Global trends in the Use of Teacher Evaluation for School Improvement:

Implications for South East Asia’s Education Systems

Professor Dr. Philip Hallinger
Faculty of Education
Chulalongkorn University
Thailand

Dr. Philip Hallinger
Professor
Chulalongkorn University
Thailand
Hallinger@gmail.com
During cycles of education reform it is not uncommon to find policymakers evidencing greater interest in various means of enhancing the efficiency and effectiveness of schools. Since 2000, in concert with pressure to hold schools accountable for learning outcomes, education policymakers have increasingly turned to tools from the corporate HR toolbox (Odden, 2003). These tools are grounded in a managerial logic which proposes that the quality of schooling will improve if we become more systematic in the preparation, recruitment, selection, development and evaluation of education personnel, in particular teachers. Since 2000, the focus on teacher evaluation has intensified in selected nations (e.g., USA, UK) as part of a broader accountability-driven quest to improve education (Danielson, 2001; Davis, Ellett & Annunziata, 2002; Donaldson, 2009; Ebmier, 2003; Ellett & Teddlie, 2003; Larsen, 2005; Marshall, 2005; Milanowski, 2004; Toch & Rothman, 2008; Sawyer, 2001). While the logic behind more systematic evaluation of teachers is seemingly irresistible, its efficacy has only found minimal empirical support in its application to the education sector (Bridges, 1990; Frase & Streshly, 1994; Hallinger, Heck & Murphy, 2014; Johnson, 1997; Kersten & Israel, 2005; Odden, 2003).

This policy brief addresses two key questions:

1. What distinguishes the ‘new generation’ of teacher evaluation models as a means of improving the quality of schooling?
2. What have we learned about the effectiveness of teacher evaluation as a strategy for school improvement?
3. What are the implications of this ‘global trend’ for school systems in South East Asia?

**New Generation Teacher Evaluation Models**

We define teacher evaluation as ‘the formal assessment of a teacher by an administrator, conducted with the intention of drawing conclusions about his/her instructional performance for the purpose of making employment decisions’ (Castetter, 1976). This contrasts with instructional supervision which we refer to as growth-oriented coaching conducted by administrators, supervisors or peers. Data gathered during the supervision process are not directly employed for employment decisions (see Attinello et al., 2006; Duke, 1990; Ellett & Teddlie, 2003; Millman, 1981; Popham, 1988).

The logic of using teacher evaluation as a strategy for school improvement is predicated on the strength of the causal relationship between teacher quality and growth in student learning (Gates Foundation, 2013; Milanowski et al., 2005). More specifically, policymakers make two key relevant assertions.

- Variations in the quality of teachers are associated with differences in the learning gains of students (e.g., Sander & Horn, 1994).
- Teaching quality is subject to reliable and valid measurement capable of distinguishing the performance of individual teachers with respect to the achievement of their students (e.g., Danielson, 2001; Hanushek, 2010; Milanowski, 2004; Wright et al., 1997).
Advocates propose that teacher evaluation will positively impact growth in student learning outcomes through three inter-related paths (Danielson, 2001; Heneman & Milanowski, 2007). The operational logic of this strategy can be expressed as follows.

1. First, performance evaluations will ‘weed out’ the weakest teachers, those failing to produce consistently positive effects on student learning (Bridges, 1990; Heneman & Milanowski, 2007; Odden & Wallace, 2008).

2. Second, performance evaluations will provide teachers with meaningful feedback, thereby resulting in improved quality of instruction and enhanced growth in student learning (e.g., Gates Foundation, 2013; Heneman & Milanowski, 2007; Odden, 2004; Wright et al., 1997).

3. Third, teacher evaluation will contribute to development of a results-oriented school culture that will support a broader set of policy interventions designed to foster quality in teaching and learning (De Fraine et al., 2002; Ellet & Teddlie, 2003; Odden, 2004).

These propositions are reflected in two features that distinguish the ‘new generation’ of teacher evaluation models. First, new generation evaluations of teachers are grounded in ‘observable standards’ designed to enhance the quality of judgments made concerning teacher effectiveness (Danielson, 2001). Administrators collect data on teacher classroom behavior through ‘low inference’ classroom observations and compare the results against stated standards (Danielson, 2001). Second, evaluations systematically incorporate data on the achievement of the teacher’s students over the preceding year (e.g., Danielson, 2001; Gates Foundation, 2013; Milanowski, 2004a; Kelly & Downey, 2010). Especially popular among this new generation of teacher evaluation models is the use of ‘value-added measures’ (VAMs) of student gains in learning for that year. VAMs are proposed to represent the individual teacher’s impact on the learning of his/her students during the prior year and describe one important dimension of the teacher’s ‘effectiveness’ (Danielson, 2001; Gates Foundation, 2013; Hallinger et al., 2014; Milanowski, 2004a).

This represents a departure from prior approaches to teacher evaluation that were typically ‘procedural’ and employed ‘high inference’ methods of performance assessment (e.g., checklists). These systems seldom, if ever, included data on student achievement (Bridges, 1990; Duke, 1990). It was previously deemed unfair to hold teachers accountable for a process-product relationship between teaching and learning that was determined to such a large degree by factors outside the individual teacher’s control.

**Empirical Research on the New Generation of Teacher Evaluation**

When we seek evidence on the efficacy of teacher evaluation as a pathway to school improvement, it becomes important to distinguish between what we know works, what we know does not work, and what remains possible but unproven. Let us begin with evidence. First, we note that in the broad literature on school improvement the concept of teacher evaluation is conspicuous by its absence. To the best of our knowledge, no significant review of research on educational effectiveness published internationally in the last 50 years has identified teacher evaluation as a critical variable in the school improvement equation (e.g., Bowles & Levin, 1968; Hattie, 2009; Purkey & Smith, 1983, 1985; Walberg, 2011). Neither does an intensified focus on teacher evaluation find
empirical support of any kind in any of the following bodies of research: effective schools, school restructuring, teacher effects, systemic school improvement, school turnaround, scaling up, comprehensive school reform, and gap reductions.

Moreover, after more than a decade of implementing ‘new generation models’ of teacher evaluation, we conclude that the ‘policy logic’ driving teacher evaluation remains considerably stronger than empirical evidence of positive results. More specifically, a recent review of evidence on the impact of the new generation of teacher yielded the following conclusions (Hallinger et al., 2014).

- Literature on the new generation of teacher evaluation is characterized by overly optimistic interpretations of the underlying studies.

- Efforts to translate ‘academic research’ on teacher effectiveness into practical tools for monitoring the performance of individual teachers fail to meet the technical and administrative requirements needed for this professional task. This is especially evident at the secondary school level where students learn under the guidance of multiple teachers, making it even more challenging to tease out their differentiated effects on growth in learning.

- There is remarkably little empirical evidence that associates the new generation of teacher evaluation with either capacity development of teachers or more consistent growth in the learning outcomes of students. Indeed, research has highlighted the complexity of achieving the necessary ‘fidelity of implementation’ in the new teacher evaluation models in schools.

- There are numerous reasons why the administrators responsible for teacher evaluation find it difficult at best, and counter-productive at worst, to intensify their efforts at teacher evaluation.

In sum, neither historical nor empirical evidence point towards teacher evaluation as a cost effective policy intervention. We highlight the criterion of cost-effectiveness because too often policymakers ignore the costs associated with interventions intended for the improvement of schooling. Yet, any intervention must be assessed not only in terms of its potential to achieve desired goals, but also in terms of its costs.

Teacher evaluation incurs two main costs. The first is the cost of training school administrators with the skills required to conduct valid performance appraisals. Here we use the term ‘valid’ with two criteria in mind: ability of teacher evaluation methods to both improve student learning and stand up in court. The cost of training school administrators with skills that go beyond the application of checklists should not be underestimated. The second cost lies in the increased time required for administrators to conduct these valid appraisals (i.e., one that hold the possibility of positively influencing the quality of teaching and learning). Beyond these limitations, there is little evidence that principals have the skills to operate the teacher evaluation machinery such that it will produce the desired results.

But even if we get past the professional barrier and the training takes hold in the workplace, there are two critical reasons why we are not likely to see much improvement in the quality of instruction in classrooms because of teacher evaluation. First, there is
very little reason to believe that the world in which school leaders currently live, and in which teacher evaluations are not done well, is likely to change anytime soon. Principals have little appetite for evaluating teachers. Moreover, their appetite for this task is unlikely to improve in the foreseeable future given the organizational realities of schooling, regardless of the resources and skills thrust upon them.

**Implications for South East Asian School Systems**

In light of this analysis of ‘new generation teacher evaluation models’ what are the implications for ASEAN school systems seeking effective means of enhancing teacher quality and learning outcomes for children? We propose two responses to this question.

The first response highlights the lack of empirical evidence in support of teacher evaluation as a school improvement strategy, either historically (Bridges, 1990; Millman, 1981) or in the current era (Hallinger et al., 2014). *Simply stated, if we consider the limited financial resources available to education, advocates for investing in teacher evaluation should be held accountable for producing evidence of the relative cost-effectiveness and suitability of this school improvement strategy for addressing the problems of this region.*

Second, we wish to suggest that characteristics of ASEAN’s education systems make the region an even less hospitable context for employing teacher evaluation as a school improvement than the USA or UK (where evidence remains lacking). Specifically, we highlight the following set of requirements or conditions that underlie the logic behind intensifying teacher evaluation.

1. The school system possesses an endorsed ‘teaching model(s)’ against which teaching performance can be assessed via standards.

2. There is a strong system of teacher preparation and professional development in place, along with a reasonable surplus supply of capable teachers.

3. There is substantial cadre of principals and middle level leaders who already possess a strong orientation towards instructional leadership, and who could be trained to the level needed to produce reliable and valid evaluations of teachers.

4. There is a performance-based system of teacher compensation in place, or able to be put into place, that differentiates teacher salaries as a means of offering incentives for teachers.

5. The legal framework of public education enables the removal of incompetent teachers with suitable safeguards but without incurring undue costs.

6. Financial resources are available to support the ongoing training of school leaders and related support personnel in the use of new data/standards-based methods of teacher evaluation.

7. If test scores are going to be used as part of the evaluation process, current tests are aligned to the curriculum in use by teachers and a technical support system exists capable of managing the information systems needed to produce value-added measures that are linked to individual teachers and their students.
A quick perusal of these ‘conditions’ would lead the author to question whether any ASEAN countries, other than possibly Singapore, possesses the conditions needed to successful implement teacher evaluation as a school improvement strategy. Having said this, however, there is no evidence that a problem exists in Singapore’s education system for which teacher evaluation is a suitable solution.

The other ASEAN nations lack too many of these ‘fundamental requirements’ to make teacher evaluation a viable strategy – even in theory – for school improvement. The author would suggest that the investment of financial and human resources needed to implement the new generation model of teacher evaluation – in whole or in part -- would be better spent building the quality of pre-service teacher preparation and in-service teacher development as means of enhancing the quality of teaching and learning. To do otherwise would be to copy a policy solution from the West that not only lacks evidence of success in developed Western nations but which is also wholly unsuited to the conditions and needs of ASEAN education systems.
References


