

Introduction to Rasch Measurement Using WINSTEPS and FACETS

**Workshop Leaders: Dr. Thomas Eckes & Frank Weiss-Motz,
TestDaF-Institut, Bochum, Germany**

LEARNING OUTCOMES

Participants will be introduced to the basic rationale of Rasch measurement. They will learn how to use the computer programs WINSTEPS and FACETS in order to analyse data from language assessments comprising two facets (examinees, items) or more than two facets (examinees, raters, criteria, etc.). The participants will also learn how to prepare the input data for a WINSTEPS or FACETS analysis and how to interpret the output from these programs.

We expect the participants to have some basic knowledge of item and test development (e.g., item types, item difficulties, rating scales, measurement error, reliability) as well as descriptive statistics (e.g., arithmetic mean, standard deviation, correlation), as typically provided in introductory courses in language testing and assessment. Prior knowledge of item response theory (IRT) or practical experience with using IRT software is not a prerequisite condition for successfully participating in our workshop.

CONTENT

The workshop will be organized into two interrelated parts: Part 1 runs from Tuesday, 22nd May, 2:00 pm, to Wednesday, 23rd, 12:30 pm. This part introduces the theoretical rationale and practical utility of a Rasch measurement approach to language assessment. We will briefly discuss the dichotomous Rasch model and use a small data set to illustrate its implementation in WINSTEPS. Key issues refer to constructing measures of examinee ability and item difficulty (Wright map), interpreting statistical indicators of data–model fit, and making use of the measurement results.

Part 2 runs from Wednesday, 23rd, 2:00 pm, to Thursday, 24th, 12:30 pm. The second part builds on Part 1 measurement issues and broadens the perspective to include facets typically involved in rater-mediated assessments. To illustrate basic concepts and procedures of a many-facet Rasch analysis as implemented in the FACETS program, we will use sample data from a language assessment, where raters evaluated the performance of examinees on a writing task according to a small set of criteria using a four-category rating scale.

METHOD

Participants will be asked to bring along their own notebooks, with pre-installed versions of WINSTEPS and FACETS. To encourage the participants to get actively involved in our workshop, they will not be required to have access to the full versions of these programs.

For the purposes of this workshop, it will be wholly sufficient to use the freely available student versions called MINISTEP and MINIFAC, respectively.

Specifically, the data sets that we will examine in due detail in Part 1 and Part 2 of the workshop are amenable to the complete functionality of WINSTEPS and FACETS. For example, the writing performance data that will be subjected to a many-facet Rasch analysis comprises a total of 1,944 responses (ratings), and thus stays below the maximum allowed number of responses in MINIFAC (i.e., 2,000 responses). Moreover, this data set has been extensively studied for illustrative purposes before (Eckes, 2009, 2015, in press) and, therefore, lends itself particularly well to demonstrate the practical implementation of the many-facet Rasch approach to the analysis and evaluation of rater-mediated assessments. In terms of a step-by-step process, we will show our workshop participants how to start WINSTEPS (MINISTEP) or FACETS (MINIFAC), guide them through building control or specification files, preparing the input data, running the analysis and, finally, interpreting the measurement results

PRE-WORKSHOP ACTIVITIES

The first thing we will ask the prospective participants to do is to download the student versions MINISTEP (Version 4.0; Linacre, 2017b) and MINIFAC (Version 3.80; Linacre, 2017a) and install these programs on their notebook (in case they do not have access to the full program versions). Both programs are available as a free download at the following address: <http://www.winsteps.com/index.htm>

We will also ask the participants to read the chapter on many-facet Rasch measurement included in the *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages* (<https://rm.coe.int/1680667a23>). Alternatively, participants may take a look at the introductory chapters in Eckes (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments* (2nd ed.) Frankfurt am Main: Peter Lang.

WORKSHOP LEADERS

Thomas Eckes holds a doctor's degree (Dr.phil.) from the University of the Saarland (Saarbrücken, Germany). He is Head of the Psychometrics and Language Testing Research Department at the TestDaF-Institut. Thomas has published in leading language assessment journals (e.g., Eckes, 2005, 2008, 2012, 2017). He is on the Editorial Board of the journals *Language Testing*, *Assessing Writing*, and *SAGE Open*. His book *Introduction to Many-Facet Rasch Measurement* (Peter Lang, Frankfurt am Main, Germany) appeared 2015 in a second, expanded edition. Recently, Thomas has finished writing a chapter on the use of many-facet Rasch measurement within language assessment contexts; this chapter is to appear in a two-volume set edited by V. Aryadoust and M. Raquel. Also, he is guest editor of a forthcoming special issue on IRT modeling of rater effects (to appear in the journal *Psychological Test and Assessment Modeling*).

Frank Weiss-Motz holds a diploma in psychology (Dipl.-Psych.) from the University of Kiel, Germany. He is a member of the Psychometrics and Language Testing Research Department at the TestDaF-Institut. Frank is responsible for data management, data analysis and evaluation regarding the TestDaF-Institut's large-scale assessments, with a focus on the TestDaF and the TestAS. He is currently involved in designing a number of validity studies regarding the TestAS, including studies of the effects that limited processing time and variations in stimulus presentation have on item statistics and examinees' solving strategies as well as studies of the extent to which the TestAS is predictive of applicants' academic success at Germany universities.