an expansive policy with regard to teacher evaluation and development, Colorado Revised Statutes 22-9-101 through 109. The

explicit language of the statute details in over 1500 words policy and procedures regarding the evaluation of licensed personnel, especially teachers. In part the statute states:

- (d) The purposes of the evaluation, which shall include but need not be limited to:
  - (I) Providing a basis for the improvement of instruction;
  - (II) Enhancing the implementation of programs of curriculum;
  - (III) Providing the measurement of satisfactory performance for individual licensed personnel and serving as documentation for an unsatisfactory performance dismissal proceeding under article 63 of this title;
  - (IV) Serving as a measurement of the professional growth and development of licensed personnel; and
  - (V) Measuring the level of performance of all licensed personnel within the school district or employed by a board of cooperative services...

(CRS 22-9-106, subsection d.)

The policy goals of the statute are consistent with the research literature on effective teaching, professional development, effective schools, teacher evaluation and teacher development (Glickman, Gordon & Ross-Gordon, 2005).

## Methods

### Sample

The study examines a cross section of the 178 school districts in Colorado. The sample is representative of urban, suburban and rural areas, and a geographic distribution within the state. A total of 30 cases were scrutinized, consisting of seven urban school districts, eight suburban school districts, and 15 rural school districts. Within this purposeful sample are represented a plurality of Colorado students and teachers.

### Variables

The study looked at 13 variables that are organized into four domains. The domains group the variables under: 1) Accountability policies and practices; 2) Teacher developmental policies and practices; 3) Student outcome consideration; 4) Contextualization of the policy and practice within larger school district goals. Additional variables associated with school district characteristics and students success are also explored within the context of the four domains.

Table 1, below, depicts the relationship between variables and domains.

Table 1, Variable & Domain Matrix

Variables	Domain 1 Accountability	Domain 2 Teacher Development	Domain 3 Student Outcomes	Domain 4 Contextualized
Formative		X		
Summative	X			
Novice, Experienced, Master— Treated differently		X		

## Teacher Development and Evaluation

Process or event				X
Goal alignment - Dist, School, Personal	X			
Differentiate Performance		X		
Linked to support and development programs		X		X
Evaluation as a compliance measure	X			
Evaluation as a growth tool		X		
Standards based			X	
Link to state assessment			X	
Student outcome focus			X	
Performance pay				X

### Analysis Techniques

A team of researchers applied a conceptual analysis to school district policy and tools related to teacher evaluation (McTavish & Pirro, 1990; Weber, 1990). Distinct from content analysis, in concept analysis researchers seek groups of words and phrases in an attempt to determine meaning and expression of concepts. Each school district's set of evaluation materials was analyzed by three researchers. A rating sheet for each school district, consisting of the 13 variables, was used by each investigator to record the presence or absence of the items in question. The degree of emphasis for the variable was also recorded along a 1 to 5 scale, with zero indicating no evidence of the variable present in the material from the school district and five acknowledging an exemplary example of the variable. Rating discrepancies of more than four points on a given variable were refereed by a fourth individual. Total raw scores awarded,

mean score for the school district, mean variable score, and four domain scores were calculated for each school district.

An overall state picture was compiled with the aggregate data based on total and average scores among the school districts. These data were further refined into the four domain areas on a state basis. Further analysis was undertaken to determine patterns among categories of school districts, e.g., enrollment, size; urban, suburban and rural; districts with and without collective bargaining units, i.e., teacher unions; relationship to student achievement based on the Colorado accountability programs.

Analysis of the 13 variables in consideration of the four domains was also undertaken. Outcome variables in the form of average student achievement on the Colorado Student Assessment System were also introduced along with other school district characteristics, like the presence of a teacher bargaining unit, i.e., union. Regression analysis was used to determine relationships among the variables, domains and outcome measures (Kleinbau, Kupper, Nizam, & Muller, 2007).

## **Results**

Findings from the study present a paradoxical picture. While the inquiry among the sample school districts revealed a wide variation along the 13 indicator variables and four domain areas, the depth, nature and extent of the variance, and the substance of these differences proved not to be significant based on our statistical examination. Efforts to link relationships to school district characteristics and outcome data proved insignificant as well. Geographic setting, i.e., urban, rural or suburban, was used to predict the overall score. The regression equation was



not significant ( $F(3, 26) = .730, p > .05$ ). Similarly, whether the school district had a bargaining unit for teachers, i.e., an employee union, was also not statistically significant ( $F(3,26) = .310, p > .05$ ) with an  $R^2$  of .035 as was whether the school district met annual yearly progress, per the state's academic accountability system.

We recognized at the outset that the potential for limited statistical conclusions might be an issue given the small sample size. However, we decided a better constructed purposeful sample, i.e., one that more closely profiled the makeup and distribution of teachers, administrators and students in Colorado better suited our inquiry. For example, a census sample or random sample would have under represented urban and suburban school districts and left out a majority of teachers affected by these evaluation systems.

Table 2. Descriptive statistics of the school districts in the study

Cases	District Size	Setting	Bargaining Unit	Overall Ratings
30	< 300 to >50,000	Urban -7		Range
		Suburban – 8	16 = yes	1.51 - 3.41
		Rural - 15	14 = no	

Furthermore, our analysis concluded that the variation we did observe among overall school district scores was almost all below the midpoint of the five point scale, see figure 1. To put it bluntly, our concept analysis found that our sample was rated from mediocre to poor. We speculate about the reason for this in the Conclusion and Implications section, below.

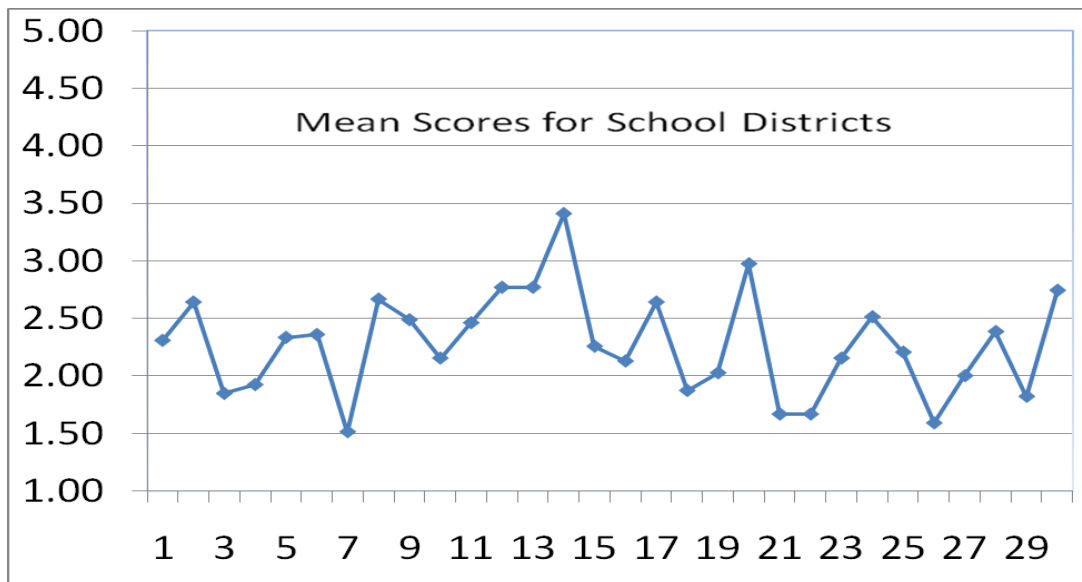


Figure 1. Overall average score for the 30 school districts in the sample

Aggregated scores from the domain areas rendered a picture of teacher evaluation across the state that emphasizes accountability, 3.08, which indicated a slant toward summative evaluations that meet established deadlines for completing required paperwork. The lowest score was for the contextualized domain, 1.64, which revealed that the teacher evaluation systems in the state are overwhelmingly stand alone processes that have little relationship to other systems in the school district, like professional development or pay for performance. The teacher development domain was also poorly rated, 2.06, from which we infer there is little differentiation among teachers being evaluated based on their career status – novice, journeyman, or master teacher. The student outcome domain averaged a modest 2.52 across the state as many of the school district did reference academic standards in their policy and evaluation instrument.

Table 3. Average State-Wide Domain Scores

Accountability	Teacher Development	Student Outcomes	Contextualized
3.08	2.06	2.52	1.64

## Discussion

Our research team reflected on the study results with great pause. The process of applying a conceptual analysis to the policies and instruments from the 30 school districts in the sample generated much discussion about the meaning of our results. This was particularly true in light of the fact that the quantitative analysis showed no significance among the factors considered. However, these findings underscore much of what we found in the literature regarding the lack of impact and bland, uniform treatment of teachers under evaluation (Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D., 2009). Few school districts in our sample distinguished themselves in any of the domain areas. Only a third of the school districts managed a mid-range overall score, leaving two-thirds of the sample below the midline.

As practitioners, our research team proffered suggestions as to why this was the case. The consensus was that teacher evaluations overwhelmingly tend to be events devoid of context. The meaningfulness of the process is diminished for all but those teachers on probationary status. Additionally, administrators are predisposed to be overwhelmed by a process that is compliance oriented, that is, focused on meeting bureaucratic demands to “evaluate” a prescribed number of teachers by a certain deadline. Finally, our observations led us to the conclusion that school district policy interprets instruction in narrowly defined terms and teacher evaluations are circumscribed by a set of forms and checklists.

Another observation from the analysis based on a holistic view of the material points to a remarkable uniformity among many of the policies and forms. In fact, it appeared that many of the school districts had adopted the same policy, procedures and forms even though the state statute called for each district to develop its process with broad stakeholder input. Like a virus, the evaluation policies and procedures were replicated from one school district to the next. We even found an example where one school district failed to change the name on all its forms of the school district from which they copied their evaluation policy.

This study has the potential to initiate a new round of dialogue within the state of Colorado with regard to the veracity and viability of existing state policy for teacher evaluation and teacher development. Individual school districts could also benefit from reviewing the study and using the information garnered from the conceptual analysis as a reference for self-study. Our plan is to move our research focus to school leaders in an effort to understand how they use their time and how teacher evaluation and teacher development play into an overall system of school improvement. This study clearly points to practice that falls far short of the State of Colorado's policy goals. However, we conclude from this conceptual analysis that the flaws in the teacher evaluation and development policies are in the system itself and not the people who are subjected to the policies.

*References*

- Chait, R. & Miller, R. (2010). *Treating different teacher differently: How state policy should act on differences in teacher performance to improve teacher effectiveness and equity*. Washington, DC: Center for American Progress.
- Colorado Revised Statutes, 22-9-106 from  
[http://www.state.co.us/gov\\_dir/leg\\_dir/OLLS/colorado\\_revised\\_statutes.htm](http://www.state.co.us/gov_dir/leg_dir/OLLS/colorado_revised_statutes.htm)
- Donaldson, M. & Peske, H. (2010). *Supporting effective teaching through evaluation: A study of teacher evaluation in five charter schools*. Washington, DC: Center for American Progress.
- Glickman, C. D., Gordon, S. P. & Ross-Gordon, J. M. (2005). *The Basic Guide to Supervision and instructional leadership*. Boston: Pearson Education.
- Kaestle, C. F., 1983. *Pillars of the republic: Common schools and American society, 1780-1860*. New York: Hill and Wang.
- Kleinbaum, D. G., Kupper, L. L., Nizam, A. & Muller, K. E. (2007). *Applied Regression Analysis and Other Multivariable Methods. 4th Ed*. Boston: Duxbury Press.
- McTavish, D. G. & Pirro, E. 1990. Contextual content analysis. *Quality and Quantity*, 24, 245-265.
- Rockoff, J., E. (2004). The impact of individual teachers on students' achievement: Evidence from panel data. *American Economic Review* 94(2), 247-252.
- Rowan, B., Correnti, R., Miller, R.J. & Camburn, E. M. (2009). *School improvement by design: Lessons from a study of comprehensive school reform programs*. Madison, WI: Consortium for policy Research in Education.
- Rivkin, S., E., Hanusheck, E. & Kain, J. (2005). Teachers, school and academic achievement. *Econometrica* 73(2), 417-458.
- Sanders, W., I. & Rivers, J., C. (1996). *Cumulative and residual effects of teachers on future students academic achievement*. University of Tennessee Values-Added Research and Assessment Center.

Scriven, M. (1987). Validity in personnel evaluation. *Journal of Personnel Evaluation in Education* 1, 9-23.

Weber, R. P., 1990. *Basic content analysis*, 2<sup>nd</sup> ed. Newbury Park, CA: Sage publications.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on teacher differences*. Brooklyn, NY: The New Teacher Project. Retrieved on June 1, 2009, from <http://widgeteffect.org/>