

Dr. David Rutkowski, Associate Professor with a joint appointment in Educational Policy and Educational Inquiry, Indiana University, USA;

Dr. Leslie A. Rutkowski, Associate Professor of Inquiry Methodology, Counselling and Educational Psychology, Indiana University, USA, Professor of Educational Measurement, University of Oslo, Norway

### **Use and misuse of international large-scale assessments: Why it matters to science education and policy**

In the 2018 cycle of PISA, nearly 80 educational systems participated, with new additions that include Brunei Darussalam, a relatively wealthy newcomer to international assessments, the Philippines, which hasn't participated in an international study since TIMSS in 2003, and Belarus, a first-time participant with a per capita GDP that is on par with Thailand and South Africa. Such a heterogeneous collection of participating educational systems poses challenges in terms of deciding what should be measured and how to measure it in a comparable way. This challenge extends to the newest instantiation of PISA: PISA for Development (PISA-D), which emphasizes economically developing countries, including Cambodia, Zambia, and Senegal. To that end, the issue of cultural heterogeneity in international assessments serves as a backdrop against which I consider several challenges to measuring such diverse populations.

To do so, I first provide a brief overview of ILSA history with a specific focus on ILSA growth and how these assessments have attempted to assess science over time. I will then focus the remainder of my talk on the two largest assessments, TIMSS and PISA, providing examples of how each assessment has been used and misused to inform educational policies over time. I will then turn my focus to a number of methodological challenges faced by ILSAs, discussing how they limit the claims that can be made with the data. In so doing, I highlight some recent operational advances in international assessment for dealing with challenges and propose some ways that ILSAs might continue to account for system-level heterogeneity. I conclude with an emphasis on the appropriate use of ILSA data and how, when used correctly, results can work as a useful tool for improving science education around the world.