Unintended effects of accountability policies and the quality of assessment and evaluation formats

Guri Skedsmo^{1,2} · Stephan Gerhard Huber³

Published online: 6 January 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

The last 2020 issue of *EAEA* covers several important topics and presents two empirical studies that describe the unintended effects of accountability policies on teachers' practice and work satisfaction. It addresses various concerns related to the quality of assessment and evaluation, including the increasing use of online modes to evaluate courses in higher education institutions, and considers an instrument that enables practitioners to self-assess and develop their assessment literacy. Finally, it discusses the challenges related to remote assessment in higher education during the Covid-19 crisis.

1 Articles in this issue of *EAEA* (4/2020)

Slomp, Marynowski, Holec and Ratcliffe open the issue by describing a study on the use of large-scale, externally mandated exit exams, which are common in many education systems. Such exit exams are considered medium stakes when their scores are blended with school-awarded marks to determine a final course grade. The authors examined the effects of policy decisions on the weighting of exam and school-awarded marks when calculating a student's final blended grade in Alberta, Canada. From a survey of grade 12 teachers in the sciences and humanities ($n = 343$), the authors conclude that exit exams profoundly narrow teachers' planning and assessment practices. The perceived breadth of measurement appears to be an important factor in determining the degree to which exit exams influence this narrowing of teaching and

---

✉ Guri Skedsmo
guri.skedsmo@phsz.ch

✉ Stephan Gerhard Huber
stephan.huber@phzg.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

³ Institute for the Management and Economics of Education, University of Teacher Education Zug, Zug, Switzerland

assessment practices. In this case, a narrowing of the exams' content sampling seemed to drive a narrowing of the curriculum, which the authors, citing Koretz and Hamilton (2006), describe as a 'washback effect' (or 'washback intensity' when considering the degree of the washback effect in one or multiple areas of teaching and learning affected by an examination) (cf. Cheng 2000).

In the second article, Smith and Holloway argue that a disproportionate emphasis on student test scores has given rise what some identify as a 'testing culture' that is fundamentally changing the work and professional identities of teachers, especially when high stakes are attached to the outcomes. Based on previous research, the authors examine whether the relationship between test-based accountability and teacher satisfaction can be partly explained by the weight given to student test scores in teacher appraisals. Using data from the 2013 Teaching and Learning International Survey, they pool data from 33 countries to evaluate the direct and indirect effect of school testing culture on teacher satisfaction. The findings suggest a direct relationship between the intensity of the testing culture and a decrease in teacher satisfaction as well as an indirect relationship, with an emphasis on test scores in teacher appraisals suppressing the appraisals' potentially positive effects on teacher satisfaction.

In the third article, Laupper, Balzer and Berger address the use of online course evaluation to assess teaching quality in higher education, which reflects increasing global demands in the accreditation of higher education institutions. Based on the pilot phase of a programme to evaluate continuing professional development courses for vocational training at a Swiss higher education institution, the authors compare two modes of evaluation questionnaire: the classic paper-and-pencil survey and a corresponding online questionnaire. The two modes were randomly assigned to 33 training courses. Analysing the data via multigroup confirmatory factor analysis (MGCFA), the authors found that the two modes yielded the same results, with no mode-inherent effects contaminating the quality of data. MGCFA, an accepted methodological approach in general mixed-methods survey research, was chosen in this study to address methodological concerns when comparing the two evaluation modes. While the two modes yielded similar results, the authors argue that, with regard to possible effects on data quality, some unresolved aspects merit further exploration, such as the online mode's flexibility on a mobile device in a setting with distractive factors.

In the fourth article, Kelly, Feistman, Dodge, St. Rose and Littenberg-Tobias report on their study testing a Performance Assessment Literacy (PAL) instrument that practitioners can use in self-assessments of their knowledge, competencies and levels of confidence regarding the assessment of students' learning. Applying exploratory and confirmatory factor analysis, the authors investigate the dimensionality of a 27-item survey instrument. The findings show that the instrument captures five reliable dimensions of PAL: valid design, reliable scoring, data analysis, fair assessment and student voice and choice. The authors suggest using the instrument as part of professional development focused on increasing assessment literacy among school staff.

Finally, Giangul, Suhail, Khalit and Khidhir explore the challenges associated in higher education in Oman with conducting assessments remotely during the Covid-19 pandemic. Using a higher education institution as a case, the authors present and discuss various formats of remote assessment. Based on their analysis of data from a quantitative survey of 50 faculties, they identify several challenges, such as academic dishonesty and the risk of cheating, infrastructure problems, the consideration of

learning outcomes as part of online assessment formats and the risk of low commitment among students in submitting assessments. As academic dishonesty and the risk of cheating seemed to be the main problems encountered by the faculties, a range of strategies are reported for minimizing the risk, such as preparing different questions for each student, using online presentation as part of the assessment and combining various assessment methods.

2 Some reflections

Regarding this issue's first topic – the unintended effects of accountability policies – Slomp and colleagues, citing Volante (2006), note that teaching to the test has been construed as a teacher problem, an unethical attempt by teachers to artificially raise test scores. Based on their empirical findings, the authors challenge this notion. One consequence is that teaching to the test can be seen as a response to accountability policies and external pressure. In discussing the implications of their findings, Smith and Holloway argue that 'there is a critical need for school leaders to grapple with the various approaches to appraisal, as well as how appraisals and student test scores might be used in more formative ways'. They suggest that training should include topics such as data literacy, assessment, accountability and the use of strategies that might help leaders navigate the complicated relationship between using test scores and supporting teachers' well-being and growth. At the same time, Smith and Holloway recognize that training and professional development can achieve only so much if policies continue to prioritize high-stakes testing to measure teacher quality.

Concerns related to the quality of evaluation and assessment are addressed in various ways in three of this issue's articles. Over the past decades, as a result of both standardized testing and assessment for learning, there is an increased need for teacher competency in the area of student assessment and evaluation, and a tool for self-assessment could represent a fruitful contribution. Moreover, online formats and modes of evaluation and assessment seem to have a range of advantages; interestingly, flexibility seems to represent both an advantage and a challenge. In the area of online and remote student assessment, there is perhaps a need to explore more holistic student assessment formats and strategies that offer no quick, easy ways to cheat or succeed.

References

- Cheng, L. (2000). Washback or backwash: a review of the impact of testing on teaching and learning. Retrieved from <https://files.eric.ed.gov/fulltext/ED442280.pdf>. Accessed 27 Nov 2020.
- Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). Westport: American Council on Education/Praeger.
- Volante, L. (2006). An alternative vision for large-scale assessment in Canada. *Journal of Teaching and Learning*, 4(1), 1–14.

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