Measuring teaching quality, designing tests, and transforming feedback targeting various education actors

Guri Skedsmo1,2 · Stephan Gerhard Huber3

Published online: 18 August 2020 · © Springer Nature B.V. 2020

Education systems across the world face various challenges which may be caused or increased by global mobility, economic and political crises and more recently, pandemics such as Covid-19. Eventually, dealing with some challenges becomes part of educators’ daily business but still requires flexibility whilst maintaining focus on core issues like educational quality. This issue of EAEA focuses on educational quality regarding (1) measuring teaching quality across countries and regions, (2) designing tests and test formats and (3) providing and transforming feedback to teachers and migrant students for learning and change.

1 Articles in this issue of EAEA 3/2020

In the first article, Aditomo and Köhler analyse data from the 2015 cycle of the Programme for International Student Assessment (PISA) to assess psychometric properties of science teaching quality indicators. Multilevel Confirmatory Factor Analysis (CFA) and Structural Equation Modelling (SEM) were conducted for each country’s or region’s scale to assess school-level reliability, factorial validity and predictive validity. The authors found that the PISA teaching quality scales (except from the classroom management) have low reliability in assessing school-level teaching quality. Moreover, Aditomo and Köhler demonstrate that several teaching scales capture meaningful
differences in teaching quality between schools, while inquiry scale assessment showed poor validity in almost all countries and regions. Based on their findings, the authors argue for using student ratings in PISA to investigate some school-level teaching quality aspects and point to several implications for policy and further research.

In the second article, Quintelier, De Maeyer and Vanhoof report on a study of 687 teachers in 80 Flemish schools that recently went through school inspection to investigate the role of teacher feedback acceptance and willingness to use feedback. Based on their analysis, the authors find that the teachers largely accepted the feedback received. Teachers were more inclined to change their own practice based on teacher core activities feedback compared to school organization feedback. As such, the authors point to the perceived relevance of feedback as an important predictor of subsequent improvement following a school inspection report. Interestingly, Quintelier, De Maeyer and Vanhoof also demonstrate that teachers believing strongly in their ability to teach reported that the feedback did not enhance their understanding of different teaching aspects and learning practices. Based on the authors’ analysis, several implications are discussed, and the authors point to further areas requiring deeper exploration.

In the third article, ShayesteFar criticizes the high-stakes nature of English language tests in many Asian countries and their direct and indirect teaching and learning impacts. While previous research has mostly focused on teaching consequences, ShayesteFar’s study addresses the combined effects of test anxiety, learning stress, attitude and motivation for learning on student achievement on Iran’s University English Requirement Tests. The test procedures were recently changed from multiple tests taking place once a year to four successive National English Achievement Tests (NEATs) that students take over their four high school years. Based on data analysis from 468 English as Foreign Language (EFL) learners, the author found that test anxiety and learning stress negatively correlated with attitude and mediated student attitude effects on achievement, and their joint effect negatively influenced motivation, which has the second strongest direct effect on achievement. Based on the analysis, the author emphasizes the need to synthesize the empirical evidence within a validity framework taking consequences into account to fully understand the impact of changing high-stakes testing.

In the fourth article, Zou and Ou examine and compare EFL reading tests used in the Chinese education context—mobile-based versus paper-based reading regarding use of metacognitive strategies and test media impact. In addition to inferiority of metacognitive regulations on screens, the quantitative and qualitative analysis revealed several complicated and intertwined factors such as reading proficiency, attitudes, adaptability, habits and personality, as well as presentation of text and questions, which may lead to poor mobile test performance.

In the fifth article, Herzog-Punzenberger et al. explore teachers’ reflections on diversity and assessment practices in light of increasing cultural diversity. Based on their analysis of interviews with 115 teachers and school leaders in 20 lower secondary schools in Austria, Ireland, Norway and Turkey, the authors find that many teachers in the selected case schools try to adapt their assessment procedures and grading to help students from diverse backgrounds demonstrate their competencies and experience success. However, such practices tend to result from individual rather than collective efforts organized by school leaders or education authorities. Such practices are also
strongly influenced and at times limited by each country’s government-mandated assessment regimes. Moreover, teachers and school leaders seem to perceive ‘proficiency in the language of instruction’ to characterize classroom diversity, and less attention is paid towards ‘cultural differences’. Based on their analysis, the authors provide several recommendations to support further development of a practicable and just practice of culturally responsive classroom assessments.

2 Some reflections

Regarding this issue’s first topic of measuring teaching quality across countries and regions, Aditomo and Köhler’s article raises awareness on several important aspects, such as validity of teaching quality scales related to country and regions as well as levels (school or classroom) for which they have relevance.

Both the articles by ShayesteFar and Zou and Ou concern the second topic on test design and formats. It is argued to include consequences of high-stakes EFL tests for student learning in validity frameworks. Moreover, media format seems to impact how students perform on EFL achievement tests. Hence, further development of new test media must include preparation regarding metacognitive training, guided practice and feedback.

While Quintelier et al. address the third topic related to providing and transforming inspection report feedback to be used by teachers to develop and change their practices, Herzog-Punzenberger et al. tackle teachers’ practices of assessing and providing feedback to migrant students. For both actors to use feedback to learn and develop, relevance is an important keyword.

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