

EALTA 2019 Pre-Conference Workshop Proposal

Nahal Khabbazzashi & Anthony Green

1. Workshop Topic

Many Facet Rasch Measurement: a practical guide using FACETS

2. Intended learning outcomes

Participants will be introduced to some of the issues associated with involving human judges in the rating process in speaking assessment contexts. They will become familiar with the basic concepts in Many-Facet Rasch Measurement (MFRM) and will learn how to run the FACETS programme and to enter, analyse, and interpret test data with more than two facets e.g. candidates, raters, scoring criteria, tasks, etc. Participants will learn to evaluate the extent to which raters are consistent in their rating behaviour and whether they display systematic differences that can affect candidate scores.

3. Workshop Content

The workshop will be divided into four interrelated sessions:

- Session 1 (afternoon of Tuesday 28th of May 2019) will cover some of the main issues related to rater-mediated contexts such as rater inconsistencies, relative severity levels of raters, systematic bias, and different tendencies in use of assessment scales. We will discuss some of the basic theoretical concepts behind MFRM, the practical contexts in which this type of analysis can be useful, and some rules of thumb for designing studies that use MFRM.
- Session 2 (morning of Wednesday 29th of May 2019) will cover the different steps involved in running MFRM with the programme FACETS using an example speaking test data set with three facets (examinees, rates, and criteria).
- Session 3 (afternoon of Wednesday 29th of May 2019) will focus on understanding and interpreting the various outputs from FACETS.
- Session 4 (morning of Thursday 30th of May 2019) consolidates previous sessions by providing participants with the opportunity to independently run, analyse, and interpret another data set. The session will end with a Q&A.

4. Method

We will ask participants to bring their own laptops with pre-installed version of MINIFAC – the freely available student version of FACETS. Alternatively, and subject to availability, we will ask organisers to provide us with an IT-suite with pre-installed versions of MINIFAC. Sample data sets will be e-mailed to participants beforehand and copies will also be available on external USB drives. We will ensure that data sets remain below the maximum allowed number of observations ($n=2000$) in MINIFAC.

Our sessions will be interactive with the first session starting off with paired/group work with a series of simple data sets and hand-on activities that will allow participants to get a concrete understanding of some of the issues related to rater-mediated contexts using an inductive approach. We will illustrate and demo the steps involved in running FACETS including preparing and formatting the data, creating control/specification files, defining models, and

running FACETS. Following the demo, participants will be asked to try out all steps independently. At the end of session 3, we will ask participants to rate a sample speaking test using the CEFR scale. We will collate these scores which would then form part of the data used for the final session; an exercise that would encourage full involvement of participants from the original rating process to data preparation, data analysis, and interpretation. Participants will be asked to complete a series of tasks related to the data set individually and then to compare responses in pairs and small groups.

5. Prior knowledge

We expect the participants to have some basic knowledge of speaking/writing tests (e.g. rating scales, assessment criteria, task difficulty, and reliability). We do not expect familiarity with theories of measurement and MFRM or experience in using FACETS.

6. Pre-workshop activities

We would ask prospective participants to download the latest student version of MINIFAC which can be accessed for free at <https://www.winsteps.com/minifac.htm>. We will also send participants a copy of the data sets to be used in the workshop and ask them to download them onto their laptops. Participants are encouraged to read the following:

- Section H of the *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages* which can be accessed at <https://rm.coe.int/1680667a23>.
- *Measuring Second Language Performance* (1996) written by T. F. McNamara and published by Addison-Wesley Longman.

7. Maximum number of participants

We suggest a maximum of 25 participants for the workshop.

8. Suitability to run such a workshop (supported by a very brief CV).

Nahal Khabbzbashi

Nahal holds a DPhil in Education from the University of Oxford. She is currently a Senior Lecturer in Language Assessment at the Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire. Before joining CRELLA, Nahal held the position of Senior Research Manager at Cambridge Assessment English where she led the research strands of a number of high-profile international projects in different educational and language testing contexts. Nahal's work appears in *Language Testing* (2017), *Language Learning Journal* (2014), as well as in the Cambridge University Press *Studies in Language Testing* series (Moeller, Creswell, and Saville, eds. 2016).

Nahal's primary research interests include the use of Many-Facet Rasch Measurement in rater-mediated assessment contexts (the focus of her doctoral thesis), the automated assessment of speaking, test validation, the use of mixed methods approaches in language testing research, and the effects of task and test-taker related variables on performance. Nahal has lectured in Language Assessment at the University of Cambridge, University of Oxford, and University of Bedfordshire. She has also delivered a number of workshops, seminars, and courses on a range of topics from foundation courses in language testing to specialised training in statistical analyses and Many-Facet Rasch Measurement on behalf of the Association for Language Testers in Europe (ALTE).

Anthony Green

Tony is a Professor of Language Assessment and the Director of the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire. He is the author of *Exploring Language Assessment and Testing* (Routledge), *Language Functions Revisited* and *IELTS Washback in Context* (both Cambridge University Press). He is President of the International Language Testing Association (ILTA) and an Expert Member of the European Association for Language Testing and Assessment (EALTA). He has published widely on language assessment and is Associate Editor of *Language Assessment Quarterly*, Executive Editor of *Assessment in Education* as well as serving on the editorial board of *Assessing Writing*.

Tony's main research interests lie in the relationship between assessment, learning and teaching, particularly in relation to written language. Current projects include defining and promoting assessment literacy among teachers, the development of English Language placement tests, the investigation of item writing and test taking processes and the development of reference level descriptions for English to supplement the Common European Framework of Reference for Languages ([English Profile](#)). In addition to his research work, Tony has extensive experience as a teacher and teacher trainer and has worked in many areas of practical test development and delivery including as a test designer, item writer and examiner for tests of speaking and writing skills. Tony has lectured in Language Assessment at the University of Cambridge, Oxford Brookes University, and the University of North London and has delivered training courses for the Association for Language Testers in Europe and British Council among many others. He has published and conducted workshops on Many-Facet Rasch Measurement.