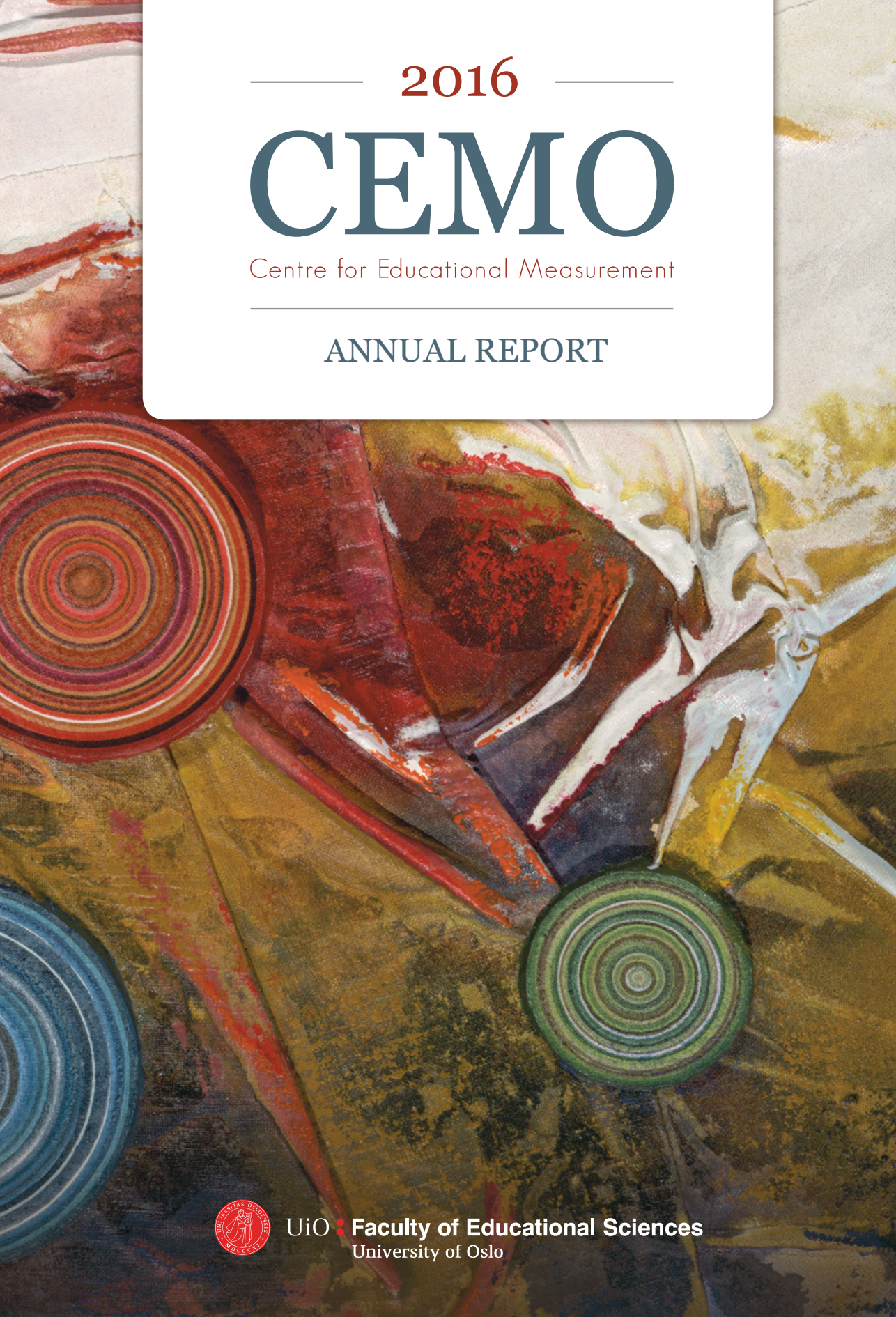
2016

CEMO

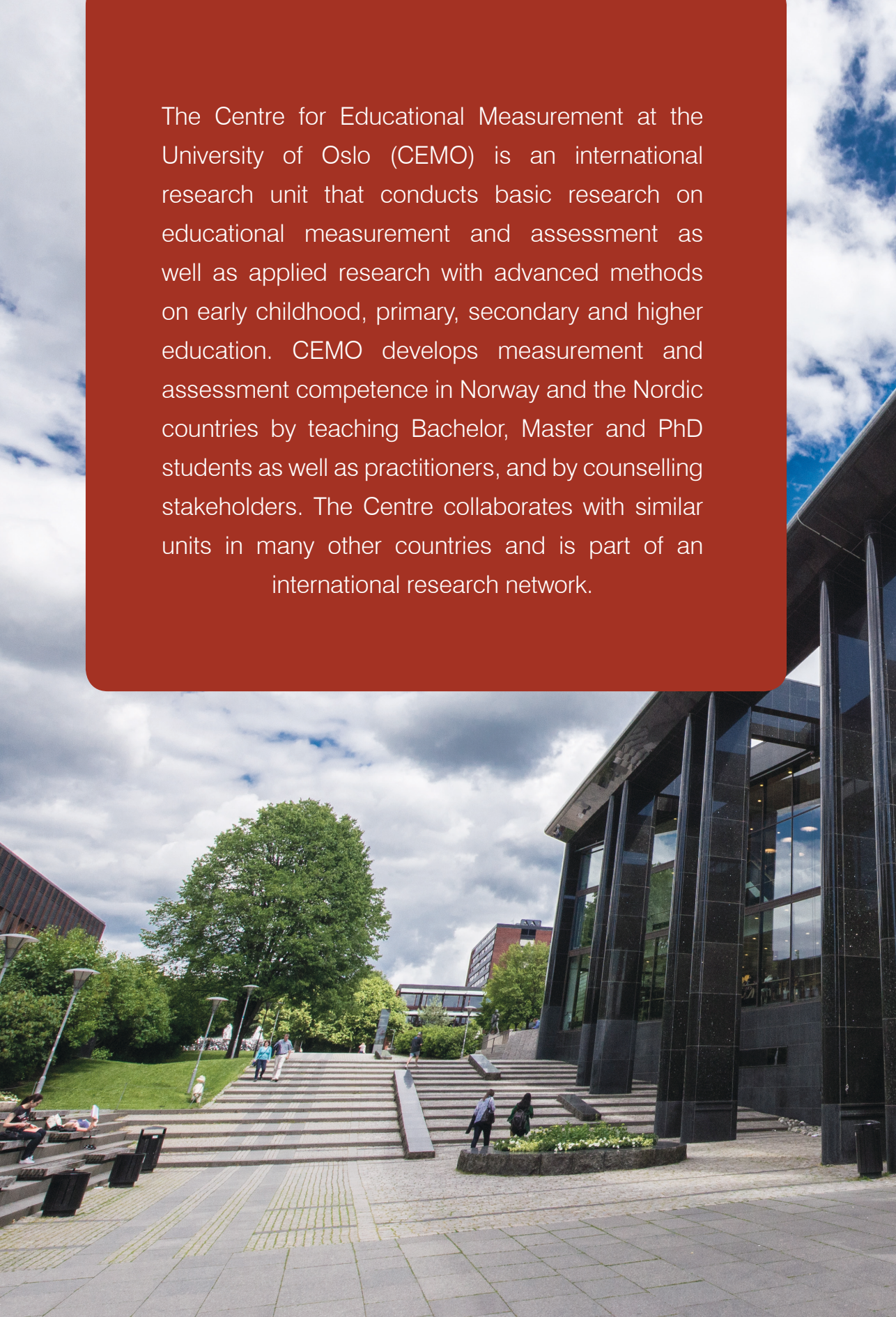
Centre for Educational Measurement

ANNUAL REPORT



UiO : Faculty of Educational Sciences
University of Oslo

The Centre for Educational Measurement at the University of Oslo (CEMO) is an international research unit that conducts basic research on educational measurement and assessment as well as applied research with advanced methods on early childhood, primary, secondary and higher education. CEMO develops measurement and assessment competence in Norway and the Nordic countries by teaching Bachelor, Master and PhD students as well as practitioners, and by counselling stakeholders. The Centre collaborates with similar units in many other countries and is part of an international research network.



CEMO 2016 IN BRIEF

The Centre for Educational Measurement at the University of Oslo (CEMO), chaired by Sigrid Blömeke, includes 25 team members: five professors, four professor IIs, six postdoctoral fellows, five PhD candidates, two administrators and three student assistants. We have thus reached our planned size, and the starting phase has come to an end. The main task of the following consolidation phase will be to develop sustainable research, teaching and outreach activities that strengthen educational measurement in Norway and the Nordic countries.

CEMO was granted five research projects in 2016: two grants by the Norwegian Research Council, in addition to participation in a European Horizon 2020 grant, in a Centre of Teaching Excellence grant awarded by NOKUT to UiO's Faculty of Mathematics and Natural Sciences, and in a grant awarded by the U.S. based Brookings Institute to the Department of Education. Our main task in 2017 will be to consolidate our research portfolio.

More than 40 scientific articles were published by CEMO (co-)authors in 2016. The majority of these appeared in journals on the highest quality level (2) according to the Norwegian publication system. We are in addition proud of a book about teacher quality, instructional quality and student outcomes written together with colleagues from the Department of Teacher Education and School Research at UiO and published in the IEA Research for Education series.

Two prestigious awards were granted to CEMO members in 2016: Sigrid Blömeke received the German Educational Research Association research award, and Leslie Rutkowski received the American Psychological Association Division 5 Anne Anastasi Early Career Award. Anders Skrondal currently serves as president of the Psychometric Society, the core association in our field. CEMO team members are in addition part of editorial boards for scientific journals and expert groups for international large-scale assessments or national research councils. We see these honors and prestigious functions as indicators that CEMO is recognized externally as a strong institution.

CEMO arranged a conference on "Differential Effectiveness" in 2016 on behalf of two Special Interest Groups of the European Association for Research on Learning and Instruction (EARLI). The three-day conference was attended by about 120 participants and provided a great opportunity for CEMO to reach out to the large community of applied researchers in education. The conference will be followed up by a special journal issue and an invited symposium at EARLI's main conference in 2017.

Teaching has become an important mission of CEMO because only through building up measurement and assessment competence, sustainability can be assured. CEMO team members taught PhD classes ranging from introduction to data management to item-response theory or advanced multilevel modeling. In addition we have contributed with PhD supervision and a range of more limited teaching activities at other departments at the Faculty of Education. A large part of our work was devoted to developing a Master program in Assessment and Evaluation that is intended to begin in 2018. The crucial task during our consolidation phase will be to implement this Master program, and in addition to reach out to practitioners in the field by teaching workshops at different places in Norway.

TABLE OF CONTENTS

1 THE DIRECTOR'S COMMENTS

Looking back at CEMO's development in 2016 is a pleasant task. The Centre has reached its crucial objectives by filling up all positions and building up a stable research environment. We have even received several external grants and awards – it is impossible to plan such a success at an early stage of a new institution but it may serve as an indicator of a successful recruiting process.

Growing so quickly is exciting but also includes challenges, in particular in terms of developing a consistent research profile, creating synergy, integrating new team members and assuring internal communication. Our new premises at Forskningsparken have proved to be valuable in this context. With a wing just for ourselves, Forskningsparken provided a home for the CEMO team so that we could become accustomed with each other and develop a joint identity. Never underestimate watercooler conversations!

CEMO currently includes 25 team members from 10 different countries and four continents. We are all very enthusiastic about the chance to work in such a multi-lingual and multi-cultural team. Yuan-Ling Liaw (from Taiwan), Yuanyue Wu (from China) and Tyler Matta (from the US) are Postdoctoral Fellows who started at CEMO during 2016. Tyler is our second Gustafsson & Skrondal Visiting Scholar and followed Sebastian Born from Germany. Kondwani Kajera Mughogho (from Malawi) and Russel Clark Boothby (from the US) started as PhD students. Andreas Frey from Germany joined our Centre as Professor II. In addition, two Norwegian colleagues joined our team in 2016. Rolf Vegar Olsen, former head of EKVA, became the first Norwegian Professor while Fredrik Helland, a Master graduate from IPED at UiO, joined CEMO as the second Norwegian PhD student.

Although we are a 10-minutes walk away from Blindern campus, collaboration with EKVA and ILS, IPED and ISP has increased substantially during the past year. Our mission is to establish educational measurement and assessment as a field that is profoundly developed at UiO and thus for the first time in the Nordic countries. In accomplishing this, CEMO has received tremendous support from the Faculty of Education and the CEMO Board representing the heads of ILS, IPED and ISP. They have served as links to existing research and teaching and provided valuable input to our work. We are looking forward to further joint activities – and we are hoping for a chance to move back to the campus latest in 2018!

Teaching will hopefully be a big part of these joint activities. CEMO has developed a Master program in Assessment and Evaluation because we noticed an urgent need for such competence in Norway and the Nordic countries. All Nordic countries take part in a broad range of international large-scale assessments and carry out national assessments without training assessment specialists. Provided that we will receive the funding necessary to run such a program, next year's activities will be focused on implementing the Master program and to advertising it to different national and international audiences.

We are all looking forward to 2017 with great expectations!

On behalf of the CEMO team,

Sigrid Blömeke, Director of CEMO

2 RESEARCH ACTIVITIES

Basic research is the primary task for CEMO, and team members are specialized on psychometrics and latent variable modelling, measurement equivalence, innovative assessment formats, and causal inference from observational data. These methodological issues frequently appear in different substantive areas where CEMO researchers are involved: from measuring socio-emotional and cognitive outcomes prior to school age, via international large-scale assessments and 21st century skills, to examination models with objective structured clinical examinations and rater effects.

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

- **Research related to educational measurement.** This involves research focusing on methodological challenges of large-scale assessments, and involving new methodologies in the context of Item Response Theory and other latent variable modelling approaches.

- **Research related to educational assessment.** This involves research addressing new formats and modes for assessments, the development of assessments of novel and hard-to-measure constructs, and the use of new types of data (e.g., log-files). In the future, we are particularly seeking to develop research on national assessments.

Research mission

CEMO strives to move the field of educational measurement forward. This includes examining the psychometric quality, fairness and effectiveness of assessments and diagnostic tools as well as developing new measurement approaches to meet the quality criteria in a better way. In addition, moving the field forward includes examining unintended consequences and side-effects of assessments as well as dissemination of knowledge about measurement issues to stakeholders and the public.

A special objective of CEMO is to contextualize educational assessments in the societal and cultural characteristics of the Nordic countries. CEMO's main contribution to reach these ambitions is through long-term and sustained activities including not only research, but also by educating new generations with psychometric qualifications and skills.

Dissemination and counselling

The activities of CEMO are of strong individual and societal relevance. Diagnosing individual strength and weaknesses plays a crucial role in all types of assignments (e.g. to educational interventions) and selection processes (e.g. for a new job). The educational sector plays a key role in preparing both persons and the economy for the future of our societies. Companies, municipalities and other actors are therefore in need of persons with foundational understanding of educational measurement, assessment and evaluation.
CEMO strives to provide this understanding through seminars, advisory activities, teaching and public discussions. In 2017, we will launch our new series of CEMO Policy Briefs where the target groups are policy makers, principals, teachers and others with a keen interest in following research from educational measurement, assessment and evaluation. Furthermore, we are organizing a series of workshops for all national centres and institutions involved in the development of assessments.

CEMO team members also have responsibilities within the larger international organisations running the large-scale international studies. In addition, CEMO has been involved as advisors both to the Norwegian Directorate for Education and Training (Utdanningsdirektoratet) and the Norwegian Agency for Quality Assurance in Education (NOKUT) regarding the national assessments that these two governmental entities are responsible for. As such, CEMO has an impact on dissemination beyond articles published in scientific journals.

Awards

Director Sigrid Blömeke received the German Educational Research Association (GERA)'s Research Award. The award was handed to Blömeke for her outstanding research on the assessment of teacher competencies including effects of teacher education on these.

Professor Leslie Rutkowski received the American Psychological Association Division 5 Anne Anastasi Early Career Award for outstanding contributions to assessment, evaluation, measurement, quantitative or qualitative research methods and/or statistics, and promise of outstanding work.

Publication examples from 2016

Blömeke, S., Jenßen, L., Grassmann, M., Dunekacke, S. & Wedekind, H. (2016). Process Mediates Structure: The Relation Between Preschool Teacher Education and Preschool Teachers' Knowledge. *Journal of Educational Psychology.* Advance online publication. http://dx.doi.org/10.1037/edu0000147 (Open Access)

Data about processes and outcomes of preschool teacher education is scarce. This paper examines prospective preschool teachers' (N=1,851) opportunities to learn during teacher education (OTL) at different program types and stages and their relation to general pedagogical knowledge (GPK), mathematics pedagogical content knowledge (MPCK), and mathematical content knowledge (MCK). Controlling for preschool teachers' background, the results revealed that OTL in general pedagogy and mathematics pedagogy were significantly positively related to GPK and MPCK. How many OTL prospective teachers received was in turn significantly related to the type and stage of a teacher education program in favor of pedagogical colleges compared to vocational schools and in favor of the last year of teacher education compared to the beginning. These findings suggest that specific OTL mediate effects of more distal factors such as the type of institution where prospective teachers are prepared. It may therefore be worthwhile to focus reforms of preschool teacher education more directly on the provision of OTL.

Braeken, J. & Van Assen, M. A. L. M. (2016). An Empirical Kaiser Criterion. *Psychological Methods.* Advance online publication. http://dx.doi.org/10.1037/met0000074 (Open Access)

In exploratory factor analysis (EFA), most popular methods for dimensionality assessment such as the screeplot, the Kaiser criterion, or—the current gold standard—parallel analysis, are based on eigenvalues of the correlation matrix. To further understanding and development of factor retention methods, results on population and sample eigenvalue distributions are introduced based on random matrix theory and Monte Carlo simulations. These results are used to develop a new factor retention method, the Empirical Kaiser Criterion. The performance of the Empirical Kaiser Criterion and parallel analysis is examined in typical research settings, with multiple scales that are desired to be relatively short, but still reliable. Theoretical and simulation results illustrate that the new Empirical Kaiser Criterion performs as well as parallel analysis in typical research settings with uncorrelated scales, but much better when scales are both correlated and short. We conclude that the Empirical Kaiser Criterion is a powerful and promising factor retention method, because it is based on distribution theory of eigenvalues, shows good performance, is easily visualized and computed, and is useful for power analysis and sample size planning for EFA.

Brunk, I., **Schauber, S.** & Georg, W. (2016). Do they know too little? An inter-institutional study on the anatomical knowledge of upper-year medical students based on multiple choice questions of a progress test. Annals of Anatomy, 209, 93-100. http://dx.doi.org/10.1016/j.aanat.2016.09.004 (Open Access)

The depth of medical students' knowledge of human anatomy is often controversially discussed. In particular, members of surgical disciplines raise concerns regarding deficits in the factual anatomical and topographical knowledge of upper-year students. The question often raised is whether or not medical students have sufficient anatomical and topographical knowledge when they graduate from medical school. Indeed, this question is highly relevant for curricular planners.

Therefore, we have addressed it by evaluating the performance of students in the 5th and 6th years of their studies on anatomical multiple choice questions from the Berlin Progress Test Medicine performed at 10 German university medical schools. Results were compared to a reference based on a standard setting (modified Angoff-procedure). The reference was established independently by 5 panels of anatomists at different universities across Germany. As the ratings were independent of university affiliation, teaching-experience or training of the anatomists, an overall cut off score could be calculated which corresponded to 60.4% correct answers for the question set used in this study. In the progress test, on average only 29.9% of the students' answers were correct, reflecting that the performance was significantly below the expected standard. On the basis of the test results it remained unclear whether acquisition or retention of anatomical information was insufficient. Further evaluation by item characteristics revealed that the students had major difficulty in applying their theoretical knowledge to practical problems in the context of a clinical setting. Thus, our results reveal deficits in the anatomical knowledge of medical students in their final years. Therefore medical curricula should not only focus on enhancing the acquisition and retention of core anatomical knowledge, but aim at improving skills applying this in a clinical setting.

Paap, M. C. S., Lenferink, L. I. M., Herzog, N., Kroeze, K. A. & van der Palen, J. (2016). The COPD-SIB: A newly developed disease-specific item bank to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *Health and Quality of Life Outcomes,* 14(1). http://doc.utwente.nl/100615/1/art%253A10.1186%252Fs12955-016-0500-0.pdf
(Open Access)

Health-related quality of life (HRQoL) is widely used as an outcome measure in the evaluation of treatment interventions in patients with chronic obstructive pulmonary disease (COPD). In order to address challenges associated with existing fixed-length measures (e.g., too long to be used routinely, too short to ensure both content validity and reliability), a COPD-specific item bank (COPD-SIB) was developed. Items were selected based on literature review and interviews with Dutch COPD patients, with a strong focus on both content validity and item comprehension. The psychometric quality of the item bank was evaluated using Mokken Scale Analysis and parametric Item Response Theory, using data of 666 COPD patients. The final item bank contains 46 items that form a strong scale, tapping into important themes that were identified based on literature review and patient interviews. The 46-item COPD-SIB has good psychometric properties and content validity. Items are available in Dutch and English. The COPD-SIB can be used as a stand-alone instrument, or to inform computerized adaptive testing.

Rutkowski, D. & Delandshere, G. (2016). Causal inferences with large scale assessment data: Using a validity framework. *Large-scale Assessments in Education* 2016, 6(4). https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-016-0019-1
(Open Access)

To answer the calls for stronger evidence by the policy community, educational researchers and their associated organizations increasingly demand more studies that can yield causal inferences. International large scale assessments (ILSAs) have been targeted as a rich data sources for causal research. It is in this context that we take

up a discussion around causal inferences and ILSAs. Although these rich, carefully developed studies have much to offer in terms of understanding educational systems, we argue that the conditions for making strong causal inferences are rarely met. To develop our argument we first discuss, in general, the nature of causal inferences and then suggest and apply a validity framework to evaluate the tenability of claims made in two well-cited studies. The cited studies exemplify interesting design features and advances in methods of data analysis and certainly contribute to the knowledge base in educational research; however, methodological shortcomings, some of which are unavoidable even in the best of circumstances, urge a more cautious interpretation than that of strict "cause and effect." We then discuss how findings from causal-focused research may not provide answers to the often broad questions posed by the policy community. We conclude with examples of the importance of the validity framework for the ILSA research community and a suggestion of what should be included in studies that wish to employ quasi-experimental methods with ILSA data.

Rutkowski, L. (2016). A Look at the Most Pressing Design Issues in International Large-Scale Assessments: A Paper Commissioned by the U.S. National Academy of Education. Washington, DC: US National Academy of Education. https://naeducation.org/wp-content/uploads/2016/12/Pressing-Methodological-Issues-in-International-Assessment-Rutkowski-2016_web-version.pdf (Open Access)

Three pressing design issues in international large-scale assessments, such as Progress in International Reading Literacy, Programme for International Student Assessment, and Trends in International Mathematics and Science Study, are outlined. In all three cases, the importance of the matter at hand and proposed solutions are set against the backdrop of educational policy. The first matter regards issues around cultural comparability of the test and context questionnaires. Cultural modifications to current studies are recommended. The second topic takes up the presence and problem of measurement error, particularly in context questionnaires, and the way that less error-prone measures might be collected in the case of key reporting variables. The final topic deals with the desire to draw causal inferences from international assessment data and the challenges therein under current designs. Two proposals are offered for strengthening the foundation upon which select causal effects might be estimated. Although none of the proposed solutions are trivial, each one offers the possibility of better meeting the demands placed on international assessments in a modern, globalized, and highly heterogeneous world.

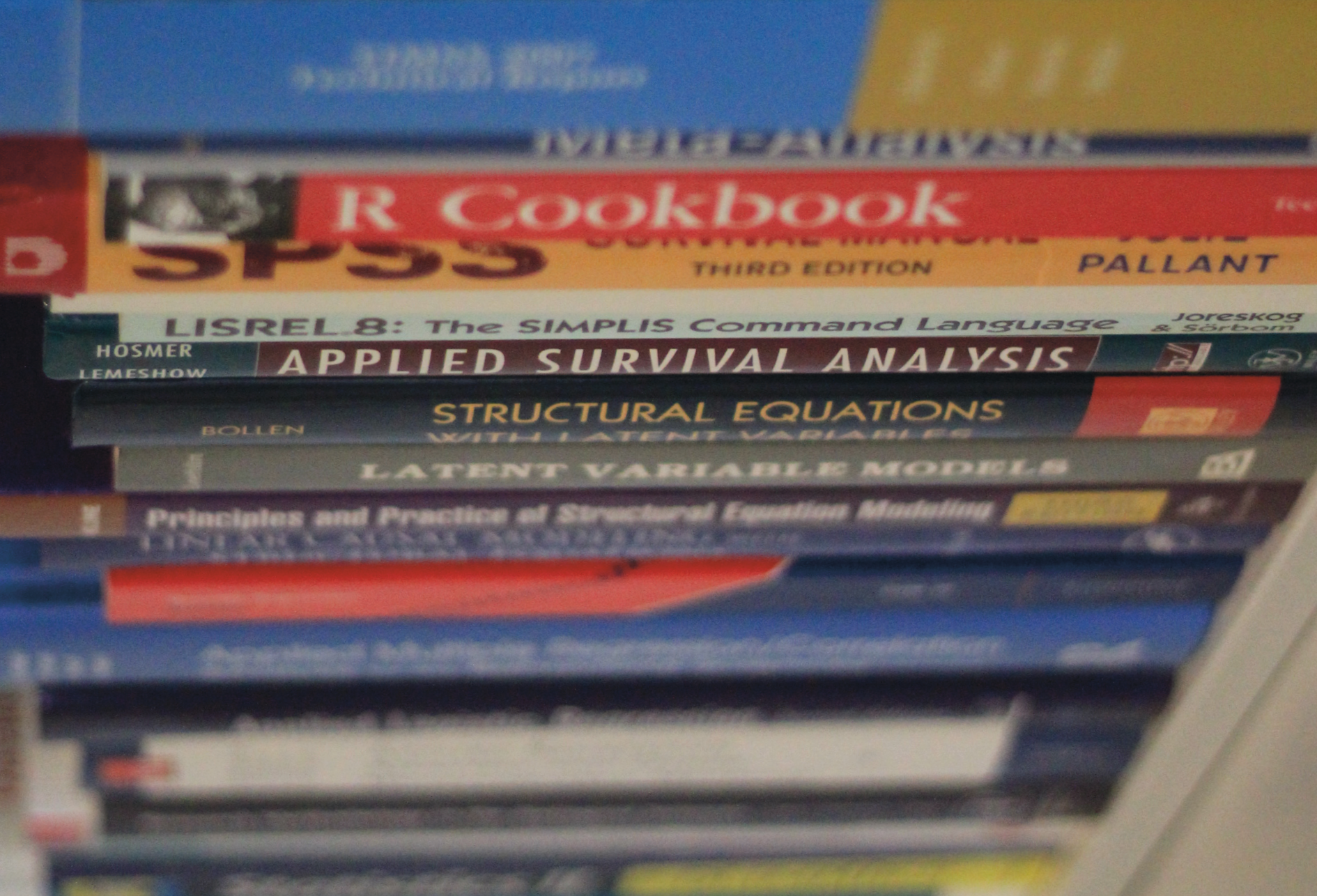
Scherer, R., Nilsen, T. & Jansen, M. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. Frontiers in Psychology, 7. http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00110/full (Open Access)

Students' perceptions of instructional quality are among the most important criteria for evaluating teaching effectiveness. The present study evaluates different latent variable modeling approaches (confirmatory factor analysis, exploratory structural equation modeling, and bifactor modeling), which are used to describe these individual percep-

tions with respect to their factor structure, measurement invariance, and the relations to selected educational outcomes (achievement, self-concept, and motivation in mathematics). On the basis of the Programme for International Student Assessment (PISA) 2012 large-scale data sets of Australia, Canada, and the USA (N=26,746 students), we find support for the distinction between three factors of individual students' perceptions and full measurement invariance across countries for all modeling approaches. In this regard, bifactor exploratory structural equation modeling outperformed alternative approaches with respect to model fit. Since bifactor models are particularly useful for disentangling the general and specific components of perceived instructional quality, we encourage researchers to consider abandoning unnecessarily strict assumptions on the factor structure. Our findings revealed in addition significant relations to the educational outcomes.

Siddiq, F., Hatlevik, O. E., **Olsen, R. V.**, Throndsen, I. & **Scherer, R.** (2016). Taking a future perspective by learning from the past: A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, 19, 58-84. http://www.sciencedirect.com/science/article/pii/S1747938X16300252 (by January 14, 2017 one of the most downloaded articles of this journal in the last 90 days)

This study systematically reviews literature on assessment instruments of primary and secondary school students' ICT literacy. It has three objectives: (1) Describe the development and characteristics of the assessments; (2) Present a synthesis of the facets of ICT literacy measured; and (3) Investigate to what extent information about reliability and validity is provided. A total of 38 tests reported in 66 studies were included. The results indicate that most of the tests target lower secondary students, comprise multiple-choice item designs, and are evaluated by quantitative methodology. The majority of the tests measure facets such as searching, retrieving, and evaluating digital information, and technical skills. In particular, the access to tests measuring digital communication, collaboration, safety, and problem solving is limited. This review demonstrates that an adequate norm for documenting and reporting test quality is lacking. Our findings point to potential future directions in developing and reporting assessments of ICT literacy: Although qualitative studies with systematic observations of students engaging with the test material may be perceived to be costly and time consuming to implement, we believe they are in great demand. We recommend that the future development of assessment of ICT literacy also targets primary and upper secondary school. In particular, assessing students at an early age may help identify the need for interventions to enable all kids to make constructive use of ICT for future learning. Without knowledge of the properties it is difficult, if not impossible, to pass any judgement on the quality of information in research studies. We therefore encourage researchers to report sufficient evidence on the reliability and the creation of a validity argument when presenting new instruments to assess ICT literacy.


R Cookbook
SPSS
THIRD EDITION
PALLANT
LISREL 8: The SIMPLIS Command Language
Joreskog & Sorbom
HOSMER
LEMESHOW
APPLIED SURVIVAL ANALYSIS
BOLLEN
STRUCTURAL EQUATIONS
LATENT VARIABLE MODELS
Principles and Practice of Structural Equation Modeling

3 TEACHING ACTIVITIES

CEMO is involved in a range of teaching and development activities from the BA through the PhD level and professional development within areas at the core of our research profile. The relationship between the quality of our research and teaching is regarded a two-way street where both are mutually dependent on the other.

At the BA level, we contribute to revising the BA programs based at the Departments of Education (IPED) and Special Education (ISP) with the objective to teach parts of the basic methods classes. At the MA level, we contribute to teaching in the international programs based at IPED. In addition, CEMO supervises several Master students from other programs who write their Master thesis as part of research projects at CEMO. We have employed several student assistants, too, who are directly involved in staff research.

Besides regular but small contributions to the teacher education program offered at the Department of Teacher Education and School Research (ILS), CEMO is not systematically involved in initial teacher education. We are more involved in the professional development of school principals (*Rektorutdanningen*), and we will substantially expand our teaching of workshops for test developers in 2017.

The development of our Master in Assessment and Evaluation has been completed, and is now on its way through the procedures regulating the establishment of new degrees at UiO. Assuming that the application is successful, we look forward to advertising this program to Norwegian, Nordic and international audiences. We are currently signing agreements with several Norwegian research institutes that will provide students the possibility to conduct their Master thesis projects in cooperation with professionals in the field and potential employers dealing with issues of measurement, assessment and evaluation.

In 2016, CEMO arranged six PhD courses with study points (Bayesian statistics for educational measurement; Computer adaptive testing; Multilevel and longitudinal modeling; Regression analysis; Introduction to educational measurement; Multilevel regression analysis) and two PhD courses without study points (Introduction to statistical thinking; Learn how to write empirical research papers). In addition to supervising our own PhD candidates, we are also involved in supervision of candidates at the other departments at the Faculty of Educational Sciences.

4 OUTREACH ACTIVITIES

An explicit assignment for CEMO is to reach out to non-specialists related to the educational sector and in need of knowledge about measurement, assessment and evaluation. Educational measurement often has profound impact both on individuals and on processes and outcomes of teaching and learning. This, in combination with the fact that educational measurement often is technically complex, generates different informational needs on the side of students, parents, teachers, school-leaders, politicians and administrative bodies.

A central part of CEMO's scientific outreach is a weekly open Brown Bag seminar where junior and senior academics informally present their research to colleagues while these have lunch. This series has proven to be a brilliant opportunity to discuss work in progress in a friendly atmosphere and a chance to talk to each other. For a full list of presenters see the last pages of this report.

CEMO has in addition continued to organize High Profile talks where renowned colleagues from other countries were invited to present hot topics of current measurement debates. This year's invitees were Bernhard Veldkamp from the University of Twente, Netherlands, and Matthias von Davier from Educational Testing Service in Princeton, USA.

Conference on Differential Educational Effectiveness

A major 2016 outreach activity was the three-day conference on differential educational effectiveness that took place in September. The main conference was preceded by a two-day junior researchers' pre-conference for PhD students. The conference was jointly organized by CEMO, the research group "Large-Scale Educational Assessments (LEA)" and the EARLI Special Interest Groups 18 "Educational Effectiveness" and 23 "Educational Evaluation, Accountability and School Improvement". The conference organizing team was led by Ronny Scherer.

The conference presented state-of-the-art research on differential accountability and effectiveness, and how this can inform and enhance school improvement in order to promote not only quality but also equity in education. Through keynotes of internationally renowned colleagues and a broad range of symposia and poster sessions,

- the international evidence was summarized for both an international and a Norwegian audience,
- an awareness of the methodological challenges involved in studying differential accountability and effectiveness was created,
- implications of differential accountability and effectiveness for policy-making, educational practice, and research were critically discussed, and
- suggestions for future research and potential

instructional and institutional approaches for teachers, principals, and school leaders to differential accountability and effectiveness, were suggested.

We are currently following-up the conference through special journal issues and an invited symposium at the main EARLI conference in Tampere, Finland in 2017.

EERA Spring School on Advanced Methods in Educational Measurement 2016

From April 18 to 22, CEMO hosted the 6th EERA Spring School on Advanced Methods in Educational Research. The theme was longitudinal structural equation modeling, and lecturer Todd D. Little (Texas Tech University) reviewed both foundations of and recent advances in longitudinal research. The participants learned about issues of design, issues of measurement, and issues of analysis. The three primary models for longitudinal data including the panel model, the growth model, and the intra-individual p-technique model were presented, and participants learned when each model is preferred and how to interpret the parameters of each model. The target audience of the Spring School was early stage researchers including PhD students, postdocs and assistant professors from any field of educational effectiveness research. The limitation of no more than 25 participants offered hands-on activities during the computer-based training modules. During the Spring School, all participants had to give a Pecha Kucha, where they presented their own research to other participants and senior researchers.

Seminar on large-scale educational analysis

On October 18, the research group LEA (Large-scale Educational Analysis) organized a seminar as part of the series of such events at the Faculty of Education to mark the Faculty's 20th anniversary. Given that results from both TIMSS and PISA were soon to be released, this seminar focused on what may be learned from participation in the large-scale international studies. The seminar was very well attended and several persons from EKVA and CEMO made significant contributions through talks and poster presentations.

External workshops by CEMO members

CEMO members have taught at international venues in addition to their classes at UiO, for instance:

- **Rutkowski, D.** & Meinck, S. (2016). Writing Policy Briefs with IEA Data, at the AEA conference, Cyprus.
- **Braeken, J.** (2016). Workshop on computerized adaptive testing, at the Norwegian military.

Oslo Science Fair

Oslo Science Fair is part of the yearly nationwide event to fuel the public's curiosity and understanding of science and research. On September 23 and 24, CEMO members participated at an information and hands-on stand together with the Department of Special Needs Education. Using a wheel of fortune and a board with stickers, CEMO members tried to help the public gained an understanding of hypothesis testing and simple statistical analysis.

Websites and social media

During 2016, CEMO further developed both the Norwegian and English websites. Numbers of followers on social media increased significantly. Both Twitter and Facebook were used several times a week to spread information about CEMO's research activities, possibilities, and cooperation. On the websites the main features remain the personal pages for each CEMO member, information about CEMO's research and teaching, upcoming events, and the list of publications. Overall, our websites had more than 60,000 hits in 2016.

5 MANAGEMENT & ADMINISTRATION

CEMO is established as a research unit hosted by the UV Faculty at UiO. The Centre is located at Gaustadalleen 30. The Norwegian Ministry of Education and Research and UiO are CEMO's main funders. They constitute the final reporting entities that define the guidelines under which CEMO operates. The UV Faculty is responsible as the main provider of administrative support.

Administrative structure

The centre is run by the director, Professor Sigrid Blömeke. In collaboration with co-director Professor Rolf Vegar Olsen and under the supervision of the CEMO Board, the director's responsibilities include strategic decisions about CEMO's research, teaching and outreach profile, about CEMO's personnel tableau, recruitment strategies and employments as well as the management of the CEMO budget. The leader team also represents CEMO at the Faculty and higher UiO levels as well as outside the university.

Senior Advisor Anne-Catherine Lehre is responsible for the daily running of CEMO. The administration also consists of Higher Executive Officer, Øystein Andresen. Operative tasks of the administration include, among other things, external communication, facilitating a good reception and stay for guests, maintenance of the websites and social media, taking minutes from board meetings, recruiting interviews, and scientific advisory board meetings, and organizing and implementing the different events like courses, seminars, and workshops. CEMO's administration also functions as permanent secretariat for the CEMO Board and the International Advisory Board.

The UV Faculty operates employments at CEMO as well as budgeting and accounting. IT support is provided by the Department of Teacher Education and School Research.

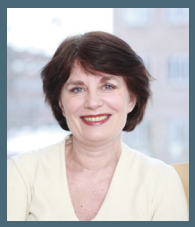
The CEMO Board and CEMO's International Scientific Advisory Board

The CEMO Board is an administrative body that meets three to four times per year to approve the CEMO budget, the director's progress reports about research, teaching and outreach activities at CEMO and the employments. In addition, the department heads and the student and emplyee representatives provide feedback to CEMO's activities from an internal perspective.

The CEMO Board	
NAME	AFFILIATION
Chair: Rita E. Hvistendahl	Head of Department of Teacher Education and School Research, UiO
Berit Rognhaug	Head of Department of Special Needs Education, UiO
Ola Erstad	Head of Department of Education, UiO
David Rutkowski	Employee representative
Celestina da Silva	Student representative

CEMO established in 2016 an International Scientific Advisory Board to receive feedback on our research, teaching and outreach activities from renowned and highly experienced international colleagues working in similar contexts in other countries. The board shall also promote research between CEMO and other international research centers. The board had its first meeting in June and discussed the CEMO portfolio including the Centre's organizational set-up during a full-day meeting. As a result, CEMO developed several activities for 2017 specifically directed towards Norwegian and Nordic contexts.

International Scientific Advisory Board	
NAME	AFFILIATION
Cees Glas, Professor of Educational Measurement	University of Twente, The Netherlands Chair of the Department of Research Methodology, Measurement and Data Analysis
Susan Embretson, Professor of Quantitative Psychology	Georgia Tech, USA
Irwin Kirsch, Director of the Center for Global Assessment	EEducational Testing Service ETS Distinguished Presidential Appointee
Sophia Rabe-Hesketh, Professor of Educational Statistics and Biostatistics	University of California, Berkeley, USA Fellow of the American Statistical Association and Elected Member of the National Academy of Education in the U.S.



Comments by the CEMO Board chair: Rita Hvistendahl

The CEMO Board had two meetings in 2016 on June 2 and November 16, both at the Department of Teacher Education and School Research. A number of issues were also decided upon in electronic board meetings on February 15, March 2 and August 3. In the first meeting, the CEMO Board welcomed David Rutkowski as employee representative on the board.

An important activity in 2016 has been to announce and fill positions. The board has welcomed Rolf Vegar Olsen as full professor and Andreas Frey as Adjunct Professor. The board has also welcomed Yuan-Ling Liaw as postdoctoral fellow and Fredrik Helland, Russel Clark Boothby and Kondwani K. Mughogho as PhD students. The board would like to thank the committees for their great efforts.

Another important issue has been to discuss a CEMO master program, and the board had the great pleasure of approving the program for the CEMO Master in Assessment and Evaluation. This is a new and innovative program involving specialized theoretical and practical coursework that is intended to strengthen the methodological competencies in Norway and in the Nordic countries in general, but the program will also accept candidates from other countries.

At a meeting with the Secretary of the Ministry of Education on September 2, the CEMO director Professor Sigrid Blömeke and Professor Rolf Vegar Olsen could report impressive scientific activities such as PhD, master's and bachelor's courses, seminars and conferences, research and dissemination, and international collaboration, all in the field of Educational Measurement. The international CEMO Master program was thoroughly presented by Associate Professor Johan Braeken.

CEMO has had a number of successful applications, nationally and internationally, for research funding and grants. These, and the many other projects, show that CEMO is actively achieving its potential as a research center.

The Ministry of Education and Research has decided to support the CEMO for another five years (2018-2022). This is based on a number of aspects, including: the reports frequently submitted by CEMO, the websites where all activities are clearly documented, the assessment of the center done by the Faculty of Educational Sciences, and the running dialogue between the Ministry and CEMO. The CEMO Board takes great pleasure in congratulating CEMO on their receiving support for another five years.

6 FINANCES

Accounting principles

The Norwegian Ministry of Education and Research (7.200 MNOK core-funding to CEMO) and UiO (several positions) are CEMO's main financial contributors.

The table below shows the financial statement and budgeted expenses for the Project 205435 (Centre for Educational Measurement).

Revenues and expenditures 2016

		Financial statement	Budgeted expenses
Opening balance		-9 356 580	-9 356 580
Total Opening balance		-9 356 580	-9 356 580
Funding	Core funding	-20 516 249	-10 454 000
	External income		
	Income from sales	-250 403	-150 000
Total funding		-20 766 652	-10 604 000
Staff expenses	Salary cost	7 188 449	7 077 712
	Overtime	12	
	Salary (variable)	247 774	24 890
	Holiday pay, payroll tax, pension	2 981 326	3 072 695
	Salary expenses	7 588 274	236 600
Total staff expenses		18 005 835	10 411 897
Operating expenses	Consultancy service	62 687	0
	Rent	203 557	150 000
	Travel costs, courses, conference	943 432	638 000
	Other operating expenses	394 272	1 534 000
Total Operating expenses		1 603 948	2 322 000
Investments	Investments	108 115	200 000
Total investments		108 115	200 000
Netto contribution	Own funding (UiO)	299 205	0
	Overhead	-466 442	-292 000
	Salary Reimbursement	-620 072	-730 972
Total netto contribution		-787 309	-1 022 972
Project closing balance	Project closing balance	-59 616	0
		52 925	
Total project closing balance		-6 691	0
Total		**-11 199 334**	**-8 049 655**

Planned revenues and expenditures 2017

		Budget
Opening balance		-11 199 334
Total Opening balance		-11 199 334
Funding	Core funding	-11 647 000
	External income	
	Rental/sales Income	
Total funding		-11 647 000
Staff expenses	Salary cost	8 478 137
	Salary (variable)	108 672
	Holiday pay, payroll tax, pension	3 616 080
	Salary expenses	
Total staff expenses		12 202 889
Operating expenses	Consultancy service	0
	Rent	402 668
	Travel costs, courses and conference	691 000
	Other operating expenses	2 656 529
Total Operating expenses		3 750 197
Investments	Investments	100 000
Total investments		100 000
Net contribution from externally funded projects	Own funding (UiO)	964 024
	Overhead	-2 223 492
	Salary Reimbursement	-2 154 048
Total net contribution from externally funded projects		-3 413 516
Total		**-10 206 764**

7 APPENDICES

CEMO team members

Name	Nationality	Position	Period
Gustafsson, Jan-Eric	Sweden	Professor II (-Jul 2014, since UV Fac.)	Oct 2012-
Lehre, Anne-Catherine WG	Norway	Senior Adviser	Jan 2013-
Scherer, Ronny	Germany	Postdoctoral Fellow	Jan 2014-
Braeken, Johan	Belgium	Associate Professor	Feb 2014-
Zachrisson, Henrik D.	Norway	Professor II	Jul 2014-
Blömeke, Sigrid	Germany	Director	Aug 2014-
Andresen, Øystein	Norway	Higher Executive Officer	Aug 2014-
Daus, Stephan	Norway	PhD Candidate	Oct 2014-
Skrondal, Anders	Norway	Professor II	Jan 2015-
Tesema, Melaku Tesfa	Ethiopia	PhD Candidate	Jan 2015-
Schauber, Stefan	Germany	Postdoctoral Fellow	Mar 2015-
Rutkowski, Leslie	USA	Professor	Sep 2015-
Rutkowski, David	USA	Professor	Sep 2015-
Paap, Muirne	Netherlands	Postdoctoral Fellow	Oct 2015 -
Olsen, Rolf Vegar	Norway	Professor	Apr 2016-
Frey, Andreas	Germany	Professor II	Sep 2016-
Helland, Fredrik	Norway	PhD Candidate	Sep 2016-
Liaw, Yuan-Ling	Taiwan	Postdoctoral fellow	Oct 2016-
Boothby, R. Clark	USA	PhD Candidate	Nov 2016-
Mughogho, Kondwani K.	Malawi	PhD Candidate	Nov 2016-

CEMO Gustafsson-Skrondal visiting scholarship

Name	Nationality	Period
Born, Sebastian	Germany	Aug - Oct
Matta, Tyler	USA	Oct-

CEMO guest researchers

Name	Nationality	Period
Kampa, Nele	Germany	Jan - Feb
Kuger, Susanne	Germany	Feb - Jun
Lotz, Christin	Germany	Apr - Jun
Emslander, Valentin	Germany	Aug - Oct
Niepel, Christoph	Germany	Sep
Wu, Yuanyue	China	Sep -

CEMO research assistants

Name	Nationality	Period
Svarva, Vivi Bull	Norway	Aug - Dec
Midthaug, Mari Bratterud	Norway	Aug - Dec
Aursand, Leah Rose	USA	Nov -

CEMO events

Conferences

International conference EARLI sig. 18 & 23 Joint conference	
Theme	Date
JURE pre-conference	Sep 26-27
Closing the Gaps? Differential Accountability and Effectiveness as a Road to School Improvement	Sep 28-30

High Profile talks

Name	Seminar title	Date
Veldkamp, Bernard	Computer Adaptive Formative Assessment in the LIBE Online Learning Environment	Mar 31
Von Davier, Matthias	Innovations in PISA Methodology	Oct 14

Brown bag seminars

Name	Seminar title	Date
Schauber, Stefan	Diagnostic performance by medical students working individually or in teams	Jan 12
Canrinus, Esther Tamara	Coherence and assignments in teacher education (CATE) study	Jan 26
Tveit, Sverre	Constructing legitimate national testing policies in Norway and Sweden	Feb 2
Kampa, Nele	Mathematical and science abilities – the interplay with g and relation to school grades in literacy and curricular large-scale assessments across grades	Feb 9
Løkken, Ingrid Midteide	Measuring children's social and emotional competence in Norwegian ECEC	Feb 16
Scherer, Ronny	Differentiation in teachers' self-efficacy - Does working experience matter?	Mar 1
Braeken, Johan	Developmental screening assessments with incrementally ordered block structure in the test design: A Bayesian item response modelling approach	Mar 8
Zachrisson, Henrik D.	Current debates about replication in psychology	Mar 15
Melby-Lervåg, Monica	A 100 new studies: working memory training does still not improve performance on measures of intelligence or other measures of “far transfer”	April 5
Bergersen, Gunnar	Ready for the 21st century jobs? Increasing realism, relevance and rigour in the assessment of programming skills	April 12
Scherer, Ronny	Argumentation in science and first-language learning – What do students think this is all about?	May 3
Helland, Fredrik	Explanatory item response modelling of an abstract reasoning assessment: A case for modern test design	May 10
Nilsen, Trude	School characteristics moderating the relation between SES and achievement	May 24
Lotz, Christin	Students’ strategy application while solving	Jun 14

	complex problems and its relation to intelligence	
Reber, Rolf	Interventions to increase interest at school	Aug 30
Born, Sebastian	Considering item position effects in computerized adaptive testing (CAT)	Sep 6
Niepel, Christoph	Determinants of academic self-concept	Sep 13
Foldnes, Njål	Consistent p-values for moment structures	Sep 20
Frey, Andreas	Application of computerized adaptive testing in large-scale assessments	Oct 4
Jensen, Ragnhild	Paper-based and on-screen reading: Does delivery mode influence students' reading comprehension?	Oct 25
Hjetland, Hanne Næss	Development of language and reading comprehension ability from 4 years to 4th grade	Nov 1
Lawrence, Joshua	Word generation: A systematic review of experimental and quasi-experimental evaluations	Nov 8
Matta, Tyler	Additive polynomial models for longitudinal data with seasonality: Formulation and implementation	Nov 15
Liaw, Yuan-Ling	When can multidimenstional Item Response Theory models be a solution for differential item functioning?	Nov 22
Roe, Astrid & Tengberg, Michael	Interrater reliability of constructed response items in standardized tests of reading	Nov 29
Skedsmo, Guri & Mausethagen, Sølvi	Data use in education: Accountability practices and integration of knowledge sources in "result meetings"	Dec 13

Courses

UV9916V1: Bayesian Statistics for Educational Research, Prof. David Kaplan, Jan 19-21
A three day course introducing the basic elements of Bayesian statistics and to show through discussion and practice, why the Bayesian perspective provides a powerful alternative to the frequentist perspective.

6th EERA Spring School on Advanced Methods in Educational Research, Prof. Todd Little, Apr 25-27
A week long course on longitudinal structural equation modeling and issues of design, of measurement, and of analysis. The panel model, the growth model, and the intra-individual p-technique model were presented, and participants learned when each model is preferred and how to interpret the parameters of each model.

UV9256: Computer Adaptive Testing, Dr. David Magis, Dr. David Stillwell, & Associate professor Johan Braeken, Apr 25-27
A three day course on issues encountered during the setup of a computerized adaptive test, starting from the design towards the actual delivery of a CAT.

UV9257: Multilevel and Longitudinal Modeling, Professor Sophia Rabe-Hesketh and Professor Anders Skrondal, May 30- Jun 2
A four day course introducing models for multilevel or clustered data, such as cross-sectional data with students nested in schools, or longitudinal data with repeated measures/panel waves nested in subjects.

Introduction to Statisical Thinking, Associate professor Johan Braeken, Sep 12, 19 and Oct 3, 10, 17, 24 and 31
The course was a series of introductory lectures on fundamental concepts in statistics and modern data-analytical practices with emphasis on statistical literacy and development of statistical thinking, stressing conceptual understanding, rather than mere procedural knowledge.

UV9225: Introduction to Educational Measurement, Professor Leslie Rutkowski, Oct 3, 17, 25, 31, Nov 7, 14, 28 and Dec 5
The course intended to be an introduction to relevant concepts, theories, and applications related to educational measurement.

UV9214 - Regression Analysis, Associate professor Johan Braeken, Oct 19, 20, 26 and 27
A four day course focusing on basic concepts and principles of simple and multiple regression, and various strategies for using multiple regression.

Learn how to write empirical research papers, organized by Professor II Henrik D. Zachrisson and Postdoctoral Researcher Ronny Scherer, Oct 21, 28, Nov 10, 22, Dec 2, Jan 12, 26 and Feb 2 (2017)
This course offered participants first-hand experiences in writing an empirical paper for an academic journal and receiving feedback from peers and scholars.

UV9253 – Multilevel Regression Analysis, Postdoctoral Researcher Ronny Scherer, Nov 14, 15 and 21
This course gave an introduction to multilevel modeling with a focus on regression analysis. In particular, it presented the basic concepts of multilevel structures and evaluated different modeling techniques.

Publications and presentations

*Contributors affiliated with CEMO in bold; * = Open Access*

Articles in peer-reviewed journals (level 2 or level 1 with Impact factor>1)

Impact factor = number of citations of articles in a given year published in the preceding two years

Arnesen, A., **Braeken, J.**, Baker, S., Meek-Hansen, W., Ogden, T. & Melby-Lervåg, M. (2016). Growth in Oral Reading Fluency in a Semitransparent Orthography: Concurrent and Predictive Relations with Reading Proficiency in Norwegian, Grades 2–5. *Reading Research Quarterly.* Doi: 10.1002/rrq.159

*** Blömeke, S.**, Busse, A., Kaiser, G., König, J. & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education: An International Journal of Research and Studies*, 56:35-46. Doi: 10.1016/j.tate.2016.02.003

*** Blömeke, S.**, Jenssen, L., Grassmann, M., Dunekacke, S. & Wedekind, H. (2016). Process Mediates Structure: The Relation Between Preschool Teacher Education and Preschool Teachers' Knowledge. *Journal of Educational Psychology.* Doi: 10.1037/edu0000147

Braeken, J. & Blömeke, S. (2016). Comparing future teachers' beliefs across countries: approximate measurement invariance with Bayesian elastic constraints for local item dependence and differential item functioning. *Assessment & Evaluation in Higher Education* 41(5): 733-749. Doi: 10.1080/02602938.2016.1161005

*Brunk, I., **Schauber, S.** & Georg, W. (2016). Do they know too little? An inter-institutional study on the anatomical knowledge of upper-year medical students based on multiple choice questions of a progress test. Annals of Anatomy, 209, 93-100. http://dx.doi.org/10.1016/j.aanat.2016.09.004

Dunekacke, S., Jenssen, L., Eilerts, K. & **Blömeke, S.** (2016). Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: a process model. *ZDM – the International Journal on Mathematics Education* 48(1-2):125-137.

Frey, A., Raphael, B. & Born, S. (2016). Umgang mit Itempositionseffekten bei der Entwicklung computerisierter adaptiver Tests. *Diagnostica (Göttingen)*. Doi: 10.1026/0012-1924/a000173

*** Frey, A.**, Seitz, N-N. & Brandt, S. (2016). Testlet-based multidimensional adaptive testing. *Frontiers in Psychology* 7: 1-14. Doi: 10.3389/fpsyg.2016.01758

Greiff, S., Niepel, C., **Scherer, R.** & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior* 61:36-46. Doi: 10.1016/j.chb.2016.02.095

* Hautz, S.C., Schuler, L., Kämmer, J.E., **Schauber, S.**, Ricklin, M.E., Sauter, T.C., Maier, V., Birrenbach, T., Exadaktylos, A. & Hautz, W.E. (2016). Factors predicting a change in diagnosis in patients hospitalised through the emergency room: A prospective observational study. *BMJ Open* 6:e011585(5). Doi: 10.1136/bmjopen-2016-011585

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J. & **Blömeke, S.** (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM – the International Journal on Mathematics Education* 48(1-2): 41-53. Doi: 10.1007/s11858-016-0759-y

Hoth, J., Schwarz, B., Kaiser, G., Busse, A., König, J. & **Blömeke, S.** (2016). Uncovering predictors of disagreement: ensuring the quality of expert ratings. *ZDM – the International Journal on Mathematics Education* 48(1-2): 83-95.

Kaiser, G., **Blömeke, S.**, Köning, J., Busse, A., Döhrmann, M. & Hoth, J. (2016). Professional competencies of (prospective) mathematics teachers: Cognitive versus situated approaches. *Educational Studies in Mathematics.* Doi: 10.1007/s10649-016-9724-5

Kjøbli, J., **Zachrisson, H. D.** & Bjørnebekk, G. (2016). Three Randomized Effectiveness Trials — One Question: Can Callous-Unemotional Traits in Children Be Altered? *Journal of Clinical Child and Adolescent Psychology.* Doi: 10.1080/15374416.2016.1178123

*** Paap, M. C. S.**, Lenferink, L. I. M., Herzog, N., Kroeze, K. A. & van der Palen, J. (2016). The COPD-SIB: a newly developed disease-specific item bank to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *Health and Quality of Life Outcomes,* 14(1). Doi: 10.1186/s12955-016-0500-0

Pankow, L., Kaiser, G., Busse, A., König, J., **Blömeke, S.**, Hoth, J. & Döhrmann, M. (2016). Early Career Teachers' ability to focus on typical students errors in relation to the complexity of a mathematical topic. *ZDM – the International Journal on Mathematics Education* 48(1-2): 55-67. Doi: 10.1007/s11858-016-0763-2

Rohatgi, A., **Scherer, R.** & Hatlevik, O. E. (2016). The role of ICT self-efficacy for students' ICT use and their achievement in a computer and information literacy test. *Computers and Education* 102: 103-116. Doi: 10.1016/j.compedu.2016.08.001

*** Rutkowski, L.** & **Rutkowski, D.** (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher* 45(4): 252-257. Doi: 10.3102/0013189X16649961

Rutten, E.A.P., Bachrach, N., van Balkom, A.J.L.M, **Braeken, J.**, Ouwens, M.A. & Bekker, M.H.J. (2016). Anxiety, depression and autonomy–connectedness: The mediating role of alexithymia and assertiveness. *Psychology and Psychotherapy: Theory, Research and Practice* 89.(4):385-401. Doi: 10.1111/papt.12083

Scheerens, J. & **Blömeke, S.** (2016). Integrating teacher education effectiveness research into educational effectiveness models. *Educational Research Review* 18: 70-87. Doi: 10.1016/j.edurev.2016.03.002

*** Scherer, R.** (2016). Learning from the Past – The Need for Empirical Evidence on the Transfer Effects of Computer Programming Skills. *Frontiers in Psychology* 7(1390). Doi: 10.3389/fpsyg.2016.01390

*** Scherer, R.**, Nilsen, T. & Jansen, M. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. *Frontiers in Psychology*, 7:110. Doi: 10.3389/fpsyg.2016.00110

Siddiq, F., Hatlevik, O. E., **Olsen, R. V.**, Throndsen, I. & **Scherer, R.** (2016). Taking a future perspective by learning from the past - A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review* 19, 58-84. Doi: 10.1016/j.edurev.2016.05.002

Siddiq, F., **Scherer, R.** & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st century education. *Computers and education*, 92-93: 1-14. Doi: 10.1016/j.compedu.2015.10.006

* Teig, N. & **Scherer, R.** (2016). Bringing formal and informal reasoning together - A new era of assessment? *Frontiers in Psychology* 7. Doi: 10.3389/fpsyg.2016.01097

Tondeur, J., van Braak, J., Siddiq, F. & **Scherer, R.** (2016). Time for a new approach to prepare future teachers for educational technology use: Its meaning and measurement. *Computers and education*, 94: 134-150. Doi: 10.1016/j.compedu.2015.11.009

Wood, S., Niven, K. & **Braeken, J.** (2016).Managerial abuse and the process of absence among mental health staff. Work, *Employment and Society* 30(5):783-801. Doi: 10.1177/0950017015613755

Articles in peer-reviewed journals (level 1)

Brevik, L. M., **Olsen, R. V.** & Hellekjær, G. O. (2016). The Complexity of Second Language Reading: Investigating the L1-L2 Relationship. *Reading in a Foreign Language*, 28(2):161-182.

Gustafsson, J-E., Nilsen, T. & Hansen, K.Y. (2016). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation.*

* Laschke, C. & **Blömeke, S.** (2016). Measurement of Job Motivation in TEDS-M: Testing for invariance across countries and cultures. *Large-scale assessments in education* 4(16). Doi: 10.1186/s40536-016-0031-5

* Lenferink, A., Effing, T., Harvey, P., Battersby, M., Frith, P., van Beurden, W., van der Palen, J. & **Paap, M.C. S.** (2016). Construct Validity of the Dutch Version of the 12-Item Partners in Health Scale: Measur-

ing Patient Self-Management Behaviour and Knowledge in Patients with Chronic Obstructive Pulmonary Disease. *PLoS ONE*, 11.(8). Doi: 10.1371/journal.pone.0161595

* Mikolajetz, A. & **Frey, A.** (2016). Differenciated assessment of mathematical competence with multidimentional adaptive testing. *Psychological Test and Assessment Modeling* 58(4): 617-693.

*** Rutkowski, D.** & Delandshere, G. (2016). Causal inferences with large scale assessment data: Using a validity framework. *Large-scale assessments in education* 6(4). Doi: 10.1186/s40536-016-0019-1

Rutkowski, L., Rutkowski, D. & Zhou, Y.(2016). Item Calibration Samples and the Stability of Achievement Estimates and System Rankings: Another Look at the PISA Model. *International Journal of Testing* 16.(1):1-20. Doi: 10.1080/15305058.2015.1036163

*** Scherer, R.**, Jansen, M., Nilsen, T., Areepattamannil, S. & Marsh, H. W. (2016). The quest for comparability: Studying the invariance of the teachers' sense of self-efficacy (TSES) measure across countries. *PLoS ONE*, 11(3). Doi: 10.1371/journal.pone.0150829

* Siddiq, F. & **Scherer, R.** (2016). The relation between teachers' emphasis on the development of students' digital information and communication skills and computer self-efficacy: the moderating roles of age and gender. *Large-scale assessments in education* 4. Doi: 10.1186/s40536-016-0032-4

Sparks, J. & **Rutkowski, D.** (2016). Exploring project sustainability: Using a multiperspectival, multidimensional approach to frame inquiry. *Development in Practice* 26.(13): 308-320. Doi: 10.1080/09614524.2016.1153041

Strietholt, R. & Rosen, M. (2016). Linking Large-Scale Reading Assessments: Measuring International Trends Over 40 Years. *Measurement* 14(1): 1-26. Doi: 10.1080/15366367.2015.1112711

Books, book chapters, and reports

* Bergem, O.K., Nilsen, T. & **Scherer, R.** (2016). Undervisningskvalitet i Matematikk. In O.K. Bergem, H. Kaarstein, & T. Nilsen (Eds.) *Vi kan lykkes i realfag – Resultater og analyser fra TIMSS 2015* (p. 120-136). Universitetsforlaget.

Blömeke, S. (2016). Der Übergang von der Schule in die Hochschule: Empirische Erkenntnisse zu mathematikbezogenen Studiengängen. In A. Hoppenbrock, R. Biehler, R. Hochmuth, & H-G. Rück (Eds.) *Lehren und Lernen von Mathematik in der Studieneingangsphase: Herausforderungen und Lösungsansätze. (=Konzepte und Studien zur Hochschuldidaktik und Lehrerbildung Mathematik)* (p.3-13). Springer

Blömeke, S. & Jenssen, L. (2016). A question of validity: Clarifying the hierarchical nature of teacher cognition. In M. Rosén, K.Y. Hansen & U. Wolff (Eds.) *Cognitive Abilities and Educational Outcomes: A Festschrift in Honour of Jan-Eric Gustafsson (Methodology of Educational Measurement and Assessment)* (p89-107). Springer

Blömeke, S. & Kaiser, G. (2016). Understanding the development of teachers' professional competencies as personally, situationally and socially determined. In Clandinin, J. D. & Husu, J. (Eds.), *International Handbook of Research on Teacher Education.* London: Sage Publishers.

*** Blömeke, S., Olsen, R.V.** & Ute, S. (2016). Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality. In T. Nilsen and J.E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (p. 21-50). Amsterdam: IEA.

*** Braeken, J.** (2016). International large-scale educational assessments: Elephants at the gate? *In Northern Lights on PISA and TALIS* (p. 195-216). Nordisk ministerråd.

Bøe, T., & **Zachrisson, H.D.** (2016). Hva betyr det for små barn å vokse opp i en fattig familie? In H.Holme, Olavsen, E.S., Valla, L., Bergum Hansen, M. *Helsestasjontjenesten. Barns psykiske helse og utvikling*. Oslo, Norway: Gyldendal Akademisk.

* Kaarstein, H., Nilsen, T., **Blömeke, S.** (2016). Lærerkompetanse. In O.K. Bergem, H. Kaarstein, & T. Nilsen (Eds.) *Vi kan lykkes i realfag - Resultater og analyser fra TIMSS 2015* (p. 97-119). Universitetsforlaget.

Lekhal, R., **Zachrisson, H.D.**, Solheim, E., Moser, T., & Drugli, M-B. (2016). *Dette vet vi om barnehagen: Betydning av kvalitet i barnehagen*. Oslo: Gyldendal/København: Dafol

* Nilsen, T., **Gustafsson, J-E.** & **Blömeke, S.** (2016). Conceptual Framework and Methodology of This Report. In T. Nilsen and J.E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (p. 1-20). Amsterdam: IEA.

* Nortvedt, G.A., **Gustafsson, J-E.** & **Lehre, A.C.** (2016). The importance of InQua for the relation between achievement in reading and mathematics. In T. Nilsen and J.E. Gustafsson (Eds.), Teacher quality, instructional quality and student outcomes (p. 97-114). Amsterdam: IEA.

Rabe-Hesketh, S. & **Skrondal, A.** (2016). Generalized linear latent and mixed modeling, In Wim J van der Linden (ed.), *Handbook of Item Response Theory, Volume One: Models.* CRC Press. ISBN 1466514310. 29. p 503 – 526.

* **Rutkowski, L.** & **Rutkowski, D.** (2016). The relation between students' perception of instructional quality and bullying victimization. In T. Nilsen and J.E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (p. 115-134). Amsterdam: IEA.

*** Scherer, R.** & Nilsen, T. (2016). The Relations Among School Climate, Instructional Quality, and Achievement Motivation in Mathematics. In T. Nilsen and J.E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes* (p. 51-80). Amsterdam: IEA.

Conference contributions

Daus, S. (January). Introduction to Item Response Theory. *Workshop*, Høyskolen i Oslo, Oslo.

Scherer, R. (January). Comments on Marcus, Kuger, & Huebener (2016): *More instructional time and the impact on student performance – Quasi-experimental evidence from PISA*. Paper presented at the first International Conference of the College for Interdisciplinary Education Research (CIDER), Berlin, Germany.

Scherer, R. & **Gustafsson, J.-E.** (March). Die psychometrische Brille der Unterrichtsforschung: Ein integrativer Ansatz zur Modellierung der Struktur und Messinvarianz von Aspekten der Instruktionsqualität [A psychometric perspective on research on teaching: An integrative approach of modeling the structure and measurement invariance of aspects of instructional quality]. Paper presented at the 4. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Berlin, Germany.

Blömeke S. & **Braeken J.** (April). Is it possible to compare teacher beliefs across countries? Annual conference of the American Educational Research Association (AERA) 2016, Washington, DC, USA.

Rutkowski, L. (April). Design considerations for planned missing auxiliary data in a latent regression context. National Council on Measurement in Education 2016 Annual Meeting. Washington, DC, USA.

Rutkowski, D., Rutkowski, L., & Wild, J. (April). The impact of poverty on US PISA achievement: A propensity score matching approach. Annual conference of the American Educational Research Association (AERA) 2016, Washington, DC, USA

Skrondal, A. (April), On the Use and Misuse of Latent Variable Scores. National Council on Measurement in Education 2016 Annual Meeting. Washington, DC, USA.

Svetina, D. & **Rutkowski, L.** (April). Measurement invariance in international large-scale assessments: Ordered-categorical outcomes in a multidimensional context. National Council on Measurement in Education 2016 Annual Meeting. Washington, DC, USA.

Braeken, J. (May). International large-scale educational assessments: Elephants at the gate? Northern Lights conference on PISA and TALIS, Copenhagen, Denmark.

Daus, S. (May). Exploring Science Content Knowledge. Symposium presentation at Competence Research in the 21st Century: Perspectives and Challenges, Berlin, Germany.

Rutkowski, D. & **Rutkowski, L.** (May). What teachers can't and can learn from ILSAs. Nordic Union Teachers Conference, Lom, Norway.

Scherer, R. (May). Educational assessments of the 21st century. Invited talk at the research

seminar of the Norwegian Reading Centre at the University of Stavanger, Norway.
Scherer, R. (May). Experiences with computer-based assessments of complex, higher-order thinking skills: The case of problem solving. Invited talk at the research seminar of the Department of Educational Technology at the University of Oslo (USIT), Norway

Rutkowski, D. & **Rutkowski, L.** (June). What teachers can't and can learn from international assessments. 2016 Nordic NLS Teacher Union Conference, Loen, Norway.

Rutkowski, L. (June). A look at the most pressing design issues in international assessment. National Academy of Education Workshop Series on International Large-Scale Assessment. Washington, DC, USA.

Scherer, R. (June). Complex problem solving – Its meaning, measurement, and potential. Talk at The Norwegian Directorate for Education and Training (UDIR), Oslo, Norway.

Zachrisson, H.D. (June). Fra barnehagebarn til elev; en god forberedelse gir barna bedre forutsetninger for å lykkes. Speech at the summer party of the Norwegian Directorate for Education and Training.

Blömeke, S. (July). Understanding mathematics teachers' competence as personally, situationally and socially influenced. Invited lecture at the 13th International Congress on Mathematics Education, Hamburg, Germany.

Niepel, C., Greiff, S., **Scherer, R.**, & Martin, R. (July). Using behavioral data from computer-generated log files to understand complex problem solving performance in a computer-based assessment. Paper presented at 31st International Congress of Psychology (ICP), Yokohama, Japan.

Tondeur, J., van Braak, J., Siddiq, F., & **Scherer, R.** (July). Effects of Preservice Training on Technological Pedagogical Content Knowledge: How Teacher Education Matters. Paper presented at the ETWC Conference, Bali, Indonesia.

Daus, S. (August). Explaining Difficulties in Large-Scale Science Assessments. Poster and presentation at European Science Education Research Association (ESERA) Summer School, České Bud□jovice, Czech Rep.

Helland, F. & **Braeken, J.** (August). Explanatory Item Response Modelling Of an Abstract Reasoning Assessment: A Case For Modern Test Design. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Olsen, R. V.; **Blömeke, S.** & Suhl, U. (August). Effects Of Teacher Quality And Instructional Quality On Student Achievement. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Scherer, R. & Nilsen, T. (August). The Relations Between Teacher Quality, Instructional Quality and Student Outcome And Their Roles In The Context Of School Climate. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Siddiq, F. & **Scherer, R.** (August). Do Age and Gender Matter for the Relation Between Teachers' Computer Self-efficacy and their Emphasis on Developing Students' Digital Skills? Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Teig, N., **Scherer, R.**, Kjærnsli, M., & **Olsen, R. V.** (August). Students' Ability to Think Scientifically: Findings from PISA 2015 and an In-depth Case Study. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Tesema, M.T. & **Braeken, J.** (August). Differential predictive validity of high school grades for future academic success as a function of opportunities to learn? Evidence from Ethiopia. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Wendt, H., Nilsen, T., Kasper, D., Van Damme, J., Bergem, O. K., Kaarstein, H., **Scherer, R.**, Hole, A., & Grønmo, L. S. (August). Assessing Instructional Quality in International Large-scale Assessments. Paper presented at the European Conference on Educational Research, Dublin, Ireland.

Daus, S. (September). Diving deeper into 'opportunity to learn' – Does the 'Opportunity to Learn' effect depend upon the content learned and the teacher? Presentation at EARLI SIGs 18 and 23 joint conference Closing the Gaps? Differential Accountability and Effectiveness as a Road to School Improvement, Oslo, Norway.

Pettersen, A & **Braeken, J.** (September). Untangling the mathematical competency demands of assessment tasks. Presentation at EARLI SIGs 18 and 23 joint conference Closing the Gaps? Differential Accountability and Effectiveness as a Road to School Improvement, Oslo, Norway.

Tesema, M.T., **Braeken, J.** & **Gustafsson, J-E** (September). Regional and gender differences in academic achievement as a function of opportunities to learn (OTL). Presentation at EARLI SIGs 18 and 23 joint conference Closing the Gaps? Differential Accountability and Effectiveness as a Road to School Improvement, Oslo, Norway.

Blömeke, S. & Jenßen, L. (October). A Question of Validity: Clarifying the Hierarchical Nature of Teacher Cognition. Challenges in Educational Measurement conference, Gothenburg, Sweden.

Blömeke S. (October) Comment on the Scandinavian Journal of Educational on the occasion of

its 60th anniversary. Perspectives on International Studies conference, Oslo, Norway.

Daus, S. (October). Explaining difficulties of science content in large-scale assessments with opportunities to learn. LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway

Tesema, M.T. (October) The role of educational opportunities in Ethiopia: Regional inequalities and Gender differences in academic achievement. LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway.

Nilsen, T. **Gustafsson, J-E** & **Scherer, R.** (October) Spotlight talks: Instructional quality. LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway

Olsen, R.V. & Frønes, T. (October). 20 år med internasjonale undersøkelser og nasjonale prøver – hva har vi lært? LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway

Paap, M.C.S., **Born, S.** & **Braeken, J.** (October) Investigating the incremental value of multidimensional over unidimensional computerized adaptive testing for average item banks in health measurement and educational testing. LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway

Paap, M.C.S, Kroeze, K.A., Terwee, C.B., Glas, C.A.W., Veldkamp, B.P & van der Palen, J. (October) Measuring Patient-reported Outcomes Adaptively: Multidimensionality Matters! Presented at the International Society of Quality of Life Research 23rd annual conference, Copenhagen, Denmark.

Paap, M.C.S, Kroeze, K.A., Terwee, C.B. & van der Palen, J. (October) Obtaining the best possible estimates of health-related quality of life in patients with COPD using computerized adaptive testing based on three PROMIS® domains. Presented at the 2016 PHO Conference: Measuring Health Outcomes Around the World, Copenhagen, Denmark.

Rutkowski, D. (October). IEA policy briefs. IEA General Assembly, Oslo, Norway.

Rutkowski, L. & **Rutkowski, D.** (October) Embracing heterogeneity. LEA seminar as part of the 20 year anniversary for the Faculty of Educational Research, University of Oslo, Norway

Scherer, R. (October). Innovative assessments of problem solving in science. Invited talk

at The Royal Society, London,

Rutkowski, D. (November). Problematizing PISA: Widening the debate about international large- scale assessments. Canadian Teacher's Association Annual Meeting, Ottawa, Canada.

Siddiq, F., & **Scherer, R.** (November). Investigating student-student interactions in an assessment of collaborative problem solving: An in-depth analysis of think-aloud protocols. Paper presented at the Annual Meeting of the Association for Educational Assessment-Europe (AEA-Europe), Limassol, Cyprus.

Teig, N., & **Scherer, R.** (November). What makes PISA items more difficult for students with minority background? Analysing the effects of item interactivity and response format in a computer-based assessment of scientific literacy. Poster presented at the Annual Meeting of the Association for Educational Assessment-Europe (AEA-Europe), Limassol, Cyprus.

Zachrisson, H.D. (December). Causal inference in non-experimental data. The case of child care and behavior problems. Keynote at the European Association for Psychiatry, division of epidemiology, biannual meeting, Gothenburg, Sweden.

Published research journalism

Andresen, Ø. (2016). – Bruk ressurser på oppfølging av nye lærere. *Forskning.no*, http://forskning.no/pedagogiske-fag-skole-og-utdanning/2016/04/bruk-ressurser-pa-oppfolging-av-nye-laerere

Gustafsson, J.-E. & Nilsen, T. (2016). Final remarks, In Trude Nilsen & Jan-Eric Gustafsson (ed.), *Teacher Quality, Instructional Quality and Student Outcome. Relationships Across Countries, Cohorts and Time*.. Springer. ISBN 978-3-319-41251-1. Chapter 7. s 135 – 147

Olsen, R. V. (2016). Mens vi venter på TIMSS og PISA. Aftenposten (morgenutg.: trykt utg.). ISSN 0804-3116. s 21- 21, http://www.aftenposten.no/meninger/debatt/Dette-kan-vi-lare-av-internasjonale-skoleundersokelser--Rolf-V-Olsen-609744b.html

Rutkowski, D. (2016). The quiet war: Challenges of educating in Afghanistan. *Pedagogisk profil.* 1, s 18- 20

Rutkowski, L. & **Rutkowski, D.** (2016). International surveys, educational ISBN: 9780470670590 DOI: 10.1002/9780470670590.wbeog949, http://onlinelibrary.wiley.com/doi/10.1002/9780470670590.wbeog949/abstract

Zachrisson, H.D., & Dearing, E. (2016). Barnehage og atferdsvansker: myter om risiko og muligheter for forebygging. forebygging.no.

CEMO

Centre for Educational Measurement

University of Oslo
Faculty of Educational Sciences
Postbox 1161 Blindern
N-0318 Oslo, Norway

postmottak@cemo.uio.no
www.uv.uio.no/cemo