



Assessment and evaluation with clarifying purposes for policy and practice

Guri Skedsmo^{1,2}

Published online: 30 May 2020
© Springer Nature B.V. 2020

Information and reporting have been a central element in school governing for centuries. Routines for reporting information to national authorities existed before comprehensive evaluation tools and systems were established (Sivesind 2008). By international comparison, innovations in school systems have been influenced, constructed and traded since the late nineteenth century, long before OECD's Programme for International Student Assessment (PISA) (Lundahl and Lawn 2014). The gathering and use of information to solve problems and implement improvements remain the responsibility of national or state educational authorities in many countries. Cuban refers to this phenomenon as 'fixed responsibility' (2004, p. 21). This concept relates to government processes and the responsibility of the state to provide public services, in this case education, to its inhabitants according to formulated aims. Cuban uses this concept to describe the shift around 1965 from 'fixed responsibility' towards developing practices of 'accountability' in American educational policy. Information is still used for improvement purposes, but holding key actors accountable will ensure that improvement indeed takes place. This issue of EAEA shows that several old tensions related to the purposes and practice of evaluation and assessment in the field of education remain relevant.

1 Articles in this issue of EAEA, 2/2020

In the first article, Ysenbaert, Van Houtte and Van Avermaet report on their case study research exploring schools' assessment policies and teachers' assessment practices in six schools in the Flemish education context in Belgium. Across the cases, several tensions are identified, such as finding the balance

✉ Guri Skedsmo
guri.skedsmo@phsz.ch

¹ Institute for Research on Professions and Professional Learning, Schwyz University of Teacher Education, Goldau, Switzerland

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

between responding to the challenges of diversity in the classroom and aligning the assessments with the requirements of centrally set attainment targets. The authors find a need for schools to use their autonomy to develop assessment policies on the school level that are adapted to the needs of diverse classrooms. Moreover, these assessment policies need to be developed along with a school-wide teaching and learning vision that also embraces diversity among students.

In the second article, Hofflinger and Hippel explore test-based accountability in Chile. In particular, they focus on school responses in terms of an inflation practice wherein low-performing students are made to miss high-stakes tests to prevent the school from being categorised as ‘inadequate’ and avoid the negative consequences of such a categorisation, such as reduced enrolment, funding and autonomy. The authors find that schools are somewhat successful in this effort to evade accountability. Between 10 and 17% of schools that would have been classified as ‘inadequate’ if all students took the test were instead classified as ‘intermediate’ or better. Furthermore, more pressure is placed on schools serving disadvantaged students, given that the same thresholds are used for all schools regardless of the socioeconomic characteristics of the student body. The authors suggest adjusting accountability targets for these schools as a possible means of addressing this issue.

In the third article, Tuytens, Devos and Vanblaere review empirical qualitative and quantitative studies on teacher evaluation. While the quantitative studies focus more on the effects of teacher evaluation on student outcomes, the authors observe that qualitative studies more often consider teacher outcomes, such as behaviour, ability and motivation. Based on the review, they identify various core elements and purposes embedded in evaluation practices which they argue to be part of a value chain. Here, they expand a previous framework and present a more comprehensive model that aligns intended and perceived purposes, individual and organisational outcomes, and individual and organisational resources and context variables. The paper contributes to the existing knowledge based on teacher evaluation by providing insights on how such routines and practices can be embedded in human resource management (HRM).

In the fourth article, Geiger, Amrein-Beardsley and Holloway use 15 cases on teacher lawsuits that occurred throughout the USA as a point of departure for their analysis. The cases were presented in Education Week in 2015 and pertain to the use of student test scores to evaluate teachers. The researchers then purposefully selected the most high profile and controversial of the 15 cases to better understand the measurement and pragmatic issues at play across the cases. The authors find that VAM scores were particularly problematic with respect to reliability and validity. In addition, they found evidence of possible teacher- and school-level bias, bias across the various measures (e.g. observations and PP measures) used and lack of transparency with respect to VAM measures. The selected case occurred in New Mexico, where consequences were attached to teachers’ VAM scores, including the flagging of teachers’ professional files if determined to be not ‘value-adding’. This ‘ineffective’ classification ultimately prevented the teachers from moving teaching positions within the state. Another issue was the teacher termination policies attached to New Mexico teachers’ VAM scores.

In the fifth article, Huber and Helm present and discuss the School Barometer, a fast survey that was conducted in Germany, Austria and Switzerland during the early weeks of the school lockdown to assess and evaluate the current school situation caused by COVID-19. The aim of the survey was to gather, analyse and present data in an exploratory way to inform policy and practice as a basis for reflection and decision-making in a time of crisis as well as for further research. The article presents some exemplary first findings and possible implications for policy, practice and research. The authors reflect on the strengths and limitations of such surveys as well as methodological options for data collection and analysis when used as a short monitoring survey approach. Specifically, they discuss the methodological challenges related to hypothesis testing, the testing of causal effects and approaches for ensuring reliability and validity. The School Barometer represents an example of how assessment and evaluation can contribute in a current crisis while also being part of future long-term crisis management when linked to further research.

2 Some reflections

In the first two articles in this issue, the authors demonstrate and discuss ways by which assessment and evaluation form a basis for and can be integrated with learning and development. In the first article, Ysenbaert et al. argue for the need for school policies, and incentives for schools to develop such policies, to develop integrated models of assessment which align assessment with teaching while providing flexibility for addressing individual needs. In the third article, Tuytens et al. present a more comprehensive model of HRM that aligns the intended and perceived purposes of teacher evaluation, individual and organisational outcomes, and individual and organisational resources and context variables. Two other articles demonstrate assessment practices where pressure and consequences, in terms of labelling of schools and school actors if the measured performance is below the threshold, do not foster learning. Rather, they lead to unfairness and create distrust in assessments (they may not provide accurate information) and the authorities in charge of the systems. In the fifth article, Huber and Helm present an evaluation approach which makes it possible to provide information quickly to inform decision-making in times of crisis.

Some of these articles present approaches where assessment and evaluation have what Popkewitz refers to as a 'policy clarification purpose' (1990, p. 104). He emphasises that policies and reform programmes aim to respond to perceived issues and problems that are not clearly defined and do not have linear outcomes. Assessments and evaluations do not tell us what policy is most efficient or useful, but they can help illuminate the tensions, contradictions and ambiguities that underlie policies and educational practices. '[...] Of the deepest value to the public debates around which schooling in a democracy is positioned (and of importance not only to policy makers) is an understanding of the strains and tensions found in the relations in school arenas [...]' (Popkewitz, 1990, p. 104). Following this, assessment and evaluation can

potentially contribute to an increased understanding which can inform policy and improve practice.

References

- Cuban, L. (2004). Looking through the rearview mirror. In K. A. Sirotnik (Ed.), *Holding accountability accountable* (pp. 18–34). New York: Routledge.
- Lundahl, C., & Lawn, M. (2014). The Swedish schoolhouse: a case study in transnational influences in education at the 1870s world's fairs. *Pedagogica Historica*. <https://doi.org/10.1080/00309230.2014.941373>.
- Popkewitz, T. S. (1990). Some problems and problematics in the production of evaluation. In M. Granheim, M. Kogan, & U. Lundgren (Eds.), *Evaluation as policymaking*. Jessica Kingsley Publishers: London.
- Sivesind, K. (2008). Reformulating reform. *Curriculum history revisited*. Thesis for the Doctor Philos Degree. Faculty of Educational Sciences, University of Oslo.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.