

Teacher evaluation—accountability and improving teaching practices

Stephan Gerhard Huber¹ · Guri Skedsmo^{1,2}

Published online: 24 May 2016

© Springer Science+Business Media New York 2016

Since the turn of the century, teacher evaluation¹ has been put forward as an important strategy to serve both accountability and school improvement purposes, particularly in countries such as England and the USA, as well as elsewhere. Out of 28 countries surveyed in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes (2013), 22 reported having national- or state-level policy frameworks for teacher evaluation. In the six remaining countries, practices to provide feedback on teachers' work are designed and implemented locally.

1 Some historical perspectives

The evaluation of teachers' work has a long history. Many countries have a long tradition of teachers being assessed by superiors—school leaders/administrators (i.e. head teachers/ principals) or regional supervisors/superintendents (or school board members in a few countries). The approaches used depended on the mandate and varied according to the professional expertise of the evaluators.

In the 1980s and 1990s, school improvement approaches promoted strategies that focused on teachers' self-evaluation or evaluation by peers instead of by superiors. This kind of assessment was undertaken by the professionals themselves and tailored to specific needs for professional development and improvement of individual instructional practices. Strategies such as action research (Altrichter et al.

¹The term 'teacher evaluation' is here used synonymously with 'teacher assessment'.

✉ Stephan Gerhard Huber
stephan.huber@phzg.ch

✉ Guri Skedsmo
guri.skedsmo@ils.uio.no

¹ Institute for the Management and Economics of Education, University of Teacher Education Zug, Zug, Switzerland

² Department of Teacher Education and School Research, University of Oslo, Oslo, Norway

2008) and teacher appraisal (Bollington et al. 1991) were particularly promoted to improve teaching as an important method for school development (Huber 1996, 1999a, b, c). These approaches strengthened the collective capacity and professional accountability of teachers through self- and peer-assisted learning cycles (Schön 1983, 1984; Kolb 1984; Huber 2011).

New approaches have emerged in recent years due to an increased emphasis on accountability policies, drawing strong attention to student learning outcomes, as well as the inputs and processes leading to these outcomes (Murphy et al. 2013). Along with new public management reforms, this type of assessment includes hierarchical relationships and an emphasis on teacher performance as determined by audits that use student performance data, inspection and performance appraisal systems (Gunter et al. 2016). Within this framework, teacher evaluation is an external tool employed to assess teacher performance. Educational authorities and superiors can use the collected information to hold teachers accountable and if possible (for example, in the USA and England), reward high-performing teachers and remove those who do not meet the required standards (Papay 2012).

The past decade has also witnessed various value-added models of teacher evaluation, which aim to estimate teachers' contributions to student achievement (Hallinger et al. 2014). Some studies have produced evidence on the positive effects of these models, while other studies have raised questions regarding the validity and accuracy of these value-added scores as assigned to individual teachers (Darling-Hammond et al. 2012), as well as the challenges related to the consistency of the ratings (Kimball and Milanowski 2009). This new generation of teacher evaluation is in most of the cases grounded in the recently increased body of knowledge of effective teaching, sometimes formulated as standards of teaching. These new approaches are mainly using a uniformed or standard-based approach of evaluation and are conducted by superiors or by external inspectors.

2 Teachers matter and make a difference—but does teacher evaluation?

A growing knowledge base shows that teachers and their instructional approaches are key factors for the effectiveness and improvement of schools, particularly for promoting student learning in terms of academic achievement results as key output variables (Reynolds et al. 2002; Scheerens et al. 2007; Creemers and Kyriakides 2008; Hattie 2009; Papay 2012). At the same time, researchers have reported wide variations in teachers' abilities to raise students' test scores (Lockwood et al. 2007).

Teacher evaluation can be regarded as one of many tools that can be used to strengthen the quality of teaching. It does so by serving different purposes, as similarly outlined by Hallinger et al. (2014): making employment decisions, providing feedback for professional reflection and professional development planning, and setting expectations to develop a performance- and output-oriented school culture. With these aims in mind, teacher evaluation is regarded as functioning at various levels of human resource management and education management, including the system, school/organisation, subject unit and personal levels, and can be differentiated as summative or formative approaches. The important questions are whether it is possible to combine these purposes in a single model of teacher evaluation, and what alternative approaches to teacher evaluation might materialise. The articles in this issue respond to these questions in different ways.

3 Articles in this issue of EAEA 2/2016

Three of the four articles in this issue focus on approaches and frameworks aimed towards evaluating teacher performance, which include taking into account contextual aspects and the complexity of teachers' work.

The first paper's authors, Meng and Munõz, take another perspective as they explore and compare teachers' perceptions of effective teaching in two different countries, China and the USA. They find that the teachers in both countries prioritise student engagement as the main characteristic of effective teaching while other aspects, such as monitoring students, their pace of learning and homework, are the least important areas. Their findings also indicate different values embedded in the teachers' perceptions. The American teachers seem more oriented towards what the authors refer to as a behaviourist belief that emphasises order and routines, repetitive teaching for reinforcement and so on. In contrast, the Chinese teachers seem to focus on factors that (according to the authors) represent moralistic values rooted in Confucianism, such as providing a safe learning environment for students, giving them feedback and regarding each teacher as a person. Interestingly, some of these factors (e.g., engaging students) are at the core of teachers' work but are still excluded from many prevailing frameworks of teacher evaluation.

Roegman and colleagues' article asks to what extent a specific framework (Danielson's 2007 Framework for Professional Practice) can serve several purposes, namely, as a tool for reflection and personal growth in the development of student teachers' practice, as well as for evaluating and grading their performance. The study is situated in a school district in the State of New York in the USA and an additional aim is to prepare the student teachers for evaluations should they teach in this particular district in the future. The authors discuss the challenges in validity and reliability due to the significant variations found in the scores and rationales. They conclude that using the framework for evaluation purposes is problematic and that it is better suited as a tool for reflection and professional growth, especially in preparation programmes. They also raise questions about the fairness of evaluating experienced, novice and student teachers by the same standards, as well as whether more nuanced measures are required according to experience levels.

Amrein-Beardsley and colleagues examine the use of an alternative achievement test in a school district in Arizona in the USA, which aims to include more value-added ineligible teachers in the district's growth and merit pay system. The district's goal is to allow its K-12 teachers to be more fairly and inclusively eligible for individual, teacher-level, value-added scores and the differential merit pay bonuses that correspond with growth. Amrein-Beardsley and colleagues find that using different tests, along with their growth estimates, does not yield similar outputs. They conclude that using the K-12 test for increased fairness and inclusivity is therefore deemed inappropriate. Moreover, they argue that growth is difficult to measure, and even if it is tempting to do so, this may have serious implications for the validity of the inferences and may generate misleading results.

Liu and colleagues also emphasise the importance of fairness in the systems and tools used for teacher evaluation. After investigating how Chinese teachers perceive various indicators and factors of teacher performance evaluation, the authors compare these results with those of a similar American study. Although student test scores are ranked highly by the samples of teachers in both countries, the Chinese teachers perceive students' evaluation of teaching performance, parents' satisfaction with

teachers, and mentoring of other teachers as more important when compared to the American teachers' views. The authors argue that Chinese educators and researchers have an increasingly stronger voice to incorporate a combination of different measures in systems for evaluating teacher performance, including self-evaluation and evaluation by peers, among others.

4 Challenges for research, policy and practice

The articles published in this issue can be regarded as responses to the currently dominant accountability-oriented approaches to teacher evaluation (including value-added models) and narrow concepts of effective teaching. Roegman et al. (in this issue), who explore the extent to which a specific framework can serve as a tool for both professional development and accountability, find that this combination proves difficult. The framework is more suitable for professional reflection and development. The articles also represent attempts to provide alternative approaches to teacher evaluation, whose various purposes are better balanced than many prevailing accountability-oriented models. These alternative methods emphasise elements such as fairness, inclusivity and core aspects of teacher work that cannot be easily measured, such as student engagement.

Questions can be raised about the links between policy and research and the extent to which policies on teacher evaluation consider empirical evidence when new models for teacher evaluations are promoted and implemented (Gunter 2002). In their EAEA article, Hallinger et al. (2014) call for an in-depth examination of the issues. Based on their analysis of the research evidence on teacher evaluation and the impact on school improvement, they conclude: "The policy logic supporting driving teacher evaluation remains considerably stronger than empirical evidence of positive results [...]. Literature on the new generation of teacher evaluation is characterized by overly optimistic interpretations of the underlying literature and a tendency to overlook important limitations of the research designs used in these studies of teacher effectiveness" (p. 21f).

It can be argued that further research is needed in this area. Meta-studies might be interesting but will not help to overcome the methodological problems raised by Hallinger et al. (2014). The methodological critique also points to the inconsistencies in rating of teachers' performance, etc. In line with the articles in this issue, more studies are required to explore other aspects of teacher evaluation, including the possible unintended consequences of evaluation models.

Moreover, we need to engage in conversations across research traditions and domains of practice, dealing with research, policy and practice on key principles of teacher evaluation and teaching practice. If the established systems of teacher evaluation and teaching standards are not perceived as sensible and useful by the key actors involved, one consequence could be a certain adaptation of teaching practices and activities to meet the expectations of these externally formulated standards at a superficial level. Such adapted practices have been characterised as 'window dressing' or 'choreographed accountability', which would not necessarily nurture student learning (Webb 2006). To avoid this situation, we argue for an increased awareness on the selection of evaluation tools. The questions to be

discussed are related to the alignment of purpose and function and the appropriate design of teacher evaluation strategies; hence ‘design follows function’. Moreover, combining too many purposes in one evaluation model might diminish the likelihood of achieving them. Finally, the interplay of teacher evaluation with other types of governing tools also needs to be considered.

References

- Altrichter, H., Feldman, A., Posch, P. & Somekh, B. (2008). *Teachers Investigate their Work. An introduction to action research across the professions*. London & New York: Routledge.
- Bollington, R., Hopkins, D., & West, M. (1991). *An introduction to teacher appraisal*. New York: Cassell.
- Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. New York: Routledge.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15.
- Gunter, H. M. (2002). Teacher appraisal 1988–1998: a case study. *School Leadership and Management*, 22(1), 61–72.
- Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (2016). NPM and educational reform in Europe. In H. M. Gunter, E. Grimaldi, D. Hall, & R. Serpieri (Eds.), *New public management and the reform of education. European lessons for policy and practice*. London: Routledge.
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: an analysis of the evidence. *Journal of Educational Assessment, Evaluation and Accountability*, 26(1), 5–28.
- Hattie, J. A. C. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Huber, S. G. (1996). *Dovetailing school effectiveness and school improvement*. Cambridge: Department of Education, University of Cambridge.
- Huber, S. G. (1999a). School effectiveness: was macht Schule wirksam? Internationale Schulentwicklungsforschung (I). *Schul-Management*, 2, 10–17.
- Huber, S. G. (1999b). School improvement: Wie kann Schule verbessert werden? Internationale Schulentwicklungsforschung (II). *Schul-Management*, 3, 7–18.
- Huber, S. G. (1999c). Effectiveness & improvement: wirksamkeit und verbesserung von schule - eine zusammenschau. *Triologie Internationaler Schulentwicklungsforschung Schul-Management*, 2.
- Huber, S. G. (2011). The impact of professional development: a theoretical model for empirical research, evaluation, planning and conducting training and development programmes. *Professional Development in Education*, 37(5), 837–853.
- Kimball, S. M., & Milanowski, A. T. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34–70.
- Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs: Prentice Hall.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1), 47–67.
- Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: the case of the missing clothes? *Educational Researcher*, 42(6), 349–354.
- OECD (2013). *Teachers for the 21st century. Using evaluation to improve teaching*. Paris.
- Papay, J. P. (2012). Refocusing the debate: assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: international perspectives on school effectiveness*. London: Routledge Falmer.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: Department of Educational Organisation and Management, University of Twente.
- Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.
- Schön, D. (1984). Leadership as reflection-in-action. In T. Sergiovanni & J. Corbally (Eds.), *Leadership and organizational culture* (pp. 36–63). Chicago: University of Illinois Press.
- Webb, P. T. (2006). The choreography of accountability. *Journal of Education Policy*, 21(2), 201–214.