

Teacher evaluation: the need for valid measures and increased teacher involvement

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

Published online: 14 February 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Since the turn of the century, teacher assessment and evaluation have been put forward as an important strategy for assuring and developing educational quality in many countries. 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 were designed and implemented locally.

The USA can be characterised as a forerunner in the implementation of teacher evaluation models. According to Reddy et al. (in this issue), 45 out of 50 states have enacted formal policies that require the use of student achievement measures as important components in teacher evaluation systems. Moreover, federal initiatives, such as Race to the Top (RTT) (see also Lavigne and Chamberlain 2017), national non-profit enterprises (e.g., The New Teacher Project [TNTP]), Measures of Effective Teaching (MET) (research projects funded by the Gates Foundation), as well as other initiatives have resulted in value-added models which attempt to estimate the relative effect of particular teacher and/or school contributions on student test scores or models which aim to combine different measures, such as student achievement and classroom observations.

In earlier issues of EAEA (please see 2016/2 and 2017/2), we discussed several issues related to models of and procedures for teacher evaluation. First, we pointed out that questions can be raised about the links between teacher evaluation policies and research and the extent to which empirical evidence is considered when promoting or

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

✉ Stephan Gerhard Huber
stephan.huber@phzg.ch

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

implementing new models of teacher evaluations (Huber and Skedsmo 2016). On the basis of their analysis of the evidence from research into teacher evaluation and its impact on school improvement, Hallinger et al. (2014) concluded that the policy logic supporting and driving teacher evaluation remains considerably stronger than the empirical evidence of positive results. Moreover, they argued that the literature on the new generation of teacher evaluation models is often characterised by overly optimistic interpretations and a tendency to overlook important limitations of the research designs employed.

Secondly, we have noted contributions in this journal that examined methods for assessing performance which often aim to combine the purposes of holding educators to account for students' learning outcomes and of providing a basis for professional development (see Huber and Skedsmo 2016; Roegman et al. 2016). Thirdly, contributions to this journal have provided insights about the use of different measures of teacher evaluation (e.g., standardised measures of student achievement and standardised procedures for teaching observations) and value-added models (see Briggs and Dadey 2017; Hallinger et al. 2014; Santelices et al. 2017; Skedsmo and Huber 2017). Fourthly, we raised awareness of contextual issues when implementing teacher evaluation models and other output-oriented policy tools (see Santelices et al. 2017). Fifthly, the role of principals in teacher evaluation and their competencies to evaluate and develop teachers' practices have been examined (see Lavigne and Chamberlain 2017). Sixth, the difficulties of using teacher evaluation measures to promote reflective and inquiry-based approaches to developing teaching practices have been addressed (*ibid.*). Finally, there have been contributions which attempted to provide alternative approaches to teacher evaluation whose various purposes are better balanced than many of the prevailing accountability oriented models. These alternative methods emphasise elements such as fairness, inclusivity, and core aspects of teacher work that cannot be easily measured, including student engagement (Amrein-Beardsley et al. 2016; Liu et al. 2016; Meng and Muñoz 2016).

In this issue, one of the contributions follows up on the topic of fair and valid measures of teaching evaluation by examining classroom observation ratings. The other three articles argue for the need for more teacher involvement in the process of designing and implementing teacher evaluation as well as systems of student test procedures and standard-based instruction.

1 Articles in this issue of EAEA 1/2018

In the first article, de Lima and Silva report on a study of the ways in which classroom observation is perceived as a teacher evaluation tool by teachers and department heads in the Azores, one of Portugal's overseas territories, where in 2007, the regional government legislated a new system of teacher evaluation. Their findings provide insights into how teachers and department heads are coping with the new mandatory procedures. The teachers doubted the potential benefits of these measures for their professional development, while the department heads found their new role—that of mid-level leaders responsible for observing and evaluating the performances of their colleagues—challenging in a professional culture built on the principles of equality and autonomy. The authors note a range of implications; for instance, the need for

leadership preparation and training in conducting the observations and, importantly, in how to engage teachers in joint reflections on their practices as well as professional development opportunities.

In the second article, Lei, Li, and Leroux report on the findings from the MET study, in which observation ratings of the CLASS instrument (covering three broad areas: emotional support, classroom organisation, and instructional support) were applied to measure the teaching quality in the classrooms of 317 schools in six large school districts in the USA over a 2-year period. The authors examined classroom-level variations in observation ratings and accounted for the teacher-level, school-level, and rater-level variations. They found that teachers' classroom observation ratings may differ when they teach multiple classrooms as the classroom-level variation may be as great as the teacher-level variation. Based on this finding, they point to several important implications for the interpretation and use of classroom observation ratings as part of teacher evaluations, including the danger of wrongly classifying teachers if their ratings are only based on the observation of one of their classrooms. Moreover, the authors argue that the classroom context needs to be considered and that further research is needed to understand classroom-level variations and to explore within-classroom variations, as they expect that teacher performance may change over time within a classroom.

In the third article, Reddy, Dudek, Peters, Alperin, Kettler, and Kurz examine teachers and school administrators' attitudes and beliefs about teacher evaluation by using the Teacher Evaluation Experience Scale (TEES). The analysis is based on data from 33 school administrators and 583 teachers from 22 schools located in four high-poverty urban districts in the USA. Moreover, both teachers and administrators agreed that their experiences with the teacher evaluation did not increase teachers' motivation to make changes to their practices. Based on these findings, the authors raise questions about the relevance of the evaluation models implemented, as they favour information from the perspective of the evaluator, such as increases in student achievement or teacher value-added methods, while direct teacher input about classroom processes is limited. The authors argue that the teacher evaluation procedures continue to remove teachers' voices from evaluation systems.

In the fourth article, Bonner, Rivera, and Chen explore teachers' beliefs about external testing and classroom assessment and the ways in which these beliefs are related to their implementation of standards-based instructional practices. In their mixed methods study, they combined survey and interview data gathered from secondary school teachers in a large urban district in the USA. The state has a long tradition of high-stakes testing, and more recently, it has adopted a set of learning standards that are aligned with the Common Core State Standards. Based on their analysis, Bonner et al. found that teachers' standards-based instructional practices, perceptions about state tests, and their classroom assessment preferences are systematically associated, though not strongly. Teachers who held positive perceptions about standards-based externally mandated testing were more likely to use standards in their instruction than those who had negative perceptions about the test-based system. Teachers who preferred alternatives to traditional formats of classroom assessment were less likely to hold positive beliefs about state standards-based testing. Bonner et al. identified different groups of teachers based on their perceptions of external testing, the alignment of their teaching to state standards and their support for alternative assessment forms in the classroom.

Further findings showed how some teachers experienced tensions when their professional values and expertise were threatened by the external mandates. Other teachers seemed to manage to align test-based formats in ways that enabled them to pursue their own ideas about good assessment and instructional practices.

2 Reflections and points for further discussion

In general, we can state that many attempts have been made to measure the quality of teachers and teaching, and thus, recently, the number of research studies investigating different models and approaches and their intended as well as unintended consequences has increased significantly. Based on the articles in this issue, we would like to draw the attention to three points. The first relates to the article by de Lima and Silva. This paper challenges the top-down implementation of performance management procedures, which due to their control-orientation and hierarchical character, seems to disrupt the existing school culture and established teacher autonomy. Moreover, the article demonstrates the importance of involving and preparing key actors, such as mid-level leaders, to take a more active role in implementation processes.

The second point concerns the measurement of teacher performance by observation, especially using observation ratings for making high-stakes decisions regarding individual teachers. Following the findings of the study by Lei et al. and their implications, educational authorities in many countries may need to invest far more resources in teacher evaluation if they use classroom observation as an evaluation tool, as the use of this tool requires systematic observations over time and across multiple classrooms to provide valid measures of teachers' performances.

The third point concerns teacher participation in designing evaluation systems. The study by Reddy et al. shows the inconsistencies between the information provided by current teacher evaluation models and the type of feedback teachers need to develop their practices. Furthermore, there are also inconsistencies between the classroom practices that teachers believe in and prefer to use and those practices mandated by policy, which materialise as external testing systems and accountability practices. It can be argued that these inconsistencies call for stronger teacher voices and participation in redesigning teacher evaluation systems, and also for the use of data to increase the relevance of these systems for improving practice.

References

- Amrein-Beardsley, A., Polasky, S., & Holloway-Libell, J. (2016). Validating “value added” in the primary grades: one district’s attempts to increase fairness and inclusivity in its teacher evaluation system. *Educational Assessment, Evaluation and Accountability*, 28(2), 139–159. <https://doi.org/10.1007/s11092-015-9234-5>.
- Briggs, D. C., & Dadey, N. (2017). Principal holistic judgments and high-stakes evaluations of teachers. *Educational Assessment, Evaluation and Accountability*, 29(2), 155–178. <https://doi.org/10.1007/s11092-016-9256-7>.
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: an analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5–28. <https://doi.org/10.1007/s11092-013-9179-5>.

- Huber, S. G., & Skedsmo, G. (2016). Teacher evaluation—accountability and improving teaching practices. *Educational Assessment, Evaluation and Accountability*, 28(2), 105–109. <https://doi.org/10.1007/s11092-016-9241-1>.
- Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: school leaders' perceptions and practices. *Educational Assessment, Evaluation and Accountability*, 29(2), 179–209. <https://doi.org/10.1007/s11092-016-9250-0>.
- Liu, S., Xu, X., & Stronge, J. H. (2016). Chinese middle school teachers' preferences regarding performance evaluation measures. *Educational Assessment, Evaluation and Accountability*, 28(2), 161–177. <https://doi.org/10.1007/s11092-016-9237-x>.
- Meng, L., & Muñoz, M. (2016). Teachers' perceptions of effective teaching: a comparative study of elementary school teachers from China and the USA. *Educational Assessment, Evaluation and Accountability*, 28(2), 179–199. <https://doi.org/10.1007/s11092-015-9230-9>.
- OECD. (2013). *Teachers for the 21st century: Using evaluation to improve teaching*. Paris.
- Roegman, R., Goodwin, A. L., & Reed, R. (2016). Unpacking the data: an analysis of the use of Danielson's (2007) Framework for Professional Practice in a teaching residency program. *Educational Assessment, Evaluation and Accountability*, 28(2), 111–137. <https://doi.org/10.1007/s11092-015-9228-3>.
- Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2017). Two teacher quality measures and the role of context: evidence from Chile. *Educational Assessment, Evaluation and Accountability*, 29(2), 111–146. <https://doi.org/10.1007/s11092-016-9247-8>.
- Skedsmo, G., & Huber, S. G. (2017). Evaluation of educators' performance—balancing various measures to improve practice. *Educational Assessment, Evaluation and Accountability*, 29(2), 107–110. <https://doi.org/10.1007/s11092-017-9262-4>.