1st Scandinavian Applied Measurement Conference

June 12-14, 2024

Kristianstad University, Sweden

https://www.hkr.se/samc2024/

Welcome message

In 2013, the Swedish Network for Psychometrics and Metrology in the health sciences (PMhealth) was initiated in response to a wish for a forum for discussions among former participants in a postgraduate course in psychometrics for the health sciences. The network was formally established in June 2013 in connection with a national workshop in psychometrics at Kristianstad University organized by Peter Hagell, Albert Westergren (Kristianstad University) and Kristofer Årestedt (Linnaeus University). Since then, the PMhealth workshop has been offered annually and typically featured an international guest speaker (including, e.g., Stephen. P. McKenna, Stefan Cano, William Fisher, Thomas Salzberger, and David Andrich). Since 2021 the workshop has been open for international participants. The purpose of PMhealth has been to provide a forum for collegial learning, exchange of experiences and problem-solving in psychometrics and metrology. The workshops have primarily been participant based and driven, with a program by and for the participants. In 2019, the idea of moving forward and organize a conference took form but was interrupted by the pandemic. However, following the 2022 PMhealth workshop, an organizing committee was formed to begin planning for a conference in 2024. Funding for the conference was obtained from FORTE (the Swedish Research Council for Health, Working Life and Welfare), and we are now pleased to welcome you to the first Scandinavian Applied Measurement Conference (SAMC). Similar to the PMhealth workshops, SAMC is intended to facilitate interaction and knowledge exchange among researchers, practitioners, and students. The SAMC theme is applied measurement, focusing on rating scales and other category-based measurement and assessment instruments. Papers presented at SAMC 2024 concern development, evaluation and quality assurance using Rasch measurement theory and related approaches, with an emphasis on the role of the individual person in the measurement process and analysis, as well as in the use and interpretation of results. The conference is multidisciplinary and cover fields such as the health, social, educational, behavioral, and related sciences. SAMC 2024 has a strong international presence. Besides presentations from Scandinavia, there will be presentations from three continents, including Australia, the United States, Denmark, Sweden, Norway, Germany, Norway, Latvia, the Netherlands, and the United Kingdom.

Organizing Committee:

Prof. Peter Hagell, Kristianstad University, Kristianstad, Sweden (Chair)

Prof. Karl Bang Christensen, University of Copenhagen, Copenhagen, Denmark

Dr. Jeanette Melin, Swedish Defence University, Karlstad, Sweden

Prof. Albert Westergren, Kristianstad University, Kristianstad, Sweden

Prof. Kristofer Årestedt, Linnaeus University, Kalmar, Sweden

SAMC program

Day 1 (June 12th)

	Registration is open 08.00-13.00			
09.00-12.00		Pre-conference workshop		
12.00-13.00	Lunch			
13.00-13.15	Opening			
13.15-14.45	Key note		Marais	
15.00-16.10	Session 1	Haaland-Øverby	Pedersen	Horton
16.30-17.15	Session 2	Andrich	Baghaei	

Day 2 (June 13th)

08.00-09.10	Session 3	Petersen	Pendrill	Lindqvist
09.30-10.40	Session 4	Melin	Gasser	Möllerberg
11.00-12.10	Session 5	Yngve	Andrich	Andrich
12.10-13.10		Lunch		
13.10-13.30		Poster pitch		
13.30-14.30		Poster session		
14.30-15.20	Session 6	Burgmanis	Lay	
15.40-16.50	Session 7	Van de Winckel	Horton	Scholtens

Day 3 (June 14th)

08.00-09.10	Session 8	Buchardt	Johansson	Christensen
09.30-10.15	Session 9	Hagell	Wahlkrantz	
10.30-11.15	Session 10	Yngve	Reinius	
11.15-11.30		Closing		
11.30-12.30		Lunch		
12.30-15.30		Post-conference workshop		

SAMC Keynote lecture

Keynote lecture

Rasch Measurement Theory and the measurement of change

Ida Marais, The University of Western Australia

This paper investigates two factors that affect the measurement of change in health outcomes assessment using Rasch Measurement Theory (RMT) and then shows how the measurement of change in intelligence and attainment tests can be enhanced using Rasch meta-metres of growth. The two factors considered in the assessment of health outcomes are response dependence and person-item alignment. In the case that the same questionnaire is used on multiple occasions, the potential bias in estimates because persons tend to respond in the same way to a greater degree than their parameter values imply arises (3). Andrich presented a mechanism for assessing, controlling, and eliminating this kind of bias in repeated measurements using RMT for dichotomous responses (1). In this paper a generalisation of the method is presented where items have more than two response categories. Person-item alignment at baseline impacts the effect of response dependence and the measurement of change. A set of simulation studies demonstrating these impacts is presented. Data from the Longitudinal Study of Australian children (LSAC), in particular longitudinal trajectories of mental health, are examined in light of the findings. One of the methodological challenges in measuring and comparing rates of growth is that growth is usually not linear. In Rasch Meta-Metres of Growth for Some Intelligence and Attainment Tests (2) we take advantage of an insight made by Georg Rasch, where a logarithmic transformation of the time variable, which he called a meta-metre, results in linear growth in the meta-metre. In particular, individuals get a single rate of growth which can then be used to compare rates of growth among groups using standard methods. In this paper examples are presented where the meta-metre helpfully describes growth of different phenomena, such as the development of intelligence, reading, and mathematics.

References

1. Andrich, D. (2017). Controlling response dependence in the measurement of change using the Rasch model. Statistical Methods in Medical Research, 1-17. https://doi.org/10.1177/0962280217710834
2. Andrich, D., Marais, I & Sappl, S. (2023) Rasch Meta-Metres of Growth for Some Intelligence and Attainment Tests. Springer Nature Singapore.
3. Marais, I. (2009). Response Dependence and the measurement of change. Journal of Applied Measurement, 10, 17-29.

Part I

Response dependence and person-item alignment: 35 (or 40) minutes with 15 (or 10) minutes for questions/discussion)

Part II

Rasch meta-metres of growth: 25 (or 30) minutes with 15 (or 10) minutes for questions/discussion).

SAMC Pre-conference workshop

Pre-conference workshop

Rasch analysis using open source software for reproducible science: RISEkbmRasch R package and Quarto

Magnus Johansson, PhD from RISE, Research Institutes of Sweden

This workshop walks participants through the process of importing data into R and making use of a simple code template for the different components of a Rasch analysis. The content will focus on various aspects of analysis of polytomous questionnaire data using the partial credit model. These include analyses of item fit, relative residual correlations, dimensionality, targeting, item hierarchy, differential item functioning, and reliability. Commonly used modifications of data, such as removing items and merging response categories, will be demonstrated. The workshop will both use the online package vignette as an example as well as live coding and analysis of a dataset provided by a workshop participant (or downloaded from an open data repository). Participants are encouraged to bring their own datasets to do work on during the workshop. The second part of the workshop focuses on participants analyzing their own data, and the workshop leader will provide hands-on support. The source code used and produced during the workshop will be shared openly on GitHub.com. By using Rstudio and Quarto, the process of analysis and its documentation can be kept in the same text file, which can be rendered into different output formats such as HTML, PDF and Word files. This ensures traceability of decisions and actions made during the analysis, as well as reproducibility of results, provided one has access to the datafile analyzed. Participants will need to have installed Rstudio (version v2023.12 or later) and R (version 4.1 or later), instructions are available here: https://github.com/pgmj/RstudioQuartoIntro. The R package should also be installed prior to the workshop, see https://github.com/pgmj/RISEkbmRasch for instructions. Program:

09.00-10.15	Using a template to set up a project and import data.
10.15-10.45	Break
10.45-12.00	Live coding & analysis of a new dataset, and participants can work with their own data and get hands-on support.

SAMC Post-conference workshop

Pre-conference workshop

Foundations of Rasch measurement theory and practice: applications using RUMM2030Plus Software

David Andrich and Ida Marais from the University of Western Australia
Sonia Sappl from the School Curriculum and Standards Authority, Perth, Australia

This workshop introduces the basic concepts of Rasch measurement theory through the interactive computer program application, RUMM2030+. This three-hour workshop will elaborate on structure of responses, the concept of the total score and its role in Rasch models, the item characteristic curve for dichotomous and polytomous items. Delegates will be shown how to set up a project and conduct various analyses. The wide range of analyses, none of which require writing macros, will be illustrated: these include, but are not limited to, deleting items, rescoring items, anchoring items, controlling for guessing, studying the various forms of fit and differential item functioning.
Delegates will also be given access to RUMM2030+ till the end of December 2024, together with manuals and three videos which explain setting up analyses. They are encouraged to have a look at the manual, see how to set the data up for reading into RUMM2030Plus, and perhaps even try using it, in advance. Those who may have used RUMM2030+ are welcome to also enrol and discuss topics.
Program:

13.00-14.15	Session 1: Setting up a project and conducting analyses
14.15-14.45	Break
14.45-16.00	Session 2: Various analyses that can be conducted interactively

SAMC day 1

Session 1 (Moderator: Jeanette Melin)

Mette Haaland -Øverby	Digital health literacy and Navigation health literacy in people with long-term conditions - Applied Rasch modeling for latent trait analysis
Henrik Pedersen	A comparison of the original Major Depression Inventory with a modified version: a Danish validation study
Mike Horton	Using a Rasch framework to measure the impact of kidney symptom burden - psychometric validation of the Kidney Symptom Burden Questionnaire (KSB-Q)

Session 2 (Moderator: Karl Bang Christensen)

David Andrich	Variance of an individual's rate of growth within a Rasch meta-metre
Purya Baghaei	Examining Testlet Effect in PIRLS 2021

SAMC day 2, morning

Session 3 (Moderator: Albert Westergren)

Jørgen Holm Petersen	Quantifying Rater Variation using a Rating Scale Model
Leslie Pendrill	Category-based interlaboratory comparisons: Clinical and medical examples
Andreas Nicolaidis Lindqvist	Integrated Assessment Models (IAMs) fit for purpose - what can the global modeling community learn from the measurement sciences?

Session 4 (Moderator: Karl Bang Christensen)

Jeanette Melin	Validity claims and modifications in the measurement system
Fredrik Gasser	Conceptual and empirical item hierarchies of person-centred outpatient care
Marie-Louise Möllerberg	Applying a measurement system analysis approach to dyadic data

Session 5 (Moderator: Peter Hagell)

Moa Yngve	Using RMT to evaluate interventions – an example from a school context
David Andrich	Assessing individual variation between Treatment and Placebo effects in clinical trials
David Andrich	The theoretic and empiric value of studying individual change in clinical trials

SAMC day 2, afternoon

Poster pitch session (Moderator: Albert Westergren)

Sara Alenius	Development and quality assessment of the psychometric properties of the Self-Efficacy in Lifestyle Counselling scale (SELC 20+20) using Rasch analysis
Mariusz Grzęda	Measuring Quality of Life in Huntington Disease
Martin Jarl	The Swedish version of nurse anaesthetists' non-technical skills: A Prospective Psychometric Evaluation Study of a behavioural marker system
Jeanette Melin	Perceived Safety and Perceived Unsafety: End Points of One Dimension or Two Separate Dimensions?
Antonio Filipe Macedo	Rasch analysis of an activity inventory to measure vision-related activity difficulties: a neovascular AMD specific study
Sabina Staal	Screening for dysphagia in elderly: Psychometric properties of the Danish 4 Question Test (4QT-DK) - A validation study

SAMC day 2, afternoon

Session 6 (Moderator: Karl Bang Christensen)

Ģirts Burgmanis	What the abberant resposes in the national exam reveal about the students' performance in mathematics
Dulce Lay	Identification and Evaluation of Inconsistent Responders in Large-Scale Questionnaire Data

Session 7 (Moderator: Kristofer Årestedt)

Ann Van de Winckel	Structural Validity of the Multidimensional Assessment of Interoceptive Awareness-2 with Rasch Measurement Theory
Mike Horton	Psychometric assessment and validation of the Modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) patient-reported outcome measure for Long COVID or Post-COVID syndrome
Sara Scholtens	Stigmatizing stigma scales? Cognitive interviews of items from the Community Attitudes to Mental Illness scale (CAMI) and Mental Illness Knowledge scale (MAKS)

SAMC Session 8

SAMC day 3

Session 8 (Moderator: Jeanette Melin)

Ann-Sophie Buchardt	New method for visualizing Rasch item fit using conditional item characteristic curves
Magnus Johansson	Analyzing polytomous data in R: a simulation study comparing TAM (MML) and eRm (CML) with different sample sizes and targeting properties using the partial credit model
Karl Bang Christensen	Calculating empirical critical values for Rasch item fit statistics using R

Session 9 (Moderator: Kristofer Årestedt)

Peter Hagell	Local dependence in health outcome measurement: Lessons from the 8-item Parkinson's disease questionnaire (PDQ-8)
Emelie Wahlkrantz	Lessons learned from multiple choice items: how can distractors be useful?

Session 10 (Moderator: Albert Westergren)

Moa Yngve	Differential item functioning in Work Environment Impact Questionnaire (WEIQ) subscales
Maria Reinius	Cognitive interviews guide the interpretation of a Rasch analysis of The Short Warwick–Edinburgh Mental Well-Being Scale with data from the Swedish public health survey

Session 1, 15.00-15.20

Digital health literacy and Navigation health literacy in people with long-term conditions - Applied Rasch modeling for latent trait analysis

Haaland-Øverby M., Guttersrud Ø., Helseth S., Skirbekk H.,
Søberg Finbråten H., Le C.

Introduction: Mapping Digital health literacy (DHL) and Navigation health literacy (NHL) in people with long-term conditions (LTC) is important for benchmarking and developing public health indicators. To optimize people's proficiency in managing their own health conditions and to navigate the health care services, health services should adapt to people's individual health literacy level. Eliciting knowledge about health literacy in people living with LTC, is a means of contributing to healthcare transformation in accordance with the needs of people with LTC. However, we would be more able to draw conclusions about important associations and enhance the usefulness of measurement data if we test the validity and reliability of the measurements. Applying Rasch modeling, the purpose of this study was to report on how recently developed measures performed when applied to people with LCT.
Methods: Seventeen countries including Norway took part in the European Health Literacy Population Survey (HLS19) to map health literacy. The cross-sectional study measured both general health literacy and specific literacies, such as DHL and NHL.
The current study is based on HLS19 data collected in Norway, including data from both the population sample and the subsample of respondents reporting LTC. We tested the psychometric properties of the DHL and NHL scales against the one-dimensional Rasch model for polytomous data using the statistical software Rumm2030 and R.
Results and discussion: We formulated and tested hypotheses targeted at the overall scale level and at the item level, such as hypotheses about data-model fit, differential item functioning, and the ordering of response categories. Our presentation will convey whether empirical data strengthened or weakened the hypotheses, and we will elaborate on practical implications following the results.

Session 10, 11.10-11.30

A comparison of the original Major Depression Inventory with a modified version: a Danish validation study

Pedersen H.

Introduction: The Major Depression Inventory (MDI) is a patient-reported outcome measure used by general practitioners to assist with diagnosing and evaluation of the severity of a patient's depression. However, recent studies have questioned the structural validity of the MDI.

Methods: We proposed a modified version (mMDI) of the MDI with fewer response categories and four rephrased items and aimed to compare the psychometric properties of the changes in a joint cohort of patients from general practice and mental health associations. We used Rasch analysis, confirmatory factor analysis, and the area under the receiver operating curve to assess the validity and reliability of the two versions. Equipercentile linking was used to compute cut-off points for the mMDI.

Results: For both versions, local dependence was found between the three item pairs (loss of interest, lack of energy), (lack of self-confidence, feelings of guilt), and (concentration problems, feeling restless/slowed down). The mMDI displayed lower measurement error in the upper end of the scale and better item level fit for three of the four reformulated items compared to the MDI. For the MDI, 5.3% of the respondents gave improbable responses; the corresponding number was 3.4% for the mMDI.

The mMDI displayed better fit to a one-factor model compared to the MDI.

When comparing the correlation of the scales with the WHO-5 instrument, the corresponding AUROC estimates for the mMDI and MDI were found to be 0.93 (0.92; 0.96) and 0.91 (0.87; 0.94), respectively.

The cut-off points for mild, moderate, and severe depression in the mMDI were found to be 17, 20, and 23, respectively.

Conclusion: The results suggest that the proposed changes of the MDI are psychometrically sound upgrades of the original version in terms of the structural validity.

Session 1, 15.40-16.00

Using a Rasch framework to measure the impact of kidney symptom burden - psychometric validation of the Kidney Symptom Burden Questionnaire (KSB-Q)

Horton M., Kyte D., Fletcher B., Damery S., Aiyegbusi OL., Anderson N., Bissell A., Calvert M., Cockwell P., Ferguson J., Paap M., Sidey-Gibbson C., Turner N., Verdi R., Slade A

Introduction: Measurement of symptoms that matter most to Chronic Kidney Disease (CKD) patients currently requires completion of multiple patient-reported outcome (PRO) measures. This may lead to questionnaire fatigue, lower levels of completion and resulting missing data. Moreover, many PROs used in CKD lack evidence of important measurement properties and were not developed using robust contemporary psychometric methods. The objective of this study was to use Rasch Measurement Theory to develop an item pool of symptom-specific items for CKD patients, alongside a single accessible short-form kidney symptom burden questionnaire (KSB-Q).

Methods: A survey containing patient-derived items was sent out to adults ($\geq$18years) with CKD stage 3-5, recruited from four sites in England (Birmingham, London, Sheffield, Nottingham). The survey contained 9 lead-in questions relating to specific symptoms, which all respondents completed. If an issue was identified with a specific symptom, then a further set of items were administered, all related to the corresponding symptom. Rasch analysis was used to assess the psychometric properties of the lead-in questions as a symptom burden scale, along with each separate set of symptom-specific items.

Results: A total of 1,464 item pool surveys were posted to patients, and 419 participants returned questionnaires (29% response rate). The sample included 60% male respondents; 70% were white, with 26% reporting other ethnic backgrounds. Rasch analysis indicated that the lead-in items representing 9 key symptom areas (fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs and shortness of breath) formed a fitting (chi-square p=0.12), well-targeted (PSI=0.80, Cronbach's alpha=0.87), unidimensional (2.3% significant p=0.05 t-tests), reliable (ICC 0.82, 95% CI 0.69-0.89) measure of somatic kidney symptom burden. However, a potential issue was identified with the response structure, although the impact of this was shown to be minor. Analysis of the symptom-specific sets was limited, due to small sample sizes.

Discussion: The KSB-Q represents a short, accessible, symptom PRO with evidence of strong psychometric properties. Further data collection is necessary to assess the symptom-specific sets of items. Future work is planned around building a CAT application of the symptom-specific item sets, with administration tailored by responses to the KSB-Q.

SAMC Session 2

Session 2, 16.30-16.50

Variance of an individual's rate of growth within a Rasch meta-metre

Andrich D., Sappl S.

Introduction: Andrich, Marais and Sappl (2023) adapted Rasch's concept of a meta-metre, which governs the rate of growth of all individuals of a group, to characterise linear growth on some intelligence and attainment tests. Within the meta-metre, individual differences in the rate of linear growth are also characterised. Comparisons among means of groups can be made with analysis of variance procedures where the variance among individuals absorbs the error variance of the estimate of each individual.

Methods: To characterise growth, an individual's measurements at different time points are required. If responses to the items of a test are available and conform to the relevant Rasch model for measurement, then each measurement has an error estimate. It is possible to write an equation for the variance of the rate of growth of an individual which is a function of the error variance of each measurement. This suggests that an estimate of the proportion of variance in the rate of growth of each individual that can be attributed to measurement error can be made.

Simulated responses over multiple time points, and analysed with the Rasch measurement model, are used to evaluate the estimate for the variance of the rate of growth attributable to measurement error.

Results: It appears that the estimate of the proportion of variance of each individual's rate of growth which can be attributed to measurement error is relatively conservative. Further work might result in a correction factor for the estimate of the variance of the rate of growth.

Discussion: Having an estimate of the variance in an individual's rate of growth which is attributable to measurement error raises the possibility of testing the degree to which an individual's rate of growth is different from some hypothesised or required rate of growth. This knowledge may be helpful in educational settings where targeted growth might be set on an individual basis.

1. Andrich, D., Marais, I. and Sappl, S. (2023). Rasch Meta-Metres of Growth for Some Intelligence and Attainment Tests. Springer: Singapore.

SAMC Session 2

Session 2, 16.50-17.20

Examining Testlet Effect in PIRLS 2021

Baghaei P., Strietholt R., Johansson S., Rosén M.

Introduction: Testlets are subsets of items clustered around the same theme or stimulus such as a passage or a diagram. Although testlets improve test efficiency, they may lead to the violation of the local independence assumption of item response theory models. Violation of local independence could introduce bias in parameter estimates and result in spurious high reliability coefficients. The purpose of this study is to examine testlet effect in PIRLS 2021 (Progress in International Reading Literacy Study).

Method: The digital version of PIRLS 2021 is structured around 18 independent passages where 11 to 18 items are presented for each passage. We fitted the Rasch testlet model to PIRLS 2021 digital test and paper test for six different countries separately.

Results: Our findings showed that the Rasch testlet model fits better than the dichotomous Rasch model where local item dependence is ignored. Examination of testlet effects showed that some items introduce substantial testlet effects. Our findings further confirmed that when local item dependence is ignored, reliability increases which is an artefact of the testlet structure.

Discussion: The consequences of testlet effect for parameter estimation and measurement precision are discussed.

SAMC Session 3

Session 3, 8.00-8.20

Quantifying Rater Variation using a Rating Scale Model

Petersen J.H.

A model-based approach to the analysis of agreement between different raters is presented. The model is relevant in a situation where a number of raters have supplied ordinal ratings of the same cases in a sample.

The ordinal rating are analysed in a Rating Scale Model, that is an ordinal regression model with random rater effects - allowing raters to have different propensities to score a given set of individuals more or less positively. The model also includes case-specific parameters that allows each case his or hers own level (disease severity). It is assumed that no 'gold standard' is available. Based on the model, the rater variation is quantified using a median odds ratio measure. This allows expressing the rater variation on the same scale as the observed ordinal data.

An important example that will serve to motivate and illustrate the proposed model is the study of breast cancer diagnosis based on screening mammograms. The purpose of the assessment is to detect early breast cancer in order to obtain improved cancer survival. In the study, mammograms from 148 women were evaluated by 110 expert radiologists. The experts were asked to rate each mammogram on a five-point scale ranging from "normal"to "probably malignant".

SAMC Session 3

Session 3, 8.20.-8.40

Category-based interlaboratory comparisons: Clinical and medical examples

Pendrill L., Korsell N.

Introduction: Interlaboratory comparisons (ILCs) – where typically an object is circulated by a pilot laboratory among a number of participating laboratories – play a key role in assuring the quality of measurements in metrology, which in turn supports quality assurance of products and processes in many areas. While established in physical metrology (Koepke et al., 2017), ILCs are less well developed where measurement system responses are on the less-quantitative ordinal and nominal scales typical of certain parts of chemical and materials metrology and in the human and social sciences (Bashkansky & Turetsky, 2016).

Methods: To our knowledge, modern measurement theory – including Rasch transformation of raw scores and measurement system analysis – has only rarely been applied to ILCs. This presentation will consider two ILC case studies of performance data from: (i) carotid artery stenosis meta-analyses of different surgical interventions (Koepke et al., 2017), (ii) pregnancy testingin a recent proposal to modernise the classifier performance tool called the Receiver Operating Characteristic (ROC)(L. R. Pendrill et al., 2023), in order to illustrate our new approach to the analysis of ILCs.

Results: Provided one applies a modernised approach combining MSA and Rasch modelling, ILCs for the less-quantitative ordinal and nominal scales should be closely analogous to traditional ILCs in physical metrology. The "circulating objects"in an ILC have quantities attributed to them (which if known can act as metrological references, for instance “task difficulty”), thus revealing a certain (finite) performance ability of each “laboratory” (or meter or person responding to a memory test item). Our presentation will show for instance how earlier ILC analyses using Forest plots (Koepke et al., 2017), can be modernised by applying the Rasch model.

Discussion: Traditional statistics deployed in ILCs physical metrology – such as key comparison reference value (KCRV) and degree of equivalence (DoE) – can, in principle, also be applied to category-based observations, but with the important requirement that modern measurement theory has beenapplied.This work is part of an on-going effort to prepare the ground for automatic machine-readable of ILC results.

1. Bashkansky, Turetsky, V. (2016), https://doi.org/10.1007/s00769-016-1208-x
2. Koepke, et al., (2017). https://doi.org/10.1088/1681-7575/aa6c0e
3. Pendrill, et al. (2023). https://doi.org/10.3390/a16050253

SAMC Session 3

Session 3, 8.40-9.00

Integrated Assessment Models (IAMs) fit for purpose -what can the global modeling community learn from the measurement sciences?

Lindqvist A.N., Cornell S., Palm C., Collste D.

Integrated assessment models (IAMs) link socio-economic features with climate and biophysical processes into one modelling framework. They are prominent tools in the global science-policy interface and play a vital role in IPCC, IPBES, and 2030 Agenda-related processes. In order to effectively support policy and decision making, IAMs need to be "fit-for-purpose". Multiple frameworks have been developed to guide model development and to evaluate model fitness. However, these frameworks often lack both a clear definition of what model fitness entails and robust ways of measuring it. For instance, most contemporary frameworks focus largely on what is "in" the model, or what steps were taken when building it, rather than measuring the qualities that constitutes fitness from a user-centered perspective. This is problematic because it does not facilitate robust assessment of fitness-for-purpose and, in turn, it increases the risk of resources being spent on the development of models that are not useful.

We propose using methods from the metrological sciences to derive a robust measurement tool for assessing IAM fitness. In this conference contribution, we present in-progress research towards this end. We use literature studies and user-surveys to develop construct maps (Wilson 2005) of five key criteria for IAM fitness -salience, accessibility, credibility, legitimacy and feasibility- with four to eight likert-type items each. We will collect data from 30-40 users (raters) of global-scale IAMs then analyze data-model fit to Rasch measurement theory.

The limited sample size aside, drawing on ongoing experience in the EU funded WorldTrans project (https://worldtranseu.wpcomstaging.com/about-us/), we suggest that the integration of metrological standards into IAM design and evaluation phases can improve the utility and impact of these models in the science-policy interface. Furthermore, we see great scientific value in exploring the untapped potential learnings and synergies between the metrological sciences and the integrated assessment modeling domain. We invite for constructive discussions around methodological choices and further development of the presented research endeavor.

SAMC Session 4

Session 4, 9.30-9.50

Validity claims and modifications in the measurement system

Melin J.

Introduction: Imagine two tests with satisfactory fit to the Rasch model claiming to measure the same latent trait of persons, but with different sets of items, and when they are being correlated to another variable, y, the degree of associations is different. This raises questions about making a valid inference about the association between y and the latent trait of interest. This presentation aims to discuss validity claims based on variations in the measurement system.

Methods: Theoretical cases and empirical data will illustrate the role of modifications in the measurement system, such as item composition, construct definitions, and measurement models.

Results: Variations in item composition, construct definitions, and measurement models can individually and together – to some degree – increase or decrease the correlation coefficient between two variables in linear regressions. This limits the comparability between studies and threatens valid inferences about the true associations between two variables. Furthermore, a higher correlation coefficient might not always be better and imply a more valid result.

Discussion: The results presented are not novel but, unfortunately, too often overlooked and may confuse inferences about the association between y and a latent trait of interest. The practical consequences will likely differ from case to case, which warrants careful consideration. Researchers must take a stronger responsibility in shedding light on such issues and advance methods to ensure valid interpretations of associations and comparability.

SAMC Session 4

Session 4, 9.50-10.10

Conceptual and empirical item hierarchies of person-centred outpatient care

Gasser F., Westergren A., Bala S-B., Ekstrand J., Hagell P.

Introduction: The Person-Centred Care instrument for outpatient care (PCCoc) is a 36-item patient-reported experience measure with 4 ordered response categories, that aims to capture the degree of perceived person-centred care (PCC) from a patient perspective among persons with long-term conditions. The PCCoc is based on a framework that conceptualises outpatient PCC from lower to higher levels of perceived PCC, from personalization via shared decision-making to empowerment, where 35 of the PCCoc items are a part of the framework´s hierarchy. This study investigates to what extent empirical item responses are consistent with the hierarchical PCCoc conceptual framework among persons with long-term conditions in outpatient care.

Methods: PCCoc data (322 responses) from persons with long-term psychiatric, cardiological, rheumatological or neurological conditions were analysed. The Rasch measurement model (RMM) was used to evaluate model fit and the empirical item ordering. Correspondence between the empirical and conceptually expected item hierarchies was assessed graphically and using the polyserial correlation between RMM derived item locations and their a-priori expected rank order.

Result: Two items showed clear misfit to the RMM (fit residuals >4.9). The polyserial correlation between empirical item locations and the expected rank order using all 35 PCCoc items was 0.64; after removing the 2 misfitting items it was 0.71. In addition, subtests (i.e., testlets consisting of a combination of all items belonging to the respective hierarchical domain) were created to account for any local dependency. Subtest locations on the hierarchical continuum indicated good correspondence between empirical data and the conceptual hierarchy, when based on 35 as well as 33 items. Both subtests had a polyserial correlation of 0.99 between testlet locations and the expected rank order.

Conclusion: The observed correspondence between empirical data and the conceptual framework indicates that the PCCoc reflects the underlying framework, and therefore can be a valuable instrument to support targeted PCC-promoting interventions.

SAMC Session 4

Session 4, 10.10-10.30

Applying a measurement system analysis approach to dyadic data

Möllerberg M-L., Årestedt K., Hagell P., Melin J.

Introduction: As families consist of at least two parts, dyadic data collections are common in family research. However, nested observations within dyads (e.g., two family members) cannot be assumed to be mutually independent as they share a common context, consequently, such violation of the independence assumption can create a bias in measurements. Therefore, the objective of this study is to present two measurement system approaches to measuring family sense of coherence, i.e., from (a) dyadic- and (b) single-informant perspectives using the 12-item Family Sense Of Coherence scale (FSOC-S12).

Methods: The FSOC-S12 is constructed as a unidimensional scale, including 12 items representing the three core components of sense of coherence (i.e., comprehensibility, manageability, meaningfulness). A secondary analysis was conducted on 151 dyads, including one patient and one family member. Rasch analyses were applied to racked (a dyadic-informant perspective) and stacked (a single informant perspective) data set-ups using Winsteps.

Results: Both set-ups showed minor deviations from the Rasch model according to fit statistics. However, most items had disordered thresholds and some problems with local dependency. Item hierarchies were similar in both set-ups and there was no differential item functioning (DIF) by role from the dyadic informant perspective, but four items showed DIF by informant role in the single-informant perspective.

Discussion: This study provides important insights into measurements in family research and alike where both dyadic- and single informant perspectives may be of interest. Depending on the clinical or research question, different set ups may be used to provide meaningful measures of family ability as a whole (dyadic-informant perspective) or measures of patient's and family member's separate views of the family ability as a whole (single-informant perspective). However, it further raises questions such as what the measured construct is in the different perspectives, weather a single-informant perspective is enough to measure family as a whole, and if more than two informants are needed to measure the family as a whole.

SAMC Session 5

Session 5, 11.00-11.20

Using RMT to evaluate interventions – an example from a school context

Yngve, M., Ekbladh E.

Introduction: Rasch Measurement Theory (RMT) transforms categorical measures into interval measurements at the person level, compensating for limitations of quasi-experimental designs and categorical measurements to generate reliable values of change in outcome between assessment time-points when evaluating an intervention. This presentation aims to exemplify how RMT can be used to evaluate an intervention within a quasi-experimental study design using an ordinal outcome measure.

Methods: Data consisted of 300 secondary school students' assessments with the School Setting Interview (SSI), before and after students received an individualized information and communication technology intervention. SSI measures 'student-environment fit' representing the students' perceived need for support in school activities based on the agreement between personal abilities and environmental demands of school activities. The 16 items of different school activities are rated on a four-step rating scale ranging from unfit to perfect fit.

Results: Ordinal SSI data was analysed in the computer software RUMM2030, organized in a stacked (by assessment time-point) format. The change in student-environment fit on an individual level was examined by calculating whether the students' follow-up measure fell outside the 95% confidence interval of their baseline measure. The individual change was calculated as the student's follow-up measure minus the baseline measure, divided by the standard error of the change. 'Significance of change' was categorized into five groups: Significant improvement = $\geq +1.96$; Non-significant improvement = $0<$Sig. change $\leq +1.95$; No change = Sig. change = 0; Non-significant worsening = $1.95\leq$Sig. change <0 and Significant worsening = Sig. change = ≤ -1.96.

Between measurement time points, the majority of students (84%, n=251) demonstrated improvement in student-environment fit, where 22% (n=66) achieved Significant improvement. Worsening in student-environment fit was noted in 16% (n=47), where 1% (n=4) showed Significant worsening.

Discussion: The one-group pretest-post test design provides limited evidence when evaluating interventions. Employing RMT ensured consistency in measuring student-environment fit before and after the intervention, offering reliable and stable individual measures across time points. The investigation of change on individual level between assessment time points provides opportunities to reliable and valid evaluations of intervention effect when categorical measurements are used.

Session 5, 11.20-11.40

Assessing individual variation between Treatment and Placebo effects in clinical trials

Hobart J., Andrich D., Marais I.

Introduction: In clinical trials individuals are expected to vary in their change. It is possible that the Treatment has a variable change among individuals even though there is no change in means. Assessing the statistical significance of individual change, and comparison between Treatment and Placebo, is possible if each measurement has a known standard error. Such errors are provided when the measuring instrument has multiple items, such as in patient reported outcomes (PROs) and analysed by methods of modern test theory.

Methods: The data analysed are simulated. They in part parallel real data from a clinical trial. The mean change between two times in Treatment and Placebo groups was the same, as were all other features, except that the variation in the change in the Treatment group was smaller, with the correlation between the two times in the Treatment and Placebo groups being respectively of the order of 0.8 and 0.5 respectively. Responses to 20 items, each with four categories, which fit the Rasch model were simulated, and the data analysed using the same model.

Results: As simulated, the Treatment and Placebo groups (both) respectively showed no mean change at the 1% level. Simultaneously, 26.00% and 17.20% of the Placebo and Treatment groups respectively showed a significant deterioration at the 5% level. The difference in these percentages is statistically significant at the 1% level.

Discussion: The substantive example that provided the context for the simulation is that of a clinical trial for treatment of Multiple Sclerosis (MS) where the status quo is deterioration in functioning, and the Treatment is intended to slow the deterioration. The manifestation of difference in individual change between a Treatment and Placebo groups is in the correlation between the measurements at the two time points. In the context of deterioration as the status quo, the higher this correlation the greater the benefit. The implication of this illustration is that in clinical trial data, it is critical to examine changes at the individual level. They are relevant for any clinical trial in which individual errors of measurement are available.

SAMC Session 5

Session 5, 11.40-12.00

The theoretic and empiric value of studying individual change in clinical trials

Hobart J., Andrich D., Marais I.

Introduction: In clinical trials individuals are expected to vary in their measured change. This variation can camouflage the significance of mean change. This effect is illustrated with the t test for dependent samples between two assessments, and then a method for assessing individual change, unconfounded by individual variation, is presented. Assessing the statistical significance of individual change is possible if each measurement has a known standard error. Such errors are provided when the measuring instrument is composed of multiple items and analysed by methods of modern test theory.
Methods: The data analysed are from a pivotal phase III randomised clinical trial of people with multiple sclerosis (MS) using a patient reported outcome (PRO) measure of physical functioning. Responses were analysed using Rasch measurement theory methods to provide each person with a measurement and its standard error at two times.
Results: After one year, the Treatment group ( ) showed a marginally significant mean deterioration, between the 5% and 1% levels. The Placebo group, of the same size, showed a greater mean deterioration at the 5% level). Simultaneously, 20.1% and 27.2% of the Treatment and Placebo groups respectively showed a significant deterioration at the 5% level, and the difference in percentages is statistically significant at the 5% level.

Discussion: In traditional methods of assessing change in means, the reference for the amount of variation due to chance is the variation among individuals. However, this variation includes the errors of measurement of each individual. In PRO measurement, where each individual measurement has an estimate of its standard error, it is possible to separate these error components, and assess the significance of change independently of the variation among individuals. This assessment in change is based on a comparison in the proportion of persons who show a change in the relevant direction rather than the change in the means. The clear implication of this illustrative study is that when analysing clinical trial PRO measure data it is critical to examine changes at the individual level. They are relevant for any clinical trial in which individual errors of measurement are available.

Poster pitch session

Development and quality assessment of the psychometric properties of the Self-Efficacy in Lifestyle Counselling scale (SELC 20+20) using Rasch analysis

Alenius S., Westergren, A., Nilsson, P., Nilsson, M., Rask, M., Behm, L.

Background: Globally as well as in Sweden, diseases that are caused by unhealthy lifestyle habits are the most common causes of death and disability. Despite guidelines that obligate all health care professionals to counsel patients about lifestyle, studies have shown that this is not prioritized within health care. One reason for this omission, among nurses, is a lack of confidence in knowledge and counselling skills. This study aimed to develop and provide a quality assessment of the psychometric properties of an instrument to measure self-efficacy in lifestyle counselling.

Methods: A measurement instrument, inspired by an existing American instrument following Bandura's recommendations for the development of self-efficacy measures, was developed according to Swedish national guidelines for disease prevention. The instrument was revised after 18 kognitive interviews with nurses, student nurses and clinical experts and was administered to 310 nursing students at different education levels. The instrument was tested with Rasch model analysis with a focus on dimensionality, local dependency, targeting, reliability, response category functioning, Rasch model fit, and differential item functioning by age, gender, educational level and previous health care education.

Results: The development of the instrument resulted in 20+20 items, 20 about self-efficacy in knowledge and 20 about self-efficacy in the ability to counsel persons about their lifestyle. The analyses showed that knowledge and ability are two different, but related constructs and that ability is more demanding than knowledge. The findings indicate (considering dimensionality and local dependency) that all 20 items within the knowledge construct as well as the 20 items within the ability construct can be summed to obtain two separate but related total scores, where knowledge (reliability 0.91) is a prerequisite for ability (reliability 0.93). The items represented lower self-efficacy than reported by the respondents. The response categories functioned as expected, the Rasch model fit was acceptable, and there was no differential item functioning.

Conclusions: The SELC 20+20 was found to be easy to understand with an acceptable respondent burden, and the instrument showed good measurement properties. The instrument can be a useful tool in professional education of health care personnel and clinically in health care.

SAMC Poster pitch session

Poster pitch session

Measuring Quality of Life in Huntington Disease

Grzęda M., Hollings S.; Johnstone E., Spray I.

Introduction: Huntington's Disease (HD) is a rare, inherited disease that causes progressive breakdown (degeneration) of nerve cells in the brain which affects motor control, cognition and behaviour of people living with this condition. It has been recognised that HD has a huge impact on daily life not only for patients but also for their relatives, and friends. Specifically, literature indicates that Quality of Life (QoL) is a crucial consideration for individuals with HD.It is a widely accepted view that the disease poses a unique challenge to healthcare providers, making it essential to address various aspects that contribute to enhanced QoL.

The nature and the progression of the disease imposes several methodological challenges related to robust measurement of the QoL of people with this disease. The main challenges include the subjective nature of the QoL construct and cognitive symptoms that affect the completion of questionnaires. One solution that could solve this problem is using a proxy-reported measure. By utilizing proxy-reported measures, researchers and caregivers can continue measuring QoL of HD individuals, even if they can no longer complete the questionnaire themselves. Due to the advantages of the Rasch measurement theory, results obtained from the proxy-reported measures can be compared with results obtained from self-reported measures which allows to track the state of the patients even when they no longer have the ability to complete a the questionnaire themselves. The current paper presents the process of development and validation of the self-reported and proxy-reported measures of QoL for people living with HD.

Methods: The 49-item draft questionnaire in a self-reported and proxy-reported format will be tested with 250 respondents recruited from the UK, Ireland, Germany, Czechia and Italy. The data set once collected will be subjected to Rasch analysis. Once the items are validated, the common item equating Rasch methodology will be applied to obtain the comparable results between self-reported and proxy-reported measures. The cross walk table for the proxy-reported measure to the equivalent self-reported measure will be created.

Results/Discussion: The data and the results will be available on the date of conference. The paper will present methodological challenges associated with measuring QoL in HD-patients.

SAMC Poster pitch session

Poster pitch session

The Swedish version of nurse anaesthetists' non-technical skills: A Prospective Psychometric Evaluation Study of a behavioural marker system

Jarl M., Cecilia Escher C., Harbut P., Helen Conte, Ulrica Nilsson

Introduction: Non-technical skills are the essential cognitive, social, and personal resource skills that contribute to safe and efficient task performance. Behavioral rating systems give an opportunity to train and evaluate non-technical skills. A behavioral rating system to evaluate the performance of nurse anesthetists' non-technical skills exists in Denmark and Norway but not in Sweden.

Aim: The study aims to translate and adapt nurse anesthetists' non-technical skills (Danish version) into Swedish and test its psychometric qualities among Swedish nurse anesthetists who have experience tutoring nurse anesthetist students and junior nurse anesthetists.

Design: This study is a prospective psychometric evaluation that includes translating, evaluating face validity, and testing the psychometric qualities.

Methods: Twelve short video clips of different situations in anesthesia were recorded in a simulated environment. Before rating the video clips, the nurse anesthetists (n=16) underwent a three-hour workshop. The nurse anesthetists rated all video clips, and after approximately four weeks, a test-retest was conducted, including five video clips. The ratings were analyzed using Cronbach alpha and intraclass correlation coefficient.

Results: Internal consistency shows an acceptable result on the element level, and Interrater reliability shows a good result. Retest reliability shows poor to moderate reliability. There is a substantial variance in the use of "Not Relevant" by the raters, the length of the video clip, and the provider that was rated.

Discussion: This initial psychometric testing shows uncertainties when rating short video clips, rating of the secondary provider, and affected by individual raters due to the high amount of rating "Not Relevant". Even so, there is good stability, inter-rater reliability, and poor to moderate test-retest reliability.

SAMC Poster pitch session

Poster pitch session

Perceived Safety and Perceived Unsafety: End Points of One Dimension or Two Separate Dimensions?

Karlsson S., Eskil M., Elefalk K., Melin J.

Introduction: Perceived safety is a multifaceted concept that has evolved in meaning over time and across different contexts and regions. Although the variation in conceptual meaning may be a natural consequence of societal changes, it poses challenges when addressing safety-related issues. This three-year research project aims to advance the practice of measuring safety perception through applying quality-assured measurements. In this study, the first of three parts, we aim to compile existing surveys on the topic in Sweden and analyze definitions, content, and measurement practices used in these surveys.

Methods: A scoping review was conducted to compile existing surveys on safety perceptions in Sweden, followed by a content analysis to categorize measurands, item content and compositions, and measurement models. Furthermore, similarities and differences between surveys and contexts were identified.

Results: Almost all surveys identified lack a clear definition of perceived safety. Many of them comprise items and scales for perceived unsafety. For instance, surveys from municipalities and authorities often address fear of crime. On the contrary, surveys in schools and workplaces often include more positively framed feelings about perceived safety. Some aspects, such as psychological safety or worries and threats, may apply across different contexts. Overall, there were vaguely defined measurement models of safety perceptions.

Discussion: Existing surveys often consider perceived safety and unsafety as the same construct. However, closer inspection of item content shows more differences than only positive or negative framing of shared content. Through the analyses, theoretical construct theories of what it means to go from low to high for perceived safety and perceived unsafety, respectively, have been identified. These theories will be further investigated utilizing Rasch analysis of existing empirical data. In turn, we will propose and test revisions on existing scales and develop new items to establish an enhanced measurement system.

Poster pitch session

Rasch analysis of an activity inventory to measure vision-related activity difficulties: a neovascular AMD specific study

Macedo AF., Nilsson I., Svanfeldt C., Mohlin C., Baskaran K.

Introduction: Our aim was to assess vision-related activity difficulties (visual ability) among patients with neovascular AMD using a Swedish version of the Massof Activity Inventory (MAI).

Methods: Participants were diagnosed with neovascular age-related macular degeneration (nAMD) and receiving treatment for the disease in a hospital in Sweden. Participants completed the Swedish version of the MAI questionnaire, best corrected distance and near visual acuity (VA) were measured. MAI is formed of 50 goals (items) and can be used to measure the overall or specific visual ability in 4 functional domains: reading, mobility, visual motor function, and visual information processing. Charlson comorbidity index weighted by age was computed.

Results: There was a total of 196 participants (mean age=78.5 years, SD=7.67, 66% female), 67 with vision impairment defined as distance VA in the better eye 0.32 logMAR or worse. Among the visually impaired, median VA was 0.53 logMAR (IQR=0.22). The median Charlson comorbidity index weighted by age was 4, range 0 to 15. After excluding participants with VA better than 0.32 logMAR, 12 items were excluded from the final Rasch analysis due to poor fitting. The mean visual ability for the visually impaired participants was 1.64 logits (SD= 1.55). The most difficult item was "sew and needlework"and the less difficult was "eat meals", items removed included, for example, "fishing"or "do electrical work". Visual ability was correlated with distance VA (p=0.043), near VA (p=0.01) and Charlson comorbidity index (p=0.026).

Discussion: The results revealed that the MAI produces reliable measures of visual ability among patients with nAMD and can be used for vision rehabilitation outcomes. Some items in the Swedish version of the MAI seem redundant for patients with nAMD. Shorter instruments are always preferrable, and we will investigate further the possibility of an item reduction for this target group.

Poster pitch session

Screening for dysphagia in elderly: Psychometric properties of the Danish 4 Question Test (4QT-DK) - A validation study

Staal S., Christensen KB., Smithard D., Westergren A., Melgaard D.

Introduction: Dysphagia refers to difficulty or impairments in swallowing and results in unsafe, ineffective, or uncomfortable swallowing. Unidentified dysphagia can result in aspiration, malnutrition, dehydration, decrease in quality of life and increased risk of mortality. Among the elderly, prevalence rates range from 15% to 50% and screening in elderly is recommended.

The 4-item questionnaire test (4QT) is a simple screening measure of dysphagia for the elderly. A positive answer to any item indicates the need for further assessment. 4QT is fast, simple to apply, and requires no training beforehand. However, it is not translated to Danish and the psychometric properties in a Danish elderly population are unknown.

The aim was to translate 4QT to Danish, determine criterion validity according to the existing measures Minimal Eating Observation Form II (MEOF-II) and Volume Viscosity Swallow Test (V-VST) and determine construct- and structural validity and reliability by exploring whether 4QT fits the assumptions of the Rasch model

Methods: The 4QT was translated and back-translated. A total of 73 participants 65 years were included and screened with the 4QT-DK and assessed using V-VST and MEOF-II. Criterion validity was assessed compared to V-VST, and MEOF-II according to sensitivity, specificity, and predictive values. Construct- and structural validity were examined using Confirmatory Factor Analysis and Rasch model analysis investigating item- and person fit, Differential Item Functioning (DIF) for sex, age, primary disease, and Local Dependency (LD). Reliability was assessed using Cronbachs coefficient alpha.

Results: For criterion validity, the 4QT-DK showed high sensitivity (84%, and 90% with V-VST and MEOF-II respectively). Specificity was as expected lower (36%, and 42%). Analyses confirmed the psychometric validity of the 4QT-DK. Reliability was low (Cronbachs Alpha=0.58) due to number of items.

Conclusion: 4QT-DK is a valid and sensitive measure for screening elderly for dysphagia. Further studies are required to assess the psychometric properties and confirm findings in a larger Danish sample.

SAMC Session 6

Session 6, 14.30-14.50

What the abberant resposes in the national exam reveal about the students' performance in mathematics

Burgmanis Ģ., Mikite M.

Introduction: In educational research abberant responses in measurement and assessment of learning is seen as an outliers that distort reliability of instrument and ability to make conclusions on teaching and learning. However, we believe that robust analysis of outliers and combing Rasch person-fit statistics and qualitative analysis of assessment can make learning visible and benefit to better understanding what student can do, what can not do and to find potential learning gaps. The purpose of this study is to provide an overview of method for revealing potential causes of students' unexpected performance in national exam of mathematics.

Methods: This study use the dataset obtained from the National Level Examination in Mathematics for grade 9 in Latvia (N=17962). First, person-fit statistics of the Partial Credit extension of the Rasch measurement model were used (person-fit values <0.5 and >1.5) to detect aberrant responses and distinguish misfitting persons (N=1908). Second, 31 out of 51 items where students displayed abberant responses were examined by four experts and grouped into six categories, based on potential reasons for mismatched answers. Finally, based on the frequency of abberant responses and their expectancy level answering on particular group of items misfitting students were profiled in three groups and causes for misfit for each profile were identified.

Results: At this phase of the study, we found three profiles of misfiting students: (1) students who gave better answers than expected because they learned only specific algorithms used in recurring constructed response items, (2) students with inconsistent performance, providing both better and worse responses, likely due to guessing from a lack of fundamental math skills, (3) students who gave more incorrect responses than expected, indicating a lack of understanding of the mathematical situation.

Discussion: Our study contribute to the idea that unexpected answers in assessments may not always result from guessing or the student's attitude. We found that combining person-fit statistics from the Rasch model with qualitative item analysis is beneficial for understanding why students consistently perform better or worse than expected in math assessments. This approach can help in making data-driven decisions at the individual level how to improve learning.

SAMC Session 6

Session 6, 14.50-15.10

Identification and Evaluation of Inconsistent Responders in Large-Scale Questionnaire Data

Lay D., Friedman T.

Inconsistent responders in questionnaires from large-scale surveys are those who are unlikely to be responding genuinely based on their contradictory responses to both positive and negatively worded items. A previous cross-national study has identified up to one-third of students as inconsistent responders across different educational systems (see Steinmann, Sanchez, van Laar, Braeken, 2021). Such high levels of dubious responding raises concerns about validity of the underlying scales. Thus, it is desirable to develop strategies for identifying and evaluating inconsistent responses in educational research. The aims for the current research are to use large-scale survey data to: 1) estimate the proportion of students who are classified as "inconsistent responders"; 2) identify the characteristics of those students; 3) investigate the extent in which adjusting inferential models are impacted by removing those respondents.

Our study uses data from the International Civic and Citizenship Education Study (ICCS) 2022 cycle, which collected data from 24 educational systems on student's civic knowledge as well as their engagement and attitudes towards aspects of civics and citizenship. In our research, we selected a set of mixed-worded statements designed to measure students' attitudes towards gender equality. Students were asked to indicate their level of agreement using a four-point likert scale ("Strongly agree" to "Strongly disagree") to six statements. Three of these statements were positively worded and three were negatively worded. Using Item Response Theory, ICCS reports a Weighted Likelihood Estimate (WLE) derived from aggregate responses (negative worded items were reverse coded).

Preliminary findings reveal inconsistent responders accounted for, on average, over 14% of students. These students exhibited lower scores on civic knowledge, with considerable variation across educational systems. Inconsistent responders displayed a propensity towards strongly agreeing to all items, rather than strongly disagreeing. Further analyses will explore the characteristics of students within this sub-group and identify the potential benefits to inferential models by removing this group of students.

The results are important for identifying students likely to provide inconsistent responses in order to target interventions to encourage accurate responding, as well as enhancing the reliability and validity of questionnaire data in international large-scale surveys.

Session 7, 16.00-16.20

Structural Validity of the Multidimensional Assessment of Interoceptive Awareness-2 with Rasch Measurement Theory

Van de Winckel A., Carpentier S.

Introduction: The Multidimensional Assessment of Interoceptive Awareness, Version-2 (MAIA-2) is a patient-reported outcome measure, evaluating interoceptive sensibility (i.e., the ability to be aware of bodily signals). The MAIA-2 has 8 dimensions underlying interoceptive sensibility: noticing, not distracting, not worrying, attention regulation, emotional awareness, self-regulation, body listening, and trusting.

Most studies reported on MAIA-2 in healthy adults, adults with mental disorders (9 studies), or with chronic pain (2 studies). Prior studies reported on internal consistency (Cronbach's alpha: 0.64-0.93 for the 8 dimensions); test-retest reliability (ICC=0.81), and structural validity with exploratory or confirmatory factor analysis. The purpose of this study was to perform Rasch Measurement Theory analysis on the MAIA-2 in healthy adults.

Methods: English-speaking community-dwelling healthy adults were recruited at the Minnesota State Fair and Highland Fest. We collected demographic and lifestyle information, as well as the MAIA-2 survey on an iPad. The MAIA-2 has 37 items with scores ranging from 0 (never) to 5 (always), with undefined labels for scoring options 1-4. A higher score on MAIA-2 is associated with greater ability in interoceptive sensibility.

Results: We recruited 556 participants, 51±17 years old (range 18-86 years), 64.0% women, 1.4% Hispanic, 88.8% White, and 11.2% of diverse race. Item fit was obtained after rescoring 4 items and deleting 18 items. Misfit was identified in 37 persons (6.65%). The remaining 19-item MAIA-2 had good targeting (person mean location: 0.35 ± 1.20 logits), and almost no floor (0.54%), or ceiling effect (0.72%). The person separation reliability (PSR) was 0.94, reflecting the ability to make individual clinical decision-making. No DIF was found for sex, current breathing or body awareness practice, or mental health issues. However, PCAR had an eigenvalue of 3.71 (19.53% variance). The paired t-tests resulted in 23.38% of participants having a significantly different location on the two subtests created by positively and negatively loaded items on the first component, indicating multidimensionality. Additionally, local item dependence (LID) was found in 21 item pairs.

Discussion: The Rasch-based 19-item MAIA-2 demonstrated good item and person fit, good targeting, and excellent PSR. Further analysis to improve unidimensionality and LID is needed.

Session 7, 16.20-16.40

Psychometric assessment and validation of the Modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) patient-reported outcome measure for Long COVID or Post-COVID syndrome

Horton M., Lawrence R., Milne R.

Introduction: The C19-YRS was the first condition-specific, validated scale published for patient assessment and monitoring in Long COVID or Post-COVID syndrome. The original C19-YRS was previously assessed using Rasch methodology, and this informed the evolution of the original C19-YRS to the modified version, where modifications were made based on a combination of psychometric evidence, clinical content relevance and feedback from patients and healthcare professionals. This study aims to psychometrically assess and validate the modified C19-YRS using newly collected data from a large-scale, multi-centre NIHR study looking to optimise Long-COVID treatments and services across the NHS (LOCOMOTION).

Methods: 1278 patients (67% Female; mean age = 48, SD 13.5) completed the C19-YRSm scale digitally on the LOCOMOTION study ELAROS PROMs platform. The psychometric properties of the Symptom Severity (SS), Functional Disability (FD) and Additional Symptoms (AS) subscales were assessed using a Rasch Measurement Theory framework, where all individual scale items were assessed for model fit, targeting, reliability, unidimensionality, local dependency, response category functioning and differential item functioning (DIF) by age group and sex. A cross-validation set-up was utilised, with the sample randomly split into three groups to assess for replication of results.

Results: Rasch analysis revealed promising psychometric properties of the modified SS and FD subscales, with both demonstrating unidimensionality, displaying good targeting and reliability (SS: PSI=0.81, Cronbach's alpha=0.82; FD: PSI=0.76, Cronbach's alpha=0.81), evidence that the modified 4-point response category structure was successful, and no consistent indication of DIF by age group or sex. However, some individual measurement anomalies were noted within the scales, indicating consistent underdiscriminations for the Cough, Smell/Taste (both SS) and Communication (FD) items. Minor pairwise local dependency was apparent between the Fatigue and Post-Exertional Malaise items of the SS scale (Q3 value = 0.27 above average). The AS scale was unidimensional, but displayed skewed targeting.

Discussion: Although some minor anomalies are apparent, the modifications to the original C19-YRS appear to have strengthened the measurement characteristics, and the clinical and conceptual relevance of the scales. Further analytic and multidisciplinary (psychometric, clinical and patient perspective) collaborative work is needed to determine further developments.

Session 7, 16.40-17.00

Stigmatizing stigma scales? Cognitive interviews of items from the Community Attitudes to Mental Illness scale (CAMI) and Mental Illness Knowledge scale (MAKS)

Scholtens S.

Introduction: The Public Health Agency of Sweden has been assigned by the government to measure and prevent stigma towards mental illness. Surveys are not only a measurement tool, they also transfer information to respondents on the topics at hand and are interpreted to reflect the knowledge and values of the survey sender (1). There is limited information on how the self-report scales used to measure stigma are understood by respondents. What messages do the scales used to measure stigma send to the respondent?

Methods: Nine individual cognitive interviews were performed on 16 items from MAKS scale and CAMI scale (fall 2021). The interviews were conducted face-to-face with participants between 20 - 70 years old (median = 43 years old). Recruitment was based on a convenience sample of people not working within health care, public health or other work places with a pronounced role in stigma prevention. Five of the participants identified as women and four identified as men. Three of the participants had an intermediate level of education and six had an advanced level of education. Three of the participants had a native language that was not Swedish.

Results: Several of the items were found to be offensive by the participants. The statements used in the scale were interpreted as proliferating stereotypical views of people with mental illness. Participants questioned the use of "people with mental illness" as a homogenous group that could be judged collectively. Participants questioned the premise of measuring community attitudes on mental illness in population based self-reports as it assumes that people taking part in the survey could not themselves have mental illness which the participants felt was an antiquated view of mental illness.

Discussion: Cognitive interviews can give important insights to the overall messaging that statement items are sending regarding the topic at hand. This aspect of the survey situation needs to tested and taken into consideration when choosing measurement scales.

1. Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. International Statistical Review/Revue Internationale de Statistique, 153-168

SAMC Session 8

Session 8, 8.00-8.20

New method for visualizing Rasch item fit using conditional item characteristic curves

Buchardt A-S., Christensen K.B.

We present a new method to visually evaluate Rasch item fit. Many statistical tests have been proposed for evaluating the fit of an empirical data set to the Rasch model. Some of these fit statistics have unknown asymptotic distributions and this complicates their use. However, even for item fit statistics with known asymptotic distributions or non-parametric item fit statistics it is perhaps not optimal that evaluation of item fit relies on p-values. We present a plot that can be used to inspect item fit for the dichotomous Rasch model, the rating scale model and the partial credit model. This is done using conditional item characteristic curves (ICC). We also show how the method can be used to investigate differential item functioning (DIF).

Session 8, 8.20-8.40

Analyzing polytomous data in R: a simulation study comparing TAM (MML) and eRm (CML) with different sample sizes and targeting properties using the partial credit model

Johansson M.

Introduction: In a recent paper (Linacre, 2022; doi: 10.5116/ijme.629d.d88f), the claim was made that the eRm package (which uses conditional maximum likelihood estimation) was unsuitable for analyzing polytomous data, while the TAM package (marginal maximum likelihood) was recommended. Since the paper did not provide sufficient details for replicating the analysis, this simulation study was designed to investigate this in detail.

Method: Item sets with 5, 8, 11, and 14 items with 4 response categories each were used to simulate datasets based on vectors of normally distributed thetas. Sample size was varied from 3 to 40 respondents per threshold estimated, while targeting was varied from being identical to item mean/SD to having a mean at -2 logits in steps of 0.5 logits (SD held constant). The design was fully crossed for all parameter variations. For each combination 200 datasets were simulated. The Partial Credit Model was used for all analyses.

Results: Preliminary results indicate that the TAM package shows systematic bias, inflating the location of upper and lower item category thresholds, while eRm is more consistent. The results were similar in a smaller simulation study varying the number of response categories from 3 to 5 with other parameters constant. A useful outcome of this study is that the figures documenting results can be used to determine statistical power in terms of item parameter precision for Rasch analysis, based on the number of item thresholds estimated, sample size, and targeting.

Discussion: The eRm package is reliable for estimating item threshold location parameters for polytomous data using the partial credit model. The source code produced for this study will be made available on GitHub to enable reproducible results and simplify future simulation studies in R with polytomous data using open source software.

SAMC Session 8

Session 8, 8.40-9.00

Calculating empirical critical values for Rasch item fit statistics using R

Christensen K. B., Buchardt A-S., Horton M.

Item fit statistics are very frequently reported in Rasch analysis. Many of them have an unknown asymptotic distribution, and the interpretation of what constitutes 'fit' relies on generalized rules which might not always be appropriate -of-thumb. Extensive simulation studies have shown that it is not possible to provide a single critical value that will apply in all settings. This is a problem for all Rasch validation studies that relies on proprietary software. We introduce an easy-to-use software application that will help in establishing critical values on a case by case basis.

Session 9, 9.30-9.50

Local dependence in health outcome measurement: Lessons from the 8-item Parkinson's disease questionnaire (PDQ-8)

Hagell P.

Introduction: The PDQ-8 is a widely used patient-reported health outcome measure in Parkinson's disease (PD) research and practice. As such, it is influential in clinical and policy decision-making processes that impact the care of persons with PD. However, its target variable is undefined and rigorous testing of its measurement properties is lacking. This study examined the measurement properties of the PDQ-8 to understand its role as an outcome measure.

Methods: Complete PDQ-8 item response data from 1289 people with PD from the Swedish national registry for Parkinson's disease were used. PDQ-8 items represent eight health-related problems and have five ordered response categories ("never" to "always"; scored 0-4). Data were analyzed according to the Rasch model using the RUMM2030plus software, focusing on targeting, reliability, response category functioning, model fit, differential item functioning (DIF) by sex and age, and local dependence (LD).

Results: The sample represented all stages of PD severity (stage 0: No signs of disease; stage 5: confined to bed/wheelchair unless aided) and consisted of 64% men. The mean (SD) age was 71 (9.5) years. The mean (SD) person location was -1.30 (0.91) logits. Reliability was 0.67. There were disordered thresholds for all items but one. Four items had significant Bonferroni-corrected chi-square statistics (P0.001), of which two had large fit residuals (-3.26 and 4.99). There were uniform DIF by age (two items) and sex (one item). Residual correlations identified LD for three item pairs. Subtests were created stepwise to absorb LD. This revealed additional LD, leading to identification of two conceptually logical subtests (four items each). Subtests showed a mean (SD) person location of -0.98 (0.7) logits, reduced reliability (0.58), ordered response categories, improved fit residuals ($\leq \pm 0.47$; $P < 0.001$), and DIF by age for one subtest but no DIF by sex.

Discussion: Identifying and resolving LD improved several aspects of its measurement properties but revealed inferior reliability and targeting remained suboptimal. Together with unclear construct validity, this argues against the appropriateness of the PDQ-8 as an outcome measure. These experiences demonstrate the central role of LD and the importance of considering LD when testing rating scale measurement properties.

SAMC Session 9

Session 9, 9.50-10.10

Lessons learned from multiple choice items: how can distractors be useful?

Wahlkrantz E., Melin J.

Introduction: Tests comprising multiple-choice items are widely used in education and other areas. Multiple-choice items are sometimes criticized for inviting guessing and focusing on lower-order thinking; however, they are efficient compared to open questions and well-suited to large-scale assessments of abilities, knowledge, and skills. An important aspect when designing multiple-choice items is the role of distractors. We aim to present insights from working with the Swedish Enlistment Battery (SEB) and recently developed tests under evaluation for a refined SEB.

Methods: Data from test takers for admission tests for basic military training in Sweden was provided by the Swedish Defense Conscription and Assessment Agency. A selection of tests measuring verbal-, logical-, and visuospatial abilities, included in the existing SEB and recently developed, were used. Item designs were qualitatively analyzed by content and distractor placement compared to the correct response and iteratively interpreted with results from Rasch analyses for each sub-test.

Results: The distractors in the studied multiple-choice items have different characteristics, such as similarities and dissimilarities with the correct response, position compared to the correct response, and information content that can be scored or not. Shared features – such as shapes, logic, and frequencies – of distractors and the correct response increased the difficulty of the items, meaning the distractors were conceivable but still incorrect. Such strong distractors compared to the correct response further affected the difficulty of an item, especially if the distractor was placed before the correct response. Some distractors contain more information than others, i.e., being partly correct, that could be used to improve measurement precision by applying polytomous scoring.

Discussion: The structure of multiple-choice items for measuring verbal-, logical-, and visuospatial abilities warrants different strategies in the design phase and analyses. Based on empirical findings from the SEB and tests for refined SEB, our insights can contribute to an improved scientific basis for the role of distractors in multiple-choice items.

SAMC Session 10

Session 10, 10.30-10.50

Differential item functioning in Work Environment Impact Questionnaire (WEIQ) subscales

Yngve M., Ekbladh E., Melin J.

Introduction: A person's ability to work arises in a dynamic relationship between the working individual, the work activity, and the work environment. Work environment assessments such as self-percieved workability may be central to enhancing a person's ability to work, reduce sick leave, and optimize work production. It is important to have valid and quality-assured measures for assessing specific work environment aspects to enable tailored interventions for individuals and workplaces. Nevertheless, the work activities and environment may differ across workplaces and settings. In turn, this can challenge measurement properties and invariant comparisons. This study aimed to compare measurement properties, focusing on differential item functioning (DIF), for the subscales in Work Environment Impact Questionnaire (WEIQ) in three workplaces.

Methods: A secondary analysis was conducted on 288 respondents from three different workplaces: assisted living personnel, vocational rehabilitation personnel, and personnel at a research institute. The WEIQ comprises four subscales: Physical and organizational prerequisites (7 items), Work-tasks (10 items), Social relations (7 items), and Motivational aspects (9 items), which were evaluated to the Rasch model in RUMM2030.

Results: The four subscales showed a reasonably good targeting to the sample, no major misfitting item (2 of 33 items, $p < 0.05$), and person separation indexes ranging from 0.64 to 0.78. Significant DIF was presented in all subscales and between all three workplaces, which qualitatively could be explained due to differences in work activities and environments. By splitting one or more items into separate items for different workplaces, DIF issues could be resolved without jeopardizing other measurement properties.

Discussion: WEIQ seems to be a valid and quality-assured measure for assessing specific aspects of the work environment. At the same time, while DIF could be resolved, the practical relevance of those actions remains. The comparability between the three workplaces may be of little priority. However, the DIF issues shown also apply within workplaces with different work activities. Furthermore, those questions can also be extended to most settings where different critical subsamples have varying preconditions and raise questions on to what extent item splits are justified and how many items can be split to provide valid and quality-assured measures.

SAMC Session 10

Session 10, 10.50-11.10

Cognitive interviews guide the interpretation of a Rasch analysis of The Short Warwick–Edinburgh Mental Well-Being Scale with data from the Swedish public health survey

Reinius, M., Busch H., Gunnarsdottir H., Scholtens S.

Introduction: The Public Health Agency of Sweden use The Short Warwick–Edinburgh Mental Well-Being Scale (SWEMWBS) to assess mental health well-being in the biannual Swedish public health survey Hälsa på lika villkor. We have explored the validity of the scale with both qualitative and quantitative methods.
Methods: Data from the Swedish public health survey 2020-2022 (n=51276) was analysed with Rasch measurement theory using the R package RISEkbmRasch. Results from cognitive interviews (7 persons aged 19-73, interviewed by SCB (1) and 33 persons aged 16-25, interviewed by Högskolan Väst (2) informed the interpretation of the Rasch analysis.
Results: In line with earlier research (3) person attribute values were positively skewed whilst item threshold attribute values had gaps and were negatively skewed. Four items had a different position in the hierarchy compared to earlier research with the item "I've been feeling useful" as the most "difficult" item in the hierarchy.
Test persons in cognitive interviews found the items in SWEMWBS to be vague and sometimes difficult to answer. Some wondered if they should think only about them-selves or also take into account what happens in the world around them when it came to items like "I've been feeling optimistic about the future". The item "I've been feeling useful" was interpreted by youths as negative and provocative which could explain why it gets the highest "difficulty" level in contrast to an "in-the-middle"-position in American studies.
Discussion: It is crucial to use both qualitative and quantitative methods in validity studies of scales. Cognitive interviews can guide selection and modification of items before surveys are sent out and inform interpretation of quantitative analysis when data has been collected.

1. Statistics Sweden (SCB), Nationella Folkhälsoenkäten (HLV) 2023 – Kognitiva intervjuer, Stockholm: Statistics Sweden; 2022
2. Folkhälsomyndigheten & Högskolan Väst, Ungas röster om psykisk hälsa, Stockholm: Folkhälsomyndigheten; 2023
3. Melin J, Lundin A, Johansson M. An off-target scale limits the utility of Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) as a measure of well-being in public health surveys. Public Health. 2022;202:43-8.

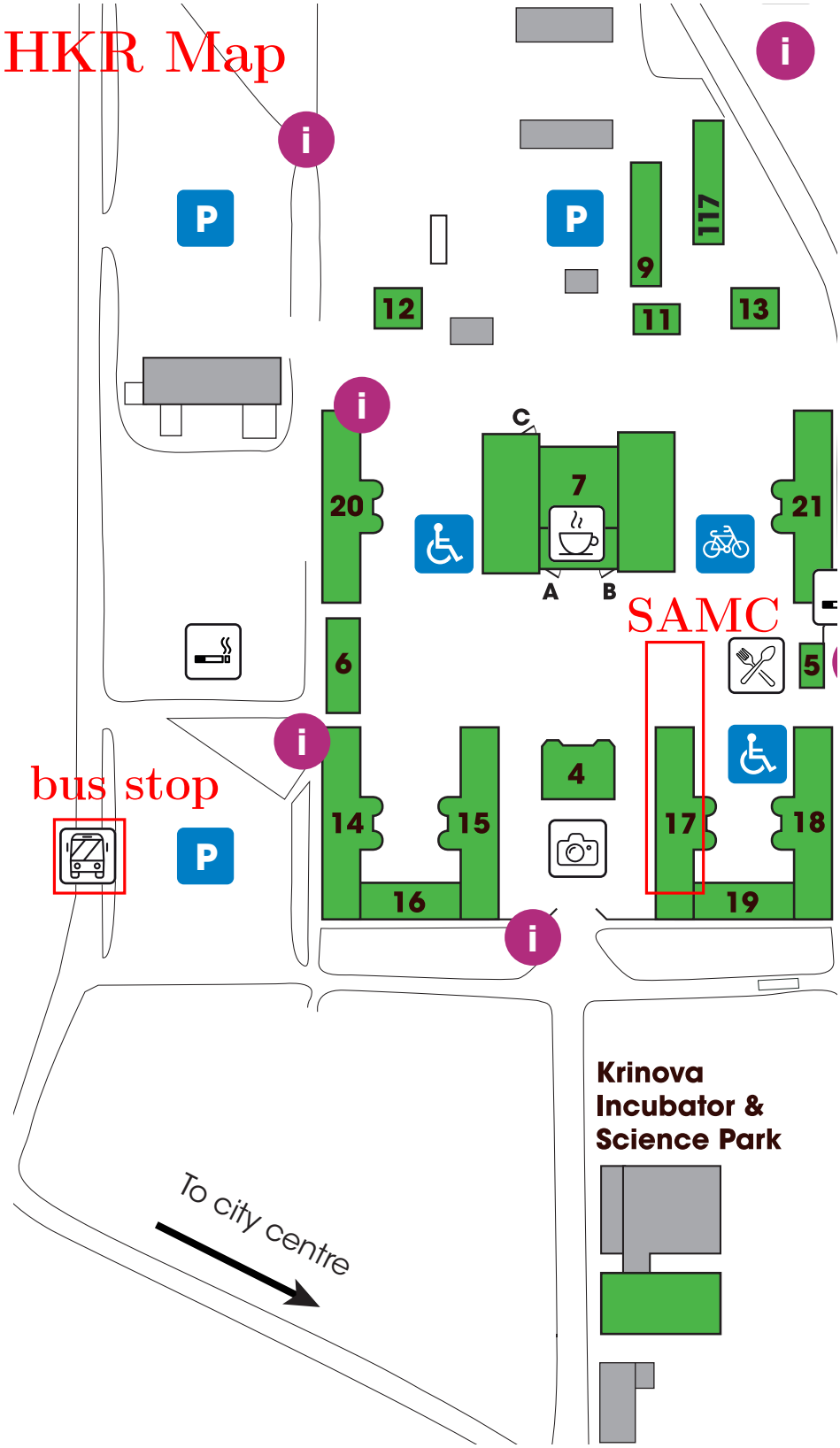
Practical information / Venue description

Internet connection

- Kristianstad University is connected eduroam. In order to connect you need an account on an affiliated organization.
- Kristianstad University offers a free guest wifi network. Connect to "HKR-Guest"and you will be automatically redirected to a portal with terms of use for the service. After accepting the terms of use, the guest is assigned internet access immediately. No registration or creation of a user account is required. Internet access is allocated in 24-hour intervals. After 24 hours, the user is automatically redirected to the guest portal for re-acceptance of the terms of use, after which a new 24-hour period of internet access is activated.

How to get here

- Kristianstad University is located approximate 20-minute walk from Kristianstad Centralstation, or 10 minutes by city bus number 4 (marked Lingenässkolan via Högskolan) from Kristianstad Centralstation. Feel free to download Skånetrafiken's app to buy a ticket for the city bus.
- By car, it's a 5-minute drive from Kristianstad city center.
- Details about the bus stop, parking lots, and the conference building are shown on the map on the next page.


HKR Map
i
P
P
117
9
12
11
13
i
C
20
7
21
A
B
SAMC
6
5
i
bus stop
4
14
15
17
18
P
16
19
i
Krinova
Incubator &
Science Park
To city centre

Conference contributors / Abstract authors

Author	Affiliation
Aiyegbusi, Olalekan Lee	University of Birmingham, Birmingham, United Kingdom
Alenius, Sara	Kristianstad University, Kristianstad, Sweden
Anderson, Nicola	University of Birmingham, Birmingham, United Kingdom
Andrich, David	The University of Western Australia, Crawley, Australia
Baghaei, Purya	IEA-Hamburg, Hamburg, Germany
Bala, Sidona-Valentina	Department of Health Sciences, Faculty of Medicine, Lund University, Lund, Sweden Department of Medicine, Section of Rheumatology, Helsingborg Central Hospital, Helsingborg, Sweden
Baskaran, Karthikeyan	Linnaeus University, Department of Medicine and Optometry, Kalmar, Sweden
Behm, Lina	Kristianstad University, Kristianstad, Sweden
Bissell, Andrew	University of Birmingham, Birmingham, United Kingdom
Buchardt, Ann-Sophie	University of Copenhagen, Copenhagen, Denmark
Burgmanis, Ģirts	University of Latvia, Interdisciplinary Centre for Educational Innovation, Riga, Latvia
Busch, Hillevi	Public Health Agency Sweden, Solna, Sweden
Calvert, Melanie	University of Birmingham, Birmingham, United Kingdom
Carpentier, Sydney	Rehabilitation Science Graduate Program, Medical School, University of Minnesota, Minneapolis, USA
Christensen, Karl Bang	Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark
Cockwell, Paul	University of Birmingham, Birmingham, United Kingdom
Collste, David	Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
Conte, Helen	Department of Neurobiology, Care Science and Society, Karolinska Institutet, Stockholm, Sweden
Cornell, Sarah	Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
Damery, Sarah	University of Birmingham, Birmingham, United Kingdom
Djukanovic, Ingrid	Health and Caring Sciences, Kalmar, Sweden
Ekbladh, Elin	Linköping University, Department of Health, Medicine and Caring Sciences, Linköping, Sweden
Ekstedt, Mirjam	Health and Caring Sciences, Kalmar, Sweden
Ekstrand, Joakim	The Pro-CARE Group, Faculty of Health Sciences, Kristianstad University, Kristianstad, Sweden
Elefalk, Kjell	Trygghet och Management AB, Stockholm, Sweden
Escher, Cecilia	Anesthesia and Intensive care Norrtälje Hospital, Norrtälje, Sweden

Author	Affiliation
Eskel, Marit	RISE, Research Institutes of Sweden, Division Safety and Transport, Division Measurement Science and Technology, Stockholm, Sweden
Fagerström,Cecilia	Region Kalmar County, Kalmar, Sweden
Ferguson, James	University of Birmingham, Birmingham, United Kingdom
Fletcher, Benjamin	University of Worcester, Worcester, United Kingdom
Friedman, Tim	Australian Council for Educational Research, Melbourne, Australia
Gasser, Fredrik	The Pro-CARE Group, Faculty of Health Sciences, Kristianstad University, Kristianstad, Sweden The Research Platform for Collaboration for Health, Faculty of Health Sciences, Kristianstad University, Kristianstad, Sweden
Grzęda, Mariusz	Galen Research, Manchester, United Kingdom
Gunnarsdottir, Hrafnhildur	Public Health Agency Sweden, Solna, Sweden
Guttersrud, Øystein	Oslo Metropolitan University (OsloMet), Oslo, Norway
Haaland-Øverby, Mette	Oslo Metropolitan University (OsloMet), Oslo, Norway
Hagell, Peter	The Pro-CARE Group, Faculty of Health Sciences, Kristianstad University, Kristianstad, Sweden
Harbut, Piotr	Department of Anesthesia and Intensive care, Danderyds University Hospital, Stockholm, Sweden
Helseth, Sølvi	Oslo Metropolitan University (OsloMet), Oslo, Norway
Hobart, Jeremy	University of Plymouth, Plymouth, United Kingdom
Hollings, Samantha-Jo	Galen Research, Manchester, United Kingdom
Horton, Mike	University of Leeds, Leeds, United Kingdom
Jarl, Martin	Department of Neurobiology, Care Science and Society, Karolinska Institutet, Stockholm, Sweden Department of Anesthesia and Intensive care. Danderyds University Hospital, Stockholm, Sweden
Johansson, Magnus	RISE Research Institutes of Sweden, Uppsala, Sweden
Johansson, Stefan	University of Gothenburg, Gothenburg, Sweden
Johnstone, Ellie	Galen Research, Manchester, United Kingdom
Juan, Andrea	University of KwaZulu-Natal, Durban, South Africa
Karlsson, Simon	RISE, Research Institutes of Sweden, Division Safety and Transport, Division Measurement Science and Technology, Stockholm, Sweden
Kazunga, Cathrine	University of KwaZulu-Natal, Durban, South Africa
Korsell, Nicklas	RISE, Research Institutes of Sweden, Division Safety and Transport, Division Measurement Science and Technology, Stockholm, Sweden
Kyte, Derek	University of Worcester, Worcester, United Kingdom
Lawrence, Román	ELAROS 24/7 Ltd., Sheffield, United Kingdom
Lay,Dulce	Australian Council for Educational Research, Melbourne, Australia
Le, Christopher	Høyskolen i Innlandet (INN), Hamar, Norway
Macedo, Antonio Filipe	Linnaeus University, Department of Medicine and Optometry, Kalmar, Sweden

Author	Affiliation
Marais, Ida	The University of Western Australia, Perth, Australia
Melgaard, Dorte	Department of Acute Medicine and Trauma Care, Aalborg University Hospital, Aalborg, Denmark Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
Melin, Jeanette	Swedish Defence University, Department of Leadership and Demand & Control, Karlstad, Sweden RISE, Research Institutes of Sweden, Division Safety and Transport, Division Measurement Science and Technology, Gothenburg, Sweden
Mikite, Marta	University of Latvia, Interdisciplinary Centre for Educational Innovation, Riga, Latvia
Milne, Ruairidh	University of Southampton, Southampton, United Kingdom
Mohlin, Camilla	Linnaeus University, Department of Chemistry and Biomedicine, Kalmar, Sweden
Möllerberg, Marie-Louise	Malmö University, Malmö, Sweden RISE Research Institutes of Sweden, Lund, Sweden
Nicolaidis Lindqvist, Andreas	Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden RISE Research Institutes of Sweden, Lund, Sweden
Nilsson Lidström, Petra	Kristianstad University, Kristianstad, Sweden
Nilsson, Ida	Linnaeus University, Department of Medicine and Optometry, Kalmar, Sweden
Nilsson, Marie	Kristianstad University, Kristianstad, Sweden
Nilsson, Ulrica	Department of Neurobiology, Care Science and Society, Karolinska Institutet, Stockholm, Sweden Perioperative Medicine and Intensive Care, Karolinska University Hospital, Stockholm, Sweden, Stockholm, Sweden

Author	Affiliation
Paap, Muirne	University of Groningen, Groningen, Netherlands
Palm, Celinda	Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
Pedersen, Henrik	The research unit for general practice, Aarhus, Denmark
Pendrill, Leslie	RISE, Research Institutes of Sweden, Division Safety and Transport, Division Measurement Science and Technology, Gothenburg, Sweden
Petersen, Jørgen Holm	Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark
Rask, Marie	Kristianstad University, Kristianstad, Sweden
Reinius, Maria	Public Health Agency Sweden, Solna, Sweden
Rosén, Monica	University of Gothenburg, Gothenburg, Sweden
Sappl, Sonia	School Curriculum and Standards Authority, Perth, Australia
Scholtens, Sara	Public Health Agency Sweden, Solna, Sweden
Sidey-Gibbons, Chris	University of Texas, Houston, USA
Sivan, Manoj	University of Leeds, Leeds, United Kingdom
Skirbekk, Helge	Oslo Metropolitan University (OsloMet), Oslo, Norway
Slade, Anita	University of Birmingham, Birmingham, United Kingdom
Smithard, David	Centre for Exercise Activity and Rehabilitation, School of Human Sciences, University of Greenwich, London, United Kingdom Elderly Care, Queen Elizabeth Hospital, Lewisham and Greenwich NHS Trust, London, United Kingdom
Spray, Isobel	Galen Research, Manchester, United Kingdom
Staal, Sabina	Department of Physio- & Occupational Therapy, Copenhagen University Hospital, Hvidovre, Denmark
Strietholt, Rolf	IEA-Hamburg, Hamburg, Germany
Svanfeldt, Cecilia	Region Kalmar Lan (Hospital), Department of Ophthalmology, Kalmar, Sweden
Søberg Finbråten, Hanne	Høyskolen i Innlandet (INN), Hamar, Norway
Turner, Neil	University of Edinburgh, Edinburgh, United Kingdom
Van de Winckel, Ann	Division of Physical Therapy and Rehabilitation Science, Medical School, University of Minnesota, Minneapolis, USA
Verdi, Rav	University of Birmingham, Birmingham, United Kingdom
Wahlkrantz, Emelie	Swedish Defence University, Department of Leadership and Demand & Control, Karlstad, Sweden
Westergren, Albert	The Pro-CARE Group, Faculty of Health Sciences, Kristianstad University, Kristianstad, Sweden
Yngve, Moa	Linköping University, Department of Health, Medicine and Caring Sciences, Linköping, Sweden
Årestedt, Kristofer	Faculty of Health and Life Sciences, Linnaeus University, Kalmar, Sweden Research Section, Region Kalmar County, Kalmar, Sweden