

Ramona Lorenz, Birgit Eickelmann & Wilfried Bos

High-stakes testing and Educational Governance – Insights and innovations in European school systems

Special Issue Editorial

International research with regard to the intended as well as to the unintended outcomes and effects of high-stakes testing shows that the impact of high-stakes tests has important consequences for the participants involved in the respective educational systems. The purpose of this special issue of the *Journal for Educational Research Online* is to examine the implementation of high-stakes testing in different national school systems and to refer to the effects in view of the concept of Educational Governance.

The use of standardized tests for high-stakes decisions regarding students, teachers and schools has increased in many countries throughout the world in recent years. Almost all countries in Europe arrange for external evaluation and accountability of their schools or teachers in order to improve the quality of education. As a measure and part of Educational Governance, the described development and accountability programs aim to raise the quality of educational systems and provide a higher degree of fairness (Altrichter & Maag Merki, 2010). From the perspective of both school effectiveness and educational economics, the assumption is that an exertion of pressure on schools, teachers, and students by implementing forms of high-stakes testing might, to a certain extent, improve students' achievement. From an organizational perspective on schooling, high-stakes tests affect the output of schools rather than the process within them. Output in an educational sense has an important influence on processes at the different levels of schooling, starting from the individual school level up to the context of the educational administration. In order to grasp the underlying complexity of the implementation of high-stakes testing by means of Educational Governance, an attempt can be made to describe the relationship between high-stakes testing at the different levels of

Dr. Ramona Lorenz (corresponding author) · Prof. Dr. Wilfried Bos, Institute for Education Research and School Development (IFS), TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany
e-mail: ramona.lorenz@tu-dortmund.de
officebos-ifs.fk12@tu-dortmund.de

Prof. Dr. Birgit Eickelmann, Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany
e-mail: birgit.eickelmann@upb.de

the educational system and the interests of educational policies and stakeholders. Educational Governance constitutes a theoretical background to frame these complex structures for the purpose of describing the impact of high-stakes testing. As the concept of Governance has been referred to in the different European school systems more or less explicitly (Amos, 2010), a dynamic process of educational research was started, which is described by Lawn and Grek (2012) as follows:

The study of common, diverse or even oppositional effects in the field of education across Europe today is a growing field of inquiry and research, as a result of many actions and programmes which have grown in volume and complexity in the last few decades, and the development of common European education policies (ibid., p. 7).

This special issue aims to discuss theoretical frameworks and empirical evidence addressing the key issues and challenges of Educational Governance and the interplay between evaluation, accountability and school improvement. The general focus of this special issue adds to these research strands, and relies on national plans and school examinations executed in different European countries. The presented research is based on the most recent empirical studies and statistical analyses. Findings from several European countries with different traditions, different ways of implementing and reasoning high-stakes testing and therefore different models of accountability will be addressed in this special issue.

Within the articles presented in this special issue, there is a focus on the evaluation of recently implemented standardized school leaving examinations in Germany. The purpose of this contribution is to analyze how newly implemented centrally executed examinations contribute to the intention of guaranteeing the increase of fairness of assessment of students' competences. Furthermore, the role of high-stakes tests in the German school system will be discussed in the context of Educational Governance and in so doing will be compared to other countries with a longer tradition of high-stakes testing and different understandings of commitments to such tests. This important focus will be extended to other countries. As a second objective, the matter of research regarding high-stakes testing in another four countries will be presented by first researching and analyzing comparisons of test preparation for state-wide exit examinations at the end of upper secondary education in three countries: Finland, Ireland, and the Netherlands. Secondly, research from single national contexts in Finland and Slovenia will provide a broadened perspective on the main topic of the special issue. In detail, the articles of this special issue broach the following topics:

Lorenz (2016) in her article focuses on the fairness of the German centralized A-level examinations (*Abitur*) by taking into account gender differences. The differences found between the results of boys and girls are discussed in the context of Educational Governance and shed light on the implementation of national *Abitur* examinations in the German school system against the background of targeting higher comparability in assessing students' achievement and providing more fair-

ness by shifting from decentralized to centralized forms of assessment. It is not yet known if the assumed functions of implementing high-stakes testing such as centralized *Abitur* examinations succeed in achieving a fair assessment for all students and different subpopulations of students, such as boys and girls. Using data from state-wide *Abitur* examinations in North Rhine-Westphalia in the school subject English (taught as a foreign language) from the year 2009, gender-specific fairness of the test items is examined by means of differential item functioning (DIF) analysis. Results indicate that parts of the examinations can be found to have a different probability of solution for boys and girls but, in all, only a few items show DIF. The findings are allocated in the context of Educational Governance and discussed in terms of both conclusive methodological indications as well as implications for improving high-stakes testing.

In her contribution, *Klein* (2016) compares teachers' attitudes and strategies in preparing their students for state-wide exit examinations in three European countries. Therefore, the goal of her study is to investigate how teachers prepare their students for state-wide exit examinations at the end of upper secondary education. This is accomplished in an exploratory international comparison of state-wide examinations at the end of upper secondary education in Finland, Ireland, and the Netherlands. The examinations differ in their overall structure, and also in the stakes attached for students and schools. The study comprises expert interviews with the examination authorities to evaluate the structure of the examinations and their consequences, and a questionnaire survey asking 385 teachers in three countries to analyze their attitudes and how they prepare their students for the examinations. Descriptive analyses of the questionnaire data suggest that the teachers in the three countries differ in how they view examination preparation and what they believe is the right amount of time that should be spent on preparation. Nevertheless, their preparation strategies are, with some exceptions, the same in all three countries. Multiple regression analyses indicate that the factors with which the strategies are connected differ across the countries, so it can be assumed that the different examination systems affect examination preparation in rather distinctive ways, but lead to comparable strategies.

Cankar (2016) explores the performance of boys and girls in external examinations in Slovenia at the beginning of upper secondary and tertiary education. He emphasizes that these examinations mirror critical points in students' educational careers at which they leave the compulsory part of schooling behind and have to decide on whether to continue with an advanced school career or, respectively, whether to enroll at a university. Analyzing these crucial points in students' lives, the Educational Governance of transitions between levels of education should, above all, assure fairness in selection procedures. At the point of transition to upper secondary schools, he explores mainly gender differences in students' achievement in mathematics and the Slovenian language at the national examination of knowledge (NPZ), which is obligatory for all students, and will compare the latter according to their school grades. Since only school grades are used as admission criteria for upper secondary schools, this investigation will allow for a compari-

son of school grades with external and more objective measures of student achievement. Should the null hypothesis of no differences between boys and girls be rejected, options for improvement in Educational Governance will be revealed.

At the point of transition to enter a university, in Slovenia, admission criteria to tertiary education usually consist of (externally assessed) Matura results and the school grades of the last two years of upper secondary school. The author compares the effects of both most commonly used measures of academic achievement on admission with a particular focus on gender differences. Again, the rejection of the null hypothesis of no differences between both criteria in view of gender differences would leave room for improvement of Educational Governance in the Slovenian educational system.

The article by *Kupiainen, Marjanen, and Hautamäki (2016)* focuses on the upper secondary matriculation examination in Finland as a school leaving and university entrance examination. The presented research addresses the question of whether increased choice of the subject-specific examinations has the potential to undermine the comparability of examination results and to direct students' choices not only in the examination but already beforehand at school. The authors refer to Finland's tradition of more than 160 years of a national examination connecting the academic track of upper secondary schools with universities. The authors explain the Finnish system by describing the adoption of a course-based (vs. class- or year-based) curriculum for the three-year upper secondary education and the subsequent reforms in the matriculation examination. This increases students' choices considerably with regard to the subject-specific examinations included in the examination (a minimum of four). As a result, high-achieving students compete against each other in the more demanding subjects while the less able share the same normal distribution of grades in the less demanding subjects. As a consequence, students tend to strategic exam-planning, which in turn affects their study choices at school, often to the detriment of the more demanding subjects and, subsequently, of students' career opportunities, endangering the traditional national objective of an all-round pre-academic upper secondary education.

This contribution provides an overview of Finnish upper secondary education and of the matriculation examination (cf. Klein, 2013) while studying three separate but related issues by using data from several years of Finnish matriculation results:

- 1) The relation of the matriculation examination and the curriculum;
- 2) The problems of comparability vis-à-vis university entry due to the increased choice within the examination (cf. Kupiainen, Marjanen, Scheinin, & Hautamäki, 2011);
- 3) The relations between students' examination choices and their course selection and achievement during upper secondary school.

With its focus on innovations in high-stakes testing in several European school systems as a measure of Educational Governance, this special issue topic mirrors the interdisciplinary and international approach of the *Journal for Educational Research Online*.

References

- Altrichter, H., & Maag Merki, K. (Eds.). (2010). *Handbuch Neue Steuerung im Schulsystem*. Wiesbaden, Germany: VS.
- Amos, S. K. (Ed.). (2010). *International Educational Governance*. Bingley, United Kingdom: Emerald.
- Cankar, G. (2016). Governing the transition to higher levels of education and differences between achievement and school grades by gender. *Journal for Educational Research Online*, 8(2), 60–86.
- Klein, E. D. (2013). *Statewide exit exams, governance, and school development. An international comparison*. Münster, Germany: Waxmann.
- Klein, E. D. (2016). How do teachers prepare their students for statewide exit exams? A comparison of Finland, Ireland, and the Netherlands. *Journal for Educational Research Online*, 8(2), 31–59.
- Kupiainen, S., Marjanen, J., & Hautamäki, J. (2016). The problem posed by exam choice on the comparability of results in the Finnish matriculation examination. *Journal for Educational Research Online*, 8(2), 87–106.
- Kupiainen, S., Marjanen, J., Scheinin, P., & Hautamäki, J. (2011, September). *Freedom of exam choice vs. commensurability of exam results: A challenge for upper secondary matriculation examinations*. Paper presented at ECER 2011, Berlin, Germany.
- Lawn, M., & Grek, S. (2012). *Europeanizing education. Governing a new policy space*. Oxford, United Kingdom: Symposium Books.
- Lorenz, R. (2016). Does gender make a difference? Gender-related fairness of high-stakes testing in A-level examinations in English as foreign language in the German state of North Rhine-Westphalia in the context of Educational Governance. *Journal for Educational Research Online*, 8(2), 10–30.