Using the CEFR Rasch and/or IRT resources: The benefits, the challenges, and the ellipsis

Charalambos Kollias

EALTA CEFR SIG, Dublin 2019

OXFORD
UNIVERSITY PRESS

The Benefits

Reference Supplement
Section B: Standard Setting (2004)

- IRT SS advantages
  - sample free item & person parameter estimation

- Overview of SS methods
  - empirical data through IRT highlighted

December 2004   DGIV/EDU/LANG (2004) 13

Reference Supplement

to the

Preliminary Pilot version of the Manual for

Relating Language examinations to the
Common European Framework of Reference for Languages:
learning, teaching, assessment

Section B: Standard Setting

Language Policy Division, Strasbourg

OXFORD
UNIVERSITY PRESS

Reference Supplement
Section G: Item Response Theory (2004)

December 2004   DGIV/EDU/LANG (2004) 13

Reference Supplement
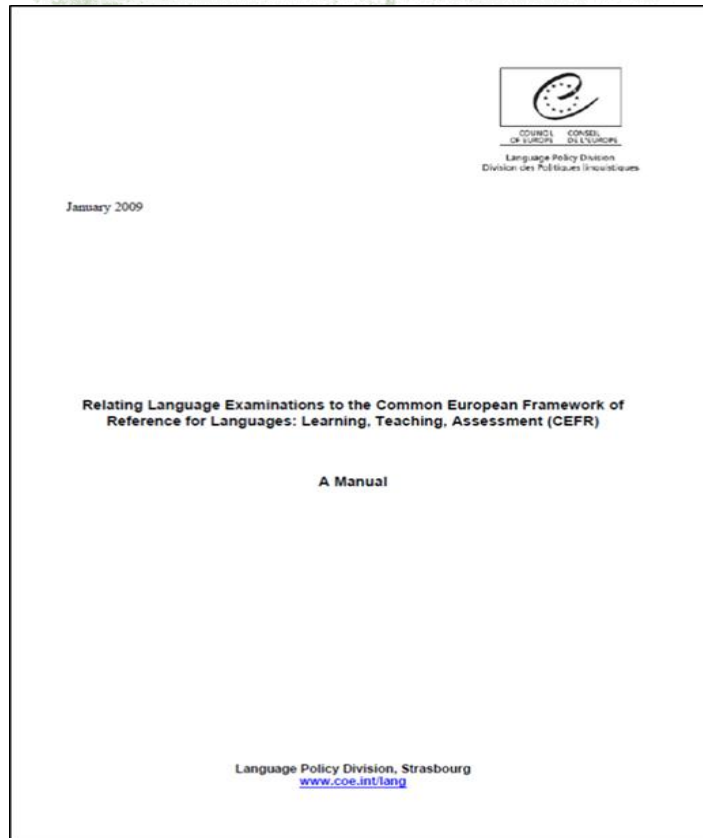
to the

Preliminary Pilot version of the Manual for

Relating Language examinations to the
Common European Framework of Reference for Languages:
learning, teaching, assessment

Section G: Item Response Theory

Language Policy Division, Strasbourg

- Basic notions of IRT explained & discussed

- IRT notions & techniques covered (technical discussion)

OXFORD
UNIVERSITY PRESS

The Manual (2009)

January 2009

Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)

A Manual

Language Policy Division, Strasbourg
www.coe.int/lang

- Concepts related to IRT
  - difficulty parameter and level, and discrimination

- IRT Standard Setting (SS) methods
  - The Bookmark Method
  - A Cito Variation of the Bookmark Method
  - reference to Item-descriptor Matching Method

- SS & test equating
  - reference to test equating through anchor items

OXFORD
UNIVERSITY PRESS

Further Material on Maintaining Standards across Languages, Contexts and Administrations by exploiting Teacher Judgment and IRT Scaling (2009)

- Creating a scale of items linked to CEFR

- Data Collection & Scale Construction
    - Rasch sample size (100 test takers)
    - Using CEFR Anchor Items

- Using CEFR Descriptors as IRT items
    - self-assessment & teacher assessment

- Benchmarking with MFRM (FACETS)

Reference Supplement
Section H: Many-Facet Rasch Measurement (2009)

- Use of Many-Facet Rasch Measurement (MFRM) for rater-mediated assessments

- Coverage of MFRM terms and notions

- MFRM & SS
  - reference to SS

Reference Supplement
Section I: Cito variation on the bookmark method (2009)

Reference Supplement

to the

Manual for

*Relating Language examinations to the
Common European Framework of Reference for Languages:
learning, teaching, assessment*

Section I: Cito variation on the bookmark method

COUNCIL
OF EUROPE   CONSEIL
DE L'EUROPE

October 2009

Language Policy Division, Strasbourg
www.coe.int/lang

- SS procedure

- Estimating RP50 & RP80
    - Rasch model; two parameter logistic model, and three parameter model

- Decision making
    - Transforming the latent scale

OXFORD
UNIVERSITY PRESS

Manual for Language Test Development and Examining (ALTE)

Rasch analysis
- minimum no. of test takers: 50 – 80

MFRM
- minimum no. of performances: 30 per task
- minimum no. of ratings per rater: 30 per rater

DIF analysis with Rasch analysis
- minimum no. of test takers: 500, with at least 100 per group

The Challenges &
the Ellipsis

CEFR Anchor Items

- Limited no. of CEFR anchor items/tasks available

- Limited no. of CEFR anchor test methods used in CEFR anchor tasks

- Few writing tasks and/or responses

- No Rasch and/or IRT calibrated writing tasks/ writing responses

- No Rasch and/or IRT calibrated speaking tasks/ responses

Setting cut scores through Rasch and/or IRT procedures

"The basic flaw of many applications of IRT modelling in language testing especially is that there is not enough evidence provided about the model-data fit, which makes the findings of these studies more or less questionable" (p.17).

(Kaftandjieva, 2004)

- No guidance on analysing cut-scores based on Rasch and/or IRT procedures

- No framework for evaluating standard setting cut scores set through Rasch and/or IRT procedures

- No framework for evaluating intra-panellist and inter-panellist consistency within Rasch model

Q & A

OXFORD
UNIVERSITY PRESS

Thank you

OXFORD
UNIVERSITY PRESS

References

ALTE (2011). *Manual for language test development and examining: For use with the CEFR*. Strasbourg: Council of Europe/Language Policy Division.

North, B. & Jones, N. (2009). *Further material on maintaining standards across languages, contexts, administrations by exploiting teacher judgments and IRT scaling: Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, Assessment (CEFR)*.

Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for languages: Learning, teaching, assessment (CEFR)*. Strasbourg: Language Policy Division.

Eckes, T. (2009). Many-Facet Rasch measurement . In S. Takala (Ed.), *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment, (Section H)*. Strasbourg: Council of Europe/Language Policy Division.

Kaftandjieva, F. (2004). : Standard setting. In S. Takala (Ed.), *Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (Section B)*. Strasbourg: Council of Europe/Language Policy Division.

OXFORD
UNIVERSITY PRESS

References

Verhelst, N. D. (2004). Item response theory. In S. Takala (Ed.), Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (Section G). Strasbourg: Council of Europe/Language Policy Division.

Van der Schoot, F. (2009). Cito variation on the bookmark method. In S. Takala (Ed.), Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (Section I). Strasbourg: Council of Europe/Language Policy Division.