

Large-scale performance assessments: problems and potentials

Rianne Janssen, Eef Ameel, Jetje De Groof and Alexia Deneire

Why AEA members should attend this workshop

Large-scale assessments are commonly limited to paper-and pencil tests or their digital alternatives. However, for some competences performance assessments are deemed necessary, despite the fact that such assessments bring their own problems. In this workshop we discuss the use of large-scale performance assessments from different perspectives encompassing theoretical considerations, quality assurance, examples of good practices, and issues brought up by the participants. AEA members should attend this workshop if they want to broaden their scope on large-scale assessments or when they need a practical framework to start implementing performance assessments themselves.

Who this Workshop is for

Policy-makers, test developers, psychometricians and experts in educational measurement.

Time schedule

| Time | Session | Presenter |
|-------|---|-----------|
| 09.00 | Coffee and registration | -- |
| 09.30 | Welcome & introduction Outline of the Workshop | |
| 09.45 | Measurement problems and potentials of performance assessments | |
| 11.00 | Break | |
| 11.30 | Quality criteria for large-scale assessments of competences | |
| 13.00 | Lunch | |
| 14.00 | Design and development of large-scale performance assessments with examples of the Flemish national assessments program | |
| 15.30 | Break | |
| 15.45 | Discussion: Introducing large-scale performance assessments in your country? Why (not)? | |
| 16.30 | Workshop close | |

Overview

Measurement issues in performance assessments

The measurement and scoring of the performance of students on such tasks are not without problems. Specific psychometric issues to be dealt with are local dependence, multidimensionality and standard setting. Standard models from Item Response Theory (IRT) may not be fully appropriate and alternative measurement models, like models for the joint classification of items and persons, may seem necessary. It will be illustrated how large-scale assessments of ICT literacy in the K-12 range dealt with these issues.

Our goal is that participants will understand the psychometric challenges of measurement and scoring in performance assessments on the one hand and see ways how to tackle these issues on the other hand.

Quality criteria for large-scale assessments of competences

In this part, we discuss an evaluation matrix, resulting from a research project issued by the Flemish Ministry of Education. The matrix aims to give a framework to evaluate the quality of large-scale performance assessments that focus on quality monitoring at the system level. Firstly, we offer an overview of essential challenges large-scale performance assessment are confronted with. Secondly, we offer insight in the building blocks and quality criteria of high-quality performance assessments. To get an idea how those quality criteria take shape in real-life situations, we analyze a worldwide range of examples of performance assessments.

Our goal is that participants will get a sense of the continuous consideration one has to make between reliability, validity and feasibility when designing large-scale performance assessments. Choosing for one of the components also means that sacrifices have to be made regarding the other components.

Design and development of large-scale performance assessments

The design and development of large-scale performance assessments in practice is the focus of the third session. Based on the experiences within the Flemish national assessments program, guidelines for test development have been set up along with a detailed project time line in order to resolve performance assessments' challenges with respect to test construction, test administration, scoring, data analysis and standard setting. Guiding principles include the translation of attainment targets into specific performance goals on the performance task, organizational feasibility, extensive pilot testing and the development of clear evaluation criteria. Throughout the presentation detailed illustrations of examples of performance assessments are shared.

Our goal is that participants will learn new ways to tackle the many problems imposed by performance assessments so that they may become inspired to consider including performance assessments in their own future assessments.

Discussion: Introducing large-scale performance assessments in your country? Why (not)?

The last time slot of the workshop is devoted to a plenary discussion on the costs and benefits of performance assessments. When is it advisable to introduce performance assessments in large-scale assessment programs and when not? What are good reasons (not) do so?

The goal of this session is to have an open discussion on the issues brought up by the participants to help them develop their own view as policy makers and test developers with respect to the use of performance assessments.

Preparation for the workshop

In order to prepare for the discussion, participants can bring their own examples of performance assessments, or descriptions of situations where the use of performance assessments is being discussed. Participants are also encouraged to read the policy brief (Tucker, 2015) of the recent report of the ETS expert commission on 'Psychometric considerations for the next generation of performance assessment' (Davey et al., 2015).