

CEMO RESEARCH PROGRAM

Methodological Challenges in Educational Measurement

CEMO's primary goal is to conduct basic and applied research seeking to generate new knowledge in the field of educational measurement. In order to achieve this goal, the centre is staffed with researchers comprising a variety of backgrounds and research interests. Although CEMO is primarily concerned with measurement in the context of education, our research is also applicable to other substantive areas.

CEMO research can be categorized into two major strands that are linked to each other:

- Basic research related to educational measurement. This includes research on both classical and modern test theory providing new insights into issues including factor analysis, item response theory, structural equation modeling or latent class analysis, as well as research on advances that combine and extend these techniques across their default boundaries.
- Applied research related to educational assessment. This involves research on the psychometric quality of existing national and international large-scale assessments as well as on methodological issues and challenges relating to novel assessment formats (e.g. CAT), measurement of new constructs (e.g. 21st century skills), and how new types of data (such as logfiles) may be used to address substantive research questions.

Exemplary research fields

Measurement (non-)invariance over time and across groups

Longitudinal modelling

In everyday educational work or clinical practice, development is something that is aspired, monitored, and worked on to make sure that students learn or that patients improve. Usually, progression addresses the question of how far a student or patient moved along a ruler and assumes that as long as the same ruler (i.e., measurement instrument) is used, scores can naturally be compared across time. If the ruler changes or what the ruler is trying to measure changes throughout the process, the common ground for comparisons disappears. Yet, in some situations, such changes would be a sign of development.

For instance, progressing from the level of a novice student-teacher to the level of an expert veteran teacher may not simply correspond to “growing” more of the same competence, but may require redefining teaching practice. Similarly, the reported quality-of-life of patients may undergo a response shift as they redefine what quality of life means for them while a disease progresses or impactful events like operations happen. In such situations, the traditional assumption of measurement invariance over time does not hold and alternative ways to measure and model developmental patterns need to be provided. CEMO aims to develop the procedures involved in the measurement of such attributes and their change over time.

Across-group comparisons

In addition to longitudinal modelling, it is essential for educational assessment to account for potentially differing interpretations of constructs and measures across cultures or subgroups within a culture. This is not only important for clearly communicating research outcomes on group differences, but it is also essential for enabling valid and fair comparisons based on these measures. CEMO focuses on research about measurement equivalence across groups in particular with respect to national and international large-scale assessments.

The issue of measurement equivalence is not a binary yes/no question because measurements will always be approximately invariant only and comparable to some degree. A core objective in CEMO's research is to bridge the gap between the methodology used to study these comparability issues and the statistical modeling used to formalize measurement equivalence (e.g., factorial invariance, differential item functioning, or linking and equating errors) and the actual application of these procedures in practice. This includes both investigating reasons why some subsets of educational assessments are not comparable across countries or subgroups, as well as assessing policy implications of measurement non-equivalence, because not every detected discrepancy needs to be directly relevant for a particular comparison. The latter research objectives connect to essential gaps in research literature and practice with respect to approximate invariance and effect sizes, and the follow-up of non-invariance and biased indicators.

Computer-based assessment

There is a growing concern in the field of measurement that assessments should do more than provide information about *how much* students know or are able to do at critical stages in their educational careers (summative assessment of largely cognitive attributes). Future developments in the methodology and practice of assessment should also increasingly aim at providing students, teachers and schools with information supporting and driving their continuous *learning process* (formative assessment including also non-cognitive attributes). In addition to this demand for increasing the information value and widening the scope of educational tests, a demand for making data collection, processing and reporting as efficient as possible exists.

Technological advances in recent years support and enable such a shift in methodology and practices of assessment. Computer-based assessment can open up opportunities to construct a new generation of educational tests, for instance with new dynamic or interactive assessment formats (possibly implemented as integrated elements of students learning activities) and with access to, modeling of and analysis of additional data

streams (such as response time or activity logs). Although computer-based assessments are already used to measure new areas of skills such as problem solving competences, the potential of these measures has not yet been fully exploited.

Core objectives in CEMO's research are to investigate and study the potential added value and implementation challenges of new assessment formats, such as dynamic and interactive formats; and to develop an expanded psychometric toolbox for dealing with these new forms of assessment. Whereas this might require new and/or adapted statistical measurement models, the new generation of tests will still have to be evaluated in terms of reliability and validity. Thus, although educational testing purposes and formats might change, several of the fundamental measurement concerns will prevail.

SUBSTANTIVE AREAS

An overall rationale of CEMO's applied research is to "unpack" the Nordic model. The Nordic countries provide a unique social and educational context. At a macro level, economies are mostly thriving, there is low income inequality and unemployment, and there is a broad consensus that education across the life-span is mostly a public responsibility. Thus, education, including both early- and higher education, is largely free or heavily subsidized, and accessible for everyone. Moreover, the Nordic pedagogical model differs in considerable ways from educational practice in most other countries. In early education, this includes a strong focus on play-based activities and children's participation. In primary, secondary and higher education, this includes a strong focus on social skills and a positive class climate. CEMO takes two main strategies to "unpack" this Nordic model: the first is through international comparisons, the other is by addressing research questions investigated in other sociopolitical contexts, and analyze them with the specifics of the Nordic contexts in focus.

Early childhood education and care (ECEC)

While considerable research on ECEC has been conducted internationally, the research base is rather scarce in the Nordic countries both regarding effects of ECEC on cognitive and language development and academic achievement, and on potentially negative side effects. Moreover, little is known about variability in the quality of children's ECEC-experiences in the Nordic countries and the consequences of this variability for children's educational attainment. Development of high-quality methodological approaches to measure socio-emotional and cognitive outcomes prior to school age, and to estimate short- and long-term causal effects of ECEC on child outcomes are on CEMO's agenda to enhance knowledge of both positive and negative effects of the Nordic ECEC model.

Primary and secondary education

In Norway, as in many other countries, international large-scale assessments such as PISA and TIMSS, have a central role in monitoring the development of the school system. There is, however, much debate and controversy with respect to the comparability of Norwegian students' performance on these tests to the results of their peers in other countries. Moreover, it is of interest, to what extent cultural differences ("are all Nordic countries the same?"), response styles, translation/language issues, within-country differences ("are all Norwegians the same?"), testing traditions, and more affect these comparisons. Such questions of measurement equivalence have high priority on CEMO's research agenda.

The dominant element of national assessment in education is trust in teachers' own judgements of students. In practical terms this means that teachers are given a mandate for both formative and summative assessment purposes. Grading starts in year 8, and students have to sit for selected central exams at the end of compulsory schooling. Early screening tests and national assessments are applied to gather standardized information about student achievement. CEMO research focuses on the reliability and validity of the national assessments for system-wide monitoring, as well as on the reliability and validity of teacher grades and central exams. In addition, a need exists for studies that investigate how the different types of assessments over grades and years can be linked in order to measure student progress and its antecedents and outcomes.

Furthermore, many research questions in the field of education are causal in nature (e.g., effects of programs or policies), and educational researchers often use observational (non-experimental data), when addressing these questions. Modeling of causes and effects requires the application of sophisticated designs. Objectives of CEMO are to develop and use state-of-the-art research designs involving experimental and longitudinal components as well as causal modeling approaches with observational data. Specifically, there is a need for research aiming at transferring to or redeveloping the latter approaches in the educational context by emphasizing more strongly the psychometrical issues so often neglected in econometric analysis of education.

Higher education

In cooperation with the Faculty of Medicine at the University of Oslo, CEMO examines the reliability and validity of the examination and grading system in medical education. This includes studying the benefits and limitations of different assessment formats as well as the structure, level and development of knowledge and skills during medical education along with its impact on workplace performance. Dimensionality and growth of medical knowledge and skills within a conceptual framework of medical competence, analyses of methods and rater effects within a multi-trait-multi-method framework, and the examination of content, criterion and construct validity will be major research topics in this context. This also includes the development of a feedback system to improve the formative purpose of assessment.