

# **Calibration of National Center Test Items against the Common European Framework of Reference for Languages**

Naoki Sugino, Michiko Nakano, Yuko Shimizu, Kojiro Shojima,  
Kenichi Yamakawa, and Hiromasa Ohba

- This presentation ...
  - is an interim progress report of our project;
  - focuses more on the methodology of compiling a can-do chart and of relating one to the CEFR.
- To share opinions and ideas as to how a can-do chart, both empirically-grounded and externally-corroborated, can be compiled.

### ● Our project

- grew out of the needs for collaboration of SLA researchers and testing researchers
- application of new test theories and statistical models
- “equating” the scales across different grammatical features: first IRT (Yamakawa et al., 2008) to LRT (Ohba et al., 2012)

- Three major background factors
  - Revisions (without sufficient review of their impact) of the *Course of Study*, the national standard for the material and curriculum development
  - Advocacy for the objectives and goals in terms of can-do statements (MEXT's Commission on the Development of Foreign Language Proficiency, 2011)
  - Advocacy for the reform of university entrance exams

**Introduction**

Previous  
Studies

The Present  
Study

Discussion

A common question



**Introduction**

Previous  
Studies

The Present  
Study

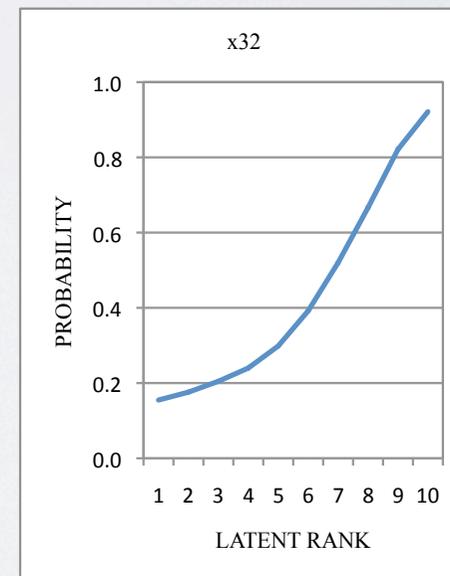
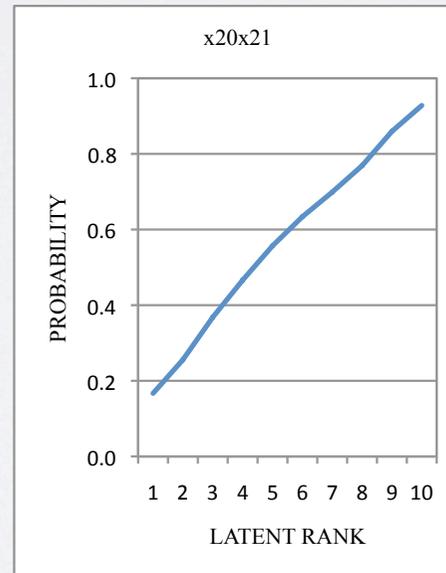
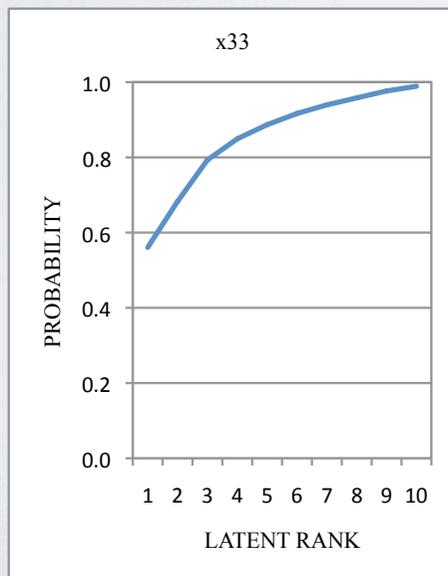
Discussion

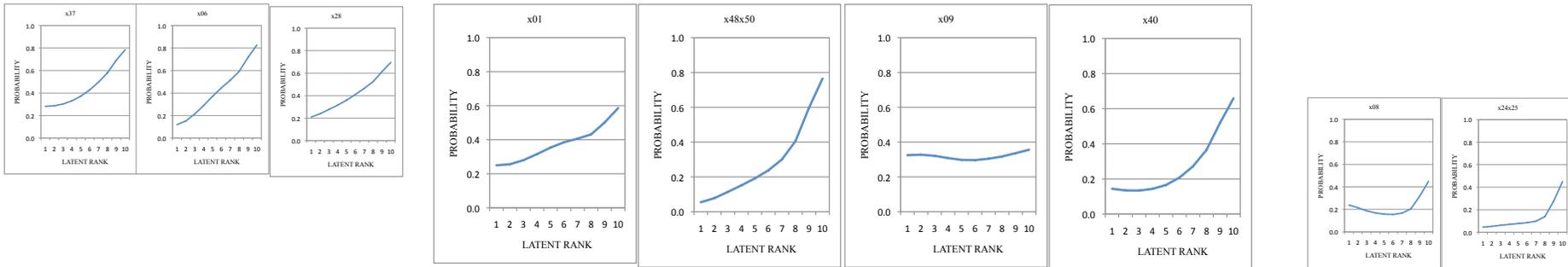
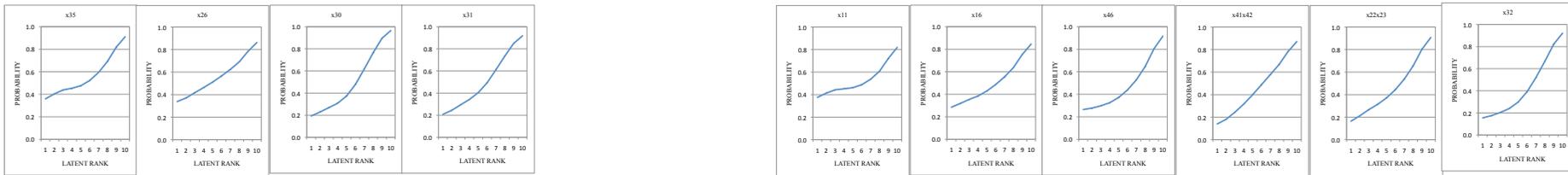
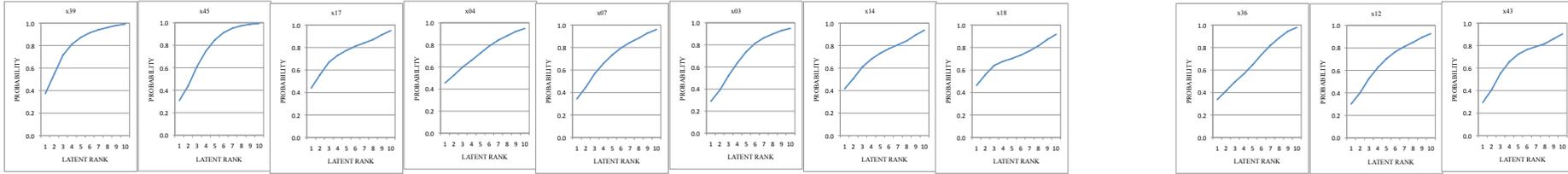
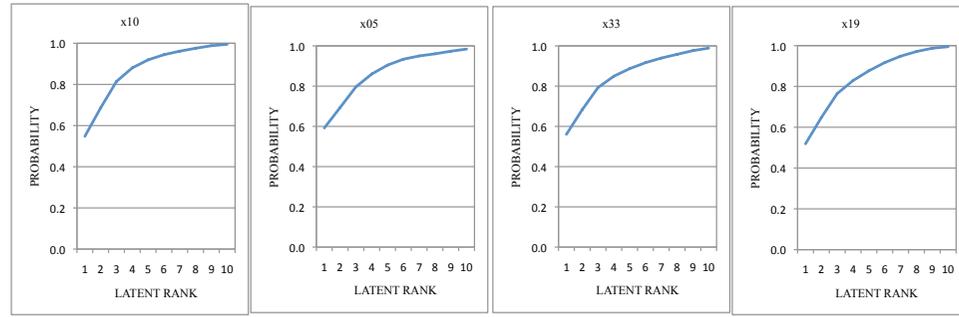
**“What does a test score mean?”**

### ● Objectives of the Project

- To align items from different tests on a single scale
    - So far, three administrations of the NCTs are equated
  - To develop a can-do system
  - To align the tests with the CEFR
- 
- To develop a CAT that yields diagnostic feedback in terms of ability descriptors

- Application of LRT to compile a can-do chart
  - IRP (Item Reference Profile)
    - A set of probabilities with which learners at a particular latent rank can answer correctly





- Application of LRT to compile a can-do chart
  - IRP (Item Reference Profile)
    - A set of probabilities with which learners at a particular latent rank can answer correctly
    - By grouping together those items with similar IRPs, we can obtain the items within the “can-do” range of that particular latent rank.



- Application of LRT to compile a can-do chart
  - IRP (Item Reference Profile)
    - A set of probabilities with which learners at a particular latent rank can answer correctly
    - By grouping together those items with similar IRPs, we can obtain the items within the “can-do” range of that particular latent rank.

► **Empirically-grounded but test-dependent**

## ● The Manual and the Grid

- for relating language tests to CEFR
  - Background and development (Martyniuk, 2010)
  - The Dutch CEFR Grid project led by Alderson (Alderson, et al., 2006)
  - The CEFR level estimates are yielded in the text dimensions, communicative language activities, and communicative language competencies and strategies.

► **The Latent Ranks to be related to CEFR**

Introduction

Previous  
Studies

**The Present  
Study**

Discussion

- Analysis of reading skill test tasks based on the Grid
  - 9 sections (53 items) out of 145 items
    - Items in each section are based on the same passages.
  - The initial analysis using the CEF-ESTIM Grid (Tardieu et al., 2010) by the presenter, confirmed by the members.

Introduction      Previous Studies      **The Present Study**      Discussion

Rank	1	2	3	4	5	6	7	8	above
Items		y90i40	y90i38 y90i39 y90i41	y90i49	y90i46 y90i48 y90i51 y90i52		y90i47 y90i50	y90i54	
		y90i42		y90i43 y90i44 y90i45	y90i53 y90i55 y90i56				
			y97i34	y97i33 v97i35			y97i37	y97i36	
	y97i40	y97i39 y97i42		y97i41					
			y97i43	y97i46 y97i47	y97i45	y97i48	y97i44 y97i49 v97i50		
		y04i33			y04i34 y04i36		y04i35	y04i37	
			y04i39		y04i41 y04i42				y04i40
			y04i45 y04i48		y04i43 y04i44 y04i49		y04i46 y04i47	y04i50	

Introduction      Previous Studies      **The Present Study**      Discussion

Rank	1	2	3	4	5	6	7	8	above	
Items		y90i40 <b>A1/A2</b>	y90i38 y90i39 y90i41	y90i49	y90i46 y90i48 y90i51 y90i52	<b>A2/B1</b>		y90i47 y90i50	y90i54	
		y90i42	<b>A2/B1</b>		y90i43 y90i44 y90i45					
			y97i34	y97i33 y97i35	<b>A2/B1</b>		y97i37	y97i36		
		y97i40	y97i39 y97i42	<b>A2/B1</b>		y97i41				
			y97i43	y97i46 y97i47	<b>A2/B1</b>		y97i45 y97i48	y97i44 y97i49 y97i50		
		y04i33			y04i34 y04i36	<b>A2</b>		y04i35	y04i37	
			y04i39			y04i41 y04i42	<b>A1/A2</b>		y04i40	
			y04i45 y04i48			y04i43 y04i44 y04i49	<b>A2/B1</b>		y04i46 y04i47	y04i50

- All items within the range between A1/A2 and A2/B1
  - conforms with the findings in Negishi, Takada, and Tono (2011); items are suitable for the Japanese EFL learners.
- The levels are not always in accordance with the latent ranks
  - The CEFR levels may not be sufficiently sensitive to the differences in their achievement; or
  - A wide range of ranks within a section is a possible cause of the mismatch?

Introduction

Previous  
Studies

The Present  
Study

**Discussion**

## ● Future research

- Item-by-item analysis?
- Combined feedback; RMP (Rank Membership Profile; another type of output of LRT about the learners) and the CEFR-levels
- CEFR-J?



Introduction

Previous  
Studies

The Present  
Study

Discussion

**Any questions, comments, suggestions, etc.  
are immensely appreciated.**

**Or, please send them to ...  
[gwisno@is.ritsumei.ac.jp](mailto:gwisno@is.ritsumei.ac.jp)**

## References

- Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The Dutch CEFR construct project. *Language Assessment Quarterly*, 3, 3–30. Retrieved from [http://dx.doi.org/10.1207/s15434311laq0301\\_2](http://dx.doi.org/10.1207/s15434311laq0301_2) on October 1, 2012.
- Béresová, J., Breton, G., Noijons, J., & Szabó, G. (2011). *Relating language examinations to Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Highlights from the manual*. Retrieved from <http://www.ecml.at/tabid/277/PublicationID/67/Default.aspx> on August 15, 2013.
- Commission on the Development of Foreign Language Proficiency. (2011). Five proposals and specific measures for developing proficiency in English for international communication (provisional translation). Retrieved from [http://www.mext.go.jp/component/english/\\_icsFiles/afieldfile/2012/07/09/1319707\\_1.pdf](http://www.mext.go.jp/component/english/_icsFiles/afieldfile/2012/07/09/1319707_1.pdf) on October 6, 2012.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK.: Cambridge University Press.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*. Retrieved from <http://www.coe.int/t/dg4/linguistic/> on August 15, 2013.
- Council of Europe. (2011). *Manual for language test development and examining: For use with the CEFR*. Retrieved from <http://www.coe.int/t/dg4/linguistic/> on August 15, 2013.
- Hildén, R., Lehmann, M., Reichert, M., & Tardieu, C. (2011). *Common European framework of reference for languages — Level estimation grid for teachers*. Retrieved from <http://www.ecml.at/tabid/277/PublicationID/74/Default.aspx> on August 15, 2013.
- Martyniuk, W. (Ed.) (2010). *Aligning Tests with the CEFR: Reflections on using the Council of Europe's draft Manual*. Cambridge, UK: Cambridge University Press.
- Negishi, M., Takada, T., & Tono, Y. (2011). A progress report on the development of CEFR-J. Presentation at the Association of Language Testers in Europe 4th International Conference. Retrieved from <http://www.alte.org/2011/presentations/index.html> on December 17, 2012.
- Ohba, H., Yamakawa, K., Shimizu, Y., Sugino, N., & Nakano, M. (2012). What can Latent Rank Theory contribute to SLA research? The JACET International Conference Proceedings—The JACET 51st International Convention, pp. 156-163.
- Shojima, K. (2008). Neural test theory: A latent rank theory for analyzing test data. *DNC Research Note*, 08-01. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/Shojima2008RN08-01.pdf> on April 30, 2012.
- Shojima, K. (2009). Neural test theory. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New trends in psychometrics* (pp. 417–426). Tokyo: University Academic Press.
- Shojima, K. (2010). Neural Test Theory. In Ueno, M., Shojima, K. (Eds.), *Gakushu Hyoka no Shin-choryu [New trends in evaluation of learning]*, pp. 83–111. Tokyo: Asakura Shoten.
- Sugino, N., Yamakawa, K., Ohba, H., Shojima, K., Shimizu, Y., & Nakano, M. (2012). Characterizing individual learners on an empirically-developed can-do system: An application of Latent Rank Theory. In W.M. Chan, K.N. Chin, S. Bhatt, & I. Walker. (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 131–150). Boston, MA: De Gruyter Mouton.
- Tardieu, C., Hildén, R., Lehmann, M., & Reichert, M. (2010). *The CEF-ESTIM Grid*. Retrieved from <http://cefestim.ecml.at/> on July 28, 2013.
- Yamakawa, K., Sugino, N., Ohba, H., Nakano, M., Shimizu, Y. (2008). Acquisition of English grammatical features by adult Japanese EFL learners: The application of Item Response Theory in SLA research. *Electronic Journal of Foreign Language Teaching*, 5, pp. 13-40.