Linking Can-do Statements with Language Tests Using Neural Test Theory

Yukie Koyama* and Tetsuo Kimura**

**Nagoya Institute of Technology **Niigata Seiryo University*

Can-do Statements

- Common European Framework of Reference for Languages (CEFR) 2001
- to establish a goal of each level of language learning
- to enhance autonomous learners' self-awareness

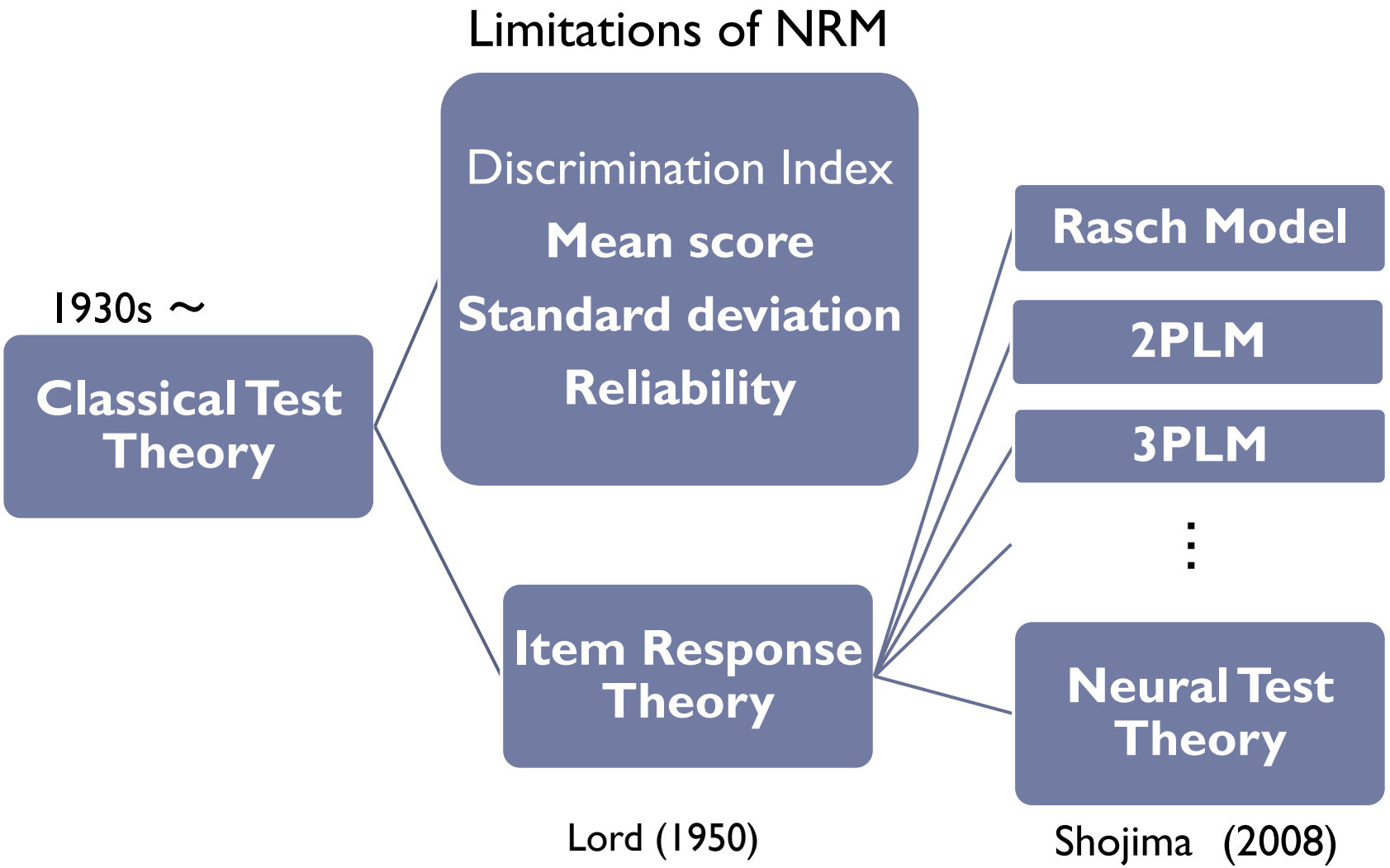
- Validity of the CDSs of CEFR : an online language test, DIALANG, as an external evaluation test

⟹ Linking CDSs and a test is important for the validation of the CDSs.

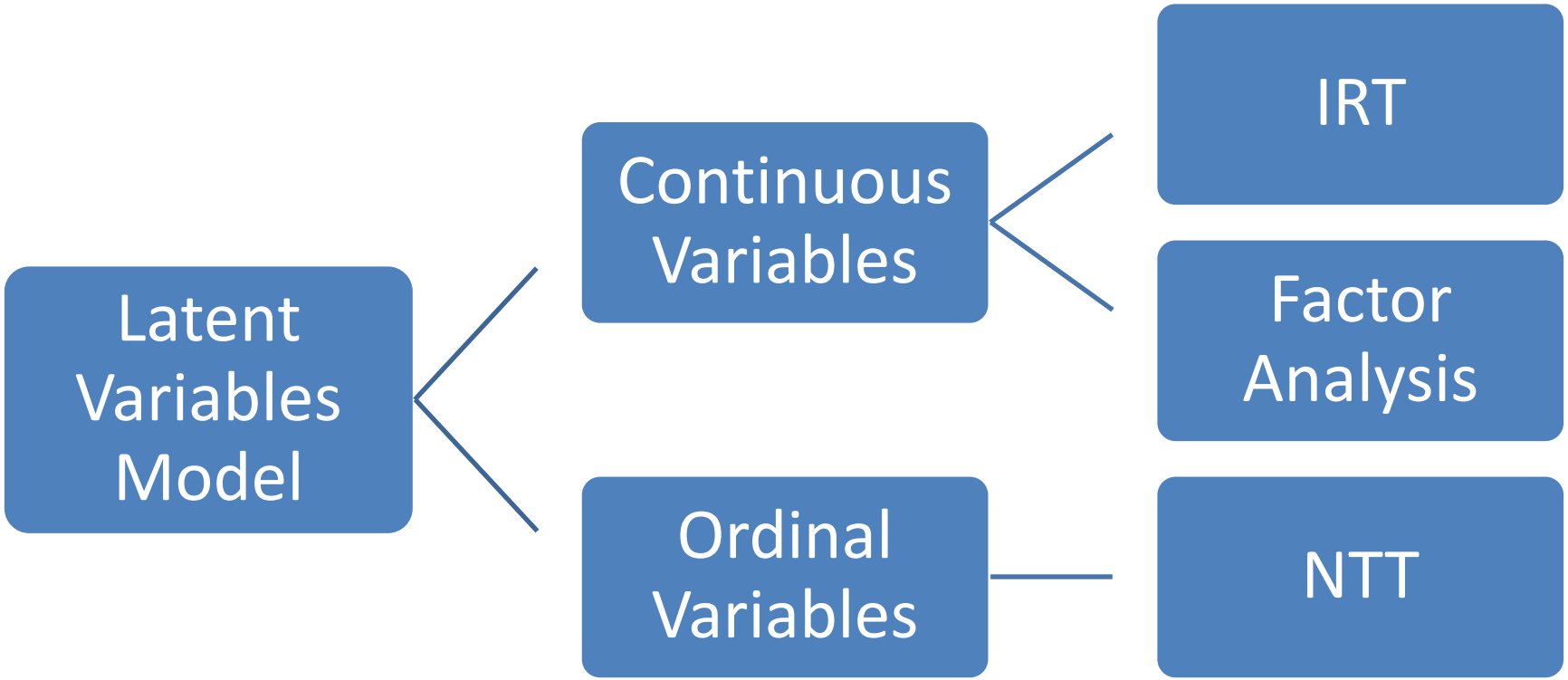
What is NTT?

- Not a tele-communication company

Testing Theory



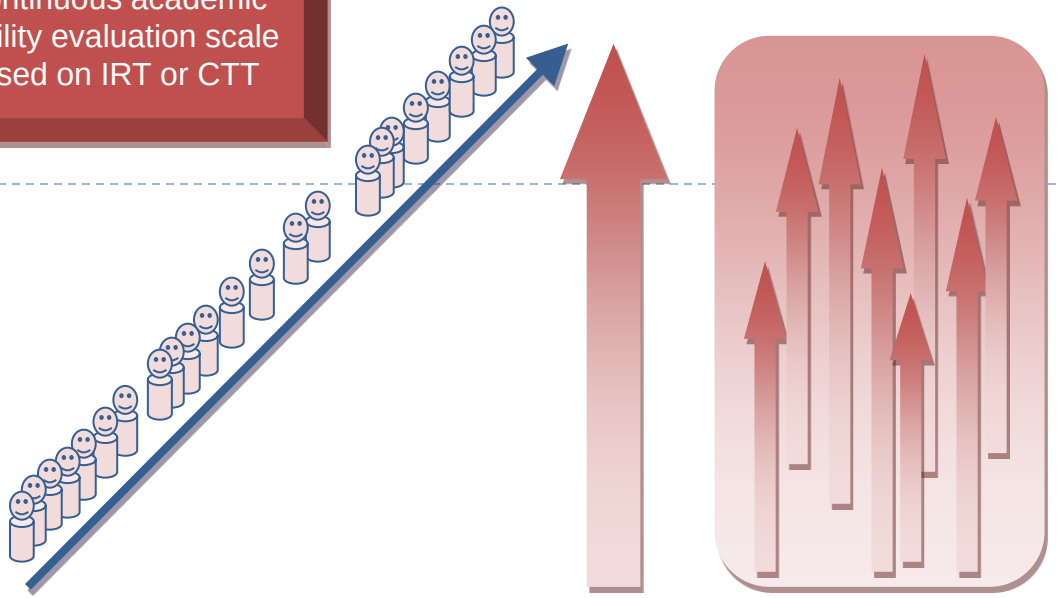
Latent Variables



IRT & NTT

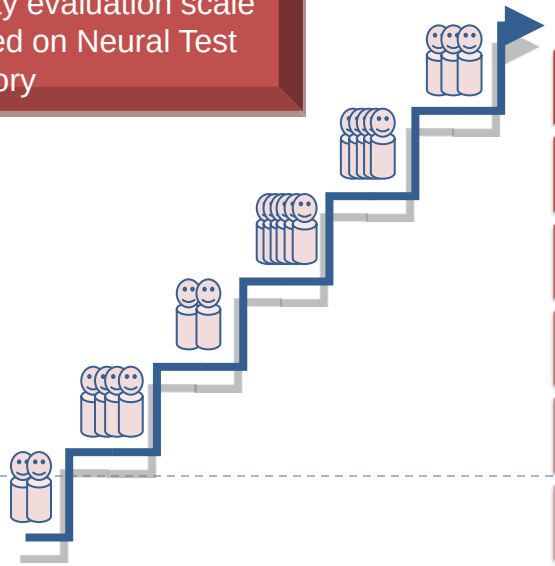
- IRT --- Continuous scale
- NTT --- Ordinal scale, Rank or Level,

Continuous academic ability evaluation scale based on IRT or CTT



For Qualifying Tests

Ordinal academic ability evaluation scale based on Neural Test Theory



- Ability to understand long sentences
- Ability to read charts and diagrams
- Conversation skills
- Grammar skills
- Vocabulary skills

Evaluation with an ordinal scale makes the evaluation less precise but improves accountability by making it easier to explain the academic ability corresponding to each grade. Along with this, the specific targets and achievements of the test become clearer.

Shojima (2009)

Objective of this research

- to examine the appropriateness of Neural Test Theory (NTT) in order to analyze Can-do Statements (CDSs) for the purpose of linking them with the corresponding language tests

Case Study 1 : EIKEN Can-Do List

- **Participants**: 220 (295) freshmen
  - engineering (64), nursing (54), and social welfare and psychology (102)

- **CDSs on the EIKEN Can-Do List**: 109

	R	L	S	W	Total
Grade Pre 1	6	4	4	4	18
Grade 2	5	5	7	6	23
Grade Pre 2	4	6	6	5	21
Grade 3	6	6	6	5	23
Grade 4	5	6	7	6	24
Total	26	27	30	26	109

Case Study 1 : EIKEN Can-Do List

- **Sample CDSs**
  - Reading
    - Can find streets, shops, and hospitals, etc. on simple maps written in English. [Grade 3]
    - Can understand the main points of lengthy texts (e.g. required readings and materials for lectures and training courses). [Grade Pre 1]
  - Listening
    - Can understand the meaning of simple instructions (e.g. "Open your textbook." / "Close the door, please."). [Grade4]
    - Can understand a speaker on the telephone, provided the content is simple (e.g. agreeing when to meet, taking short messages). [Grade Pre 2]

Case Study 1 : EIKEN Can-Do List

- **Sample CDSs**
  - Speaking
    - Can talk briefly about something that he/she is interested in (e.g. his/her hobbies, club activities). [Grade3]
    - Can make a complaint about products or services (e.g. about damaged products or unsatisfactory service). [Grade Pre 1]
  - Writing
    - Can write sentences using English word order, provided that the sentences are short (e.g. "I went to the park yesterday."). [Grade 4]
    - **G2**: Can describe the details of memorable experiences (e.g. school events, trips). [Grade 2]

Case Study 1 : EIKEN Can-Do List

- **Placement Test**
  - Vocabulary & Grammar (Vgm)
  - Listening comprehension with dialogue (Dlg)
  - Listening comprehension with monologue (Mlg)
  - Reading comprehension (Rdg)

  All the items were adopted from the EIKEN Test Grade 3, Grade pre 2, Grade 2 and Grade pre 1 in 2007 and 2008, under the permission of STEP.

  Four versions were constructed :equated with common anchor items that were calibrated in the previous study (Kimura, 2009).

Case Study 1 : EIKEN Can-Do List

- **Procedure**

	Model	Condition
CDSs	Dichotomous model NTT -SOM	The number of rank : 5 The target latent rank distribution: not specified
PT Vgm		
PT Listening (Dlg, Mlg)		
PT Reading	Graded model NTT-SOM	

NTT-SOM :NTT that uses a self-organizing map mechanism

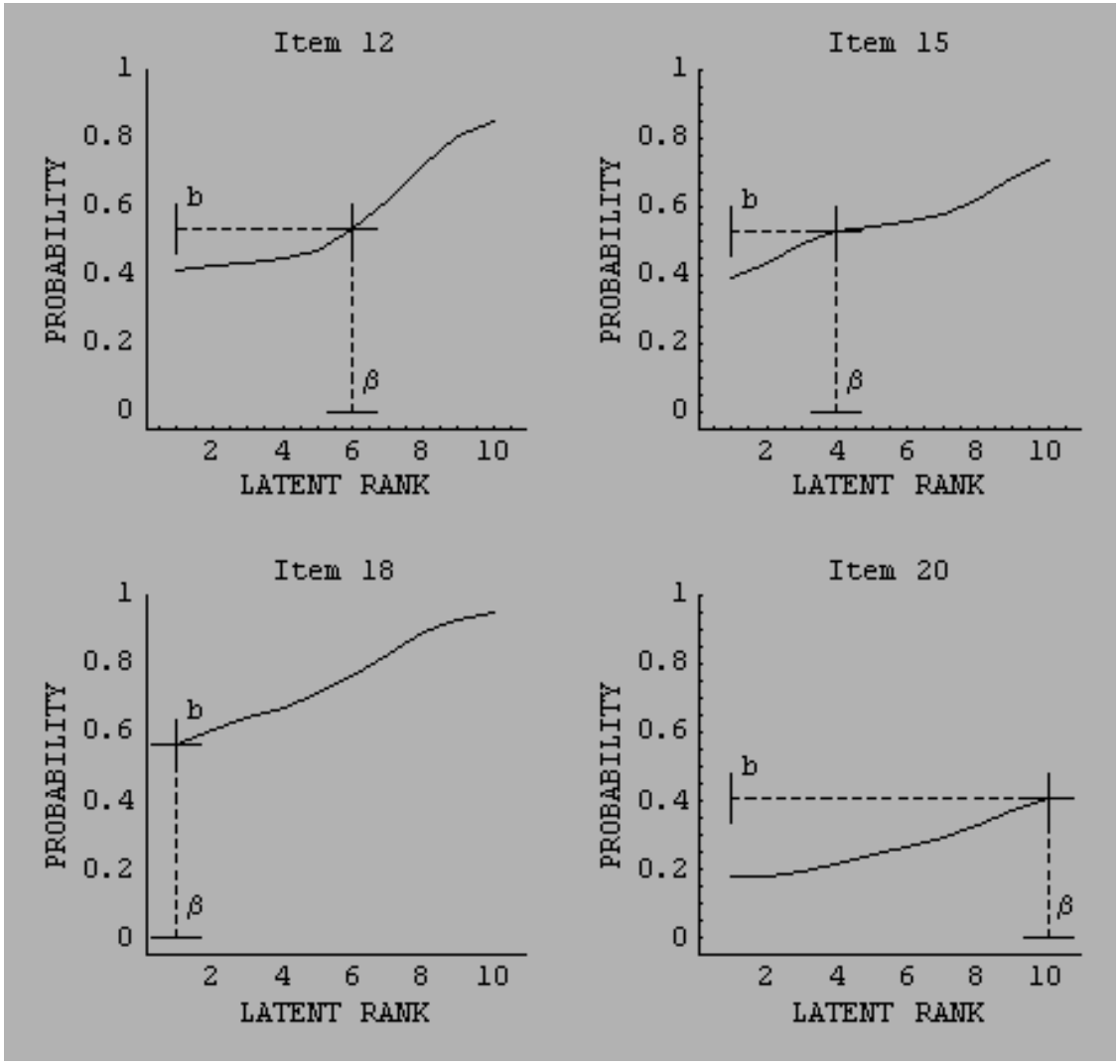
Software: Exametrika Ver.4.4 (Shojima, 2010)

Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - **CDSs' Difficulty (IRP index β) and EIKEN Grade**
    β is the location of the latent rank when the IRP value is closest to 0.5, simply expresses item difficulty.

  - **Spearman's rho (ρ)**
    - Reading ------- .93
    - Listening ------- .94
    - Speaking ------- .94
    - Writing --------- .95

Case Study 1 : EIKEN Can-Do List

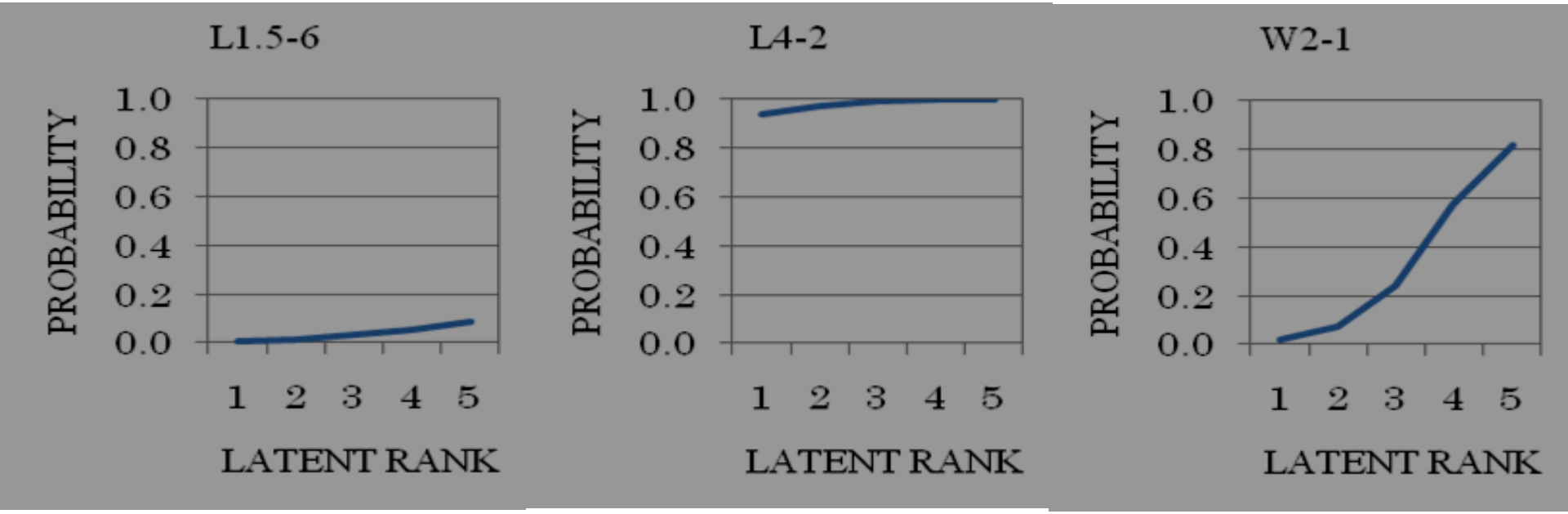


IRP index β and b

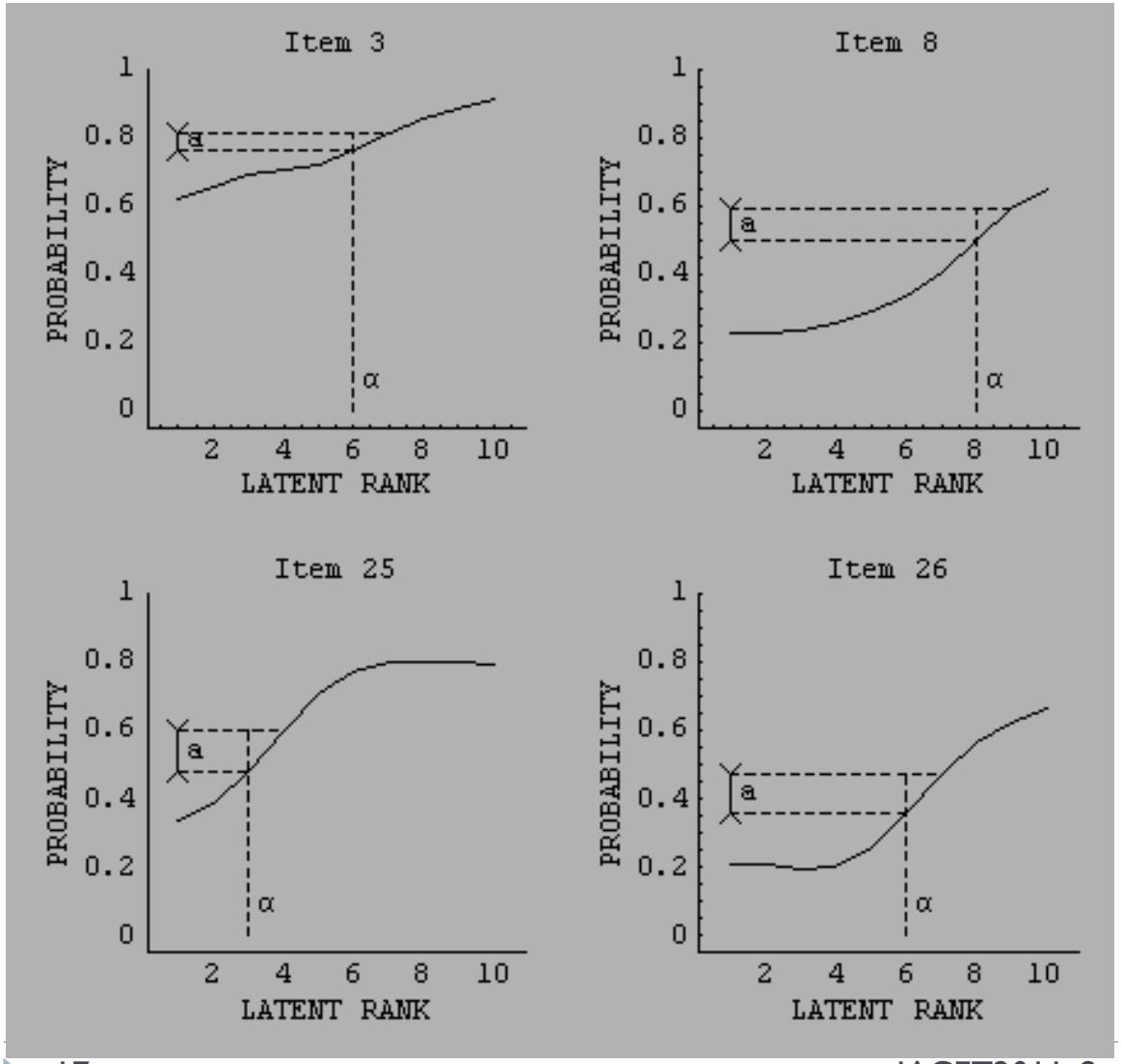
Kumagai(2007)

Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - **CDSs' Discrimination power (IRP index *a*) : .02-.34**
    Index *a* is the maximum difference in the IRP value among all adjoining rank pairs.



Case Study 1 : EIKEN Can-Do List

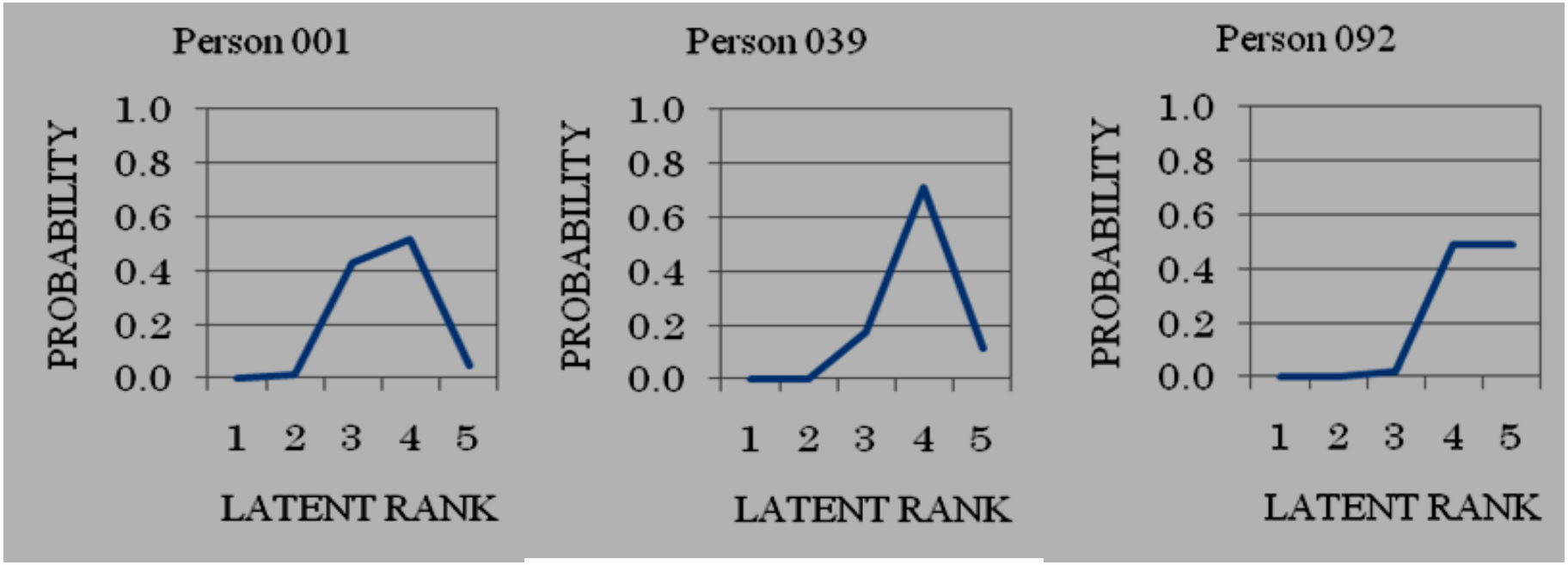


IRP index *α* and *a*

Kumagai(2007)

Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - The estimation of latent ability analyzed by NTT is best described in rank membership profile (RMP).

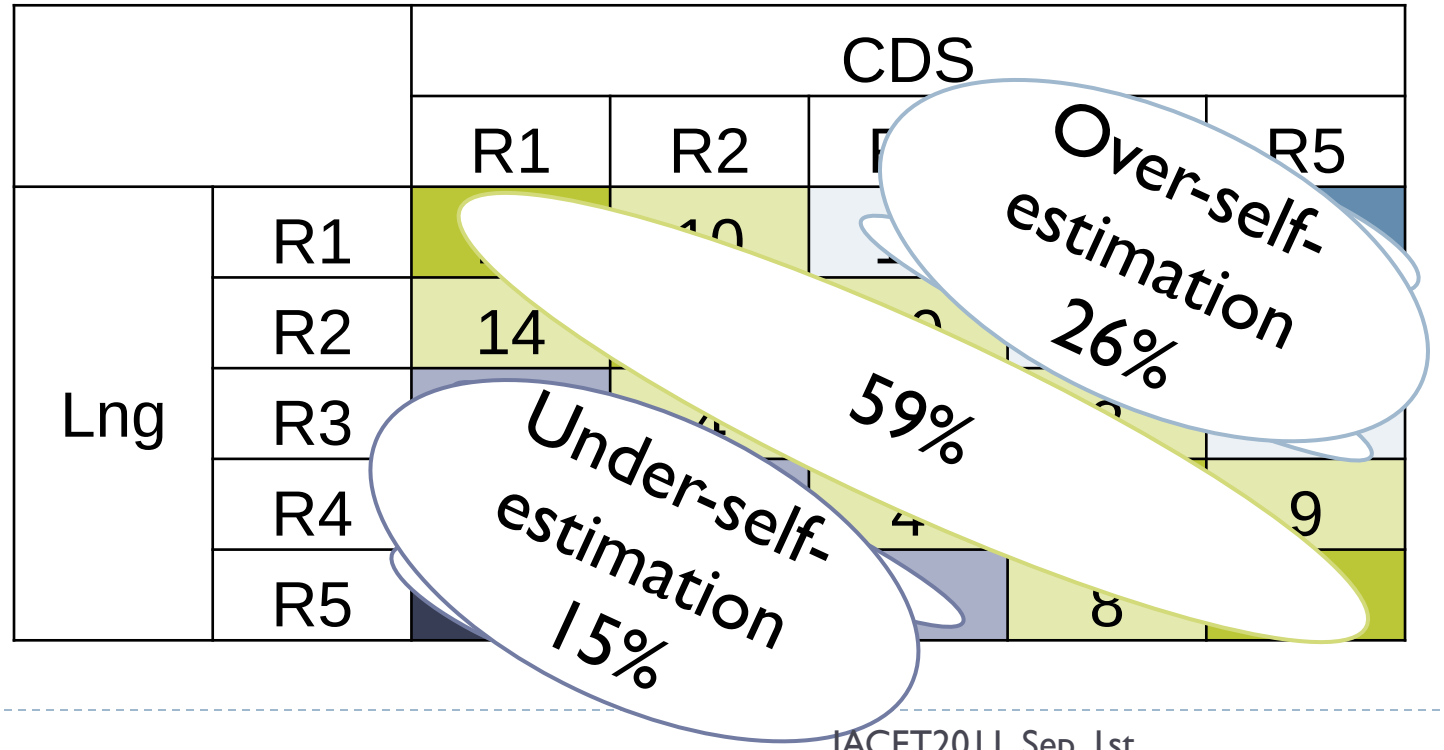


Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - **Self-evaluation based on CDSs and result of placement tests: Spearman's rho (ρ) was low**
    - CDSs Reading & PT Reading -------.22
    - CDSs Listening & PT Listening ---- .28
  - **Alpha reliability coefficient was high**
    - CDSs Reading ----- .87
    - CDSs Listening ----- .89
    - CDSs Speaking ----- .91
    - CDSs Writing ------- .92
    - PT Vgm ---------- .81
    - PT Listening ----- .88
    - PT Reading -------- .82

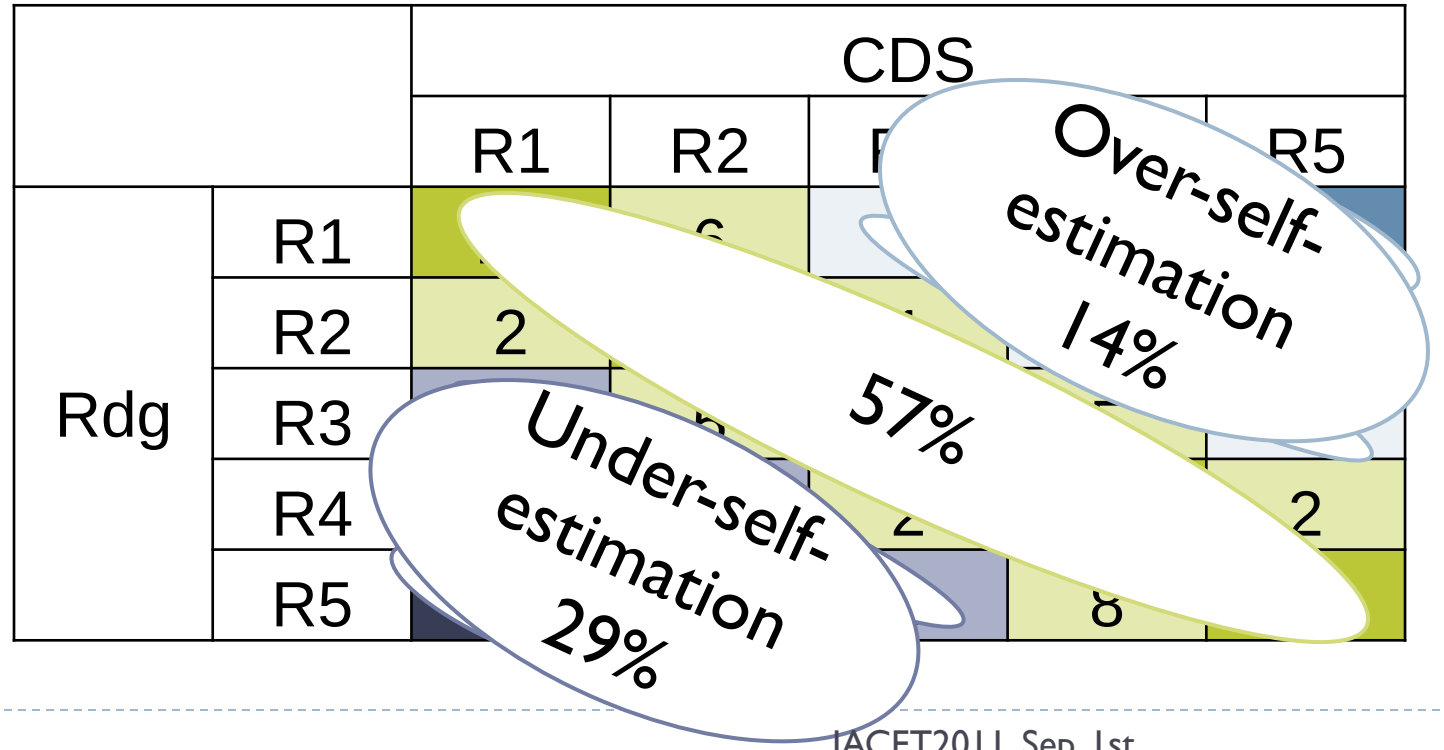
Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - **Self-evaluation based on CDSs and result of placement tests (Listening)**



Case Study 1 : EIKEN Can-Do List

- **Results and Discussion**
  - **Self-evaluation based on CDSs and result of placement tests (Reading)**



Case Study 1 : EIKEN Can-Do List

- **Conclusion**
  - **CDSs on the EIKEN Can-Do List are reliable**
    - Alpha reliability coefficient: ρ = .87 ～.92
    - CDSs' Difficulty (IRP index β) and EIKEN Grade are highly correlated: ρ = .93 ～.95
  - **Correlations between self-evaluation based on CDSs and result of placement tests are low:** ρ = .22 ～ .28
    - Many students over- or under-estimate their skills
    - More tend to over-estimate their listening skills
    - More tend to under-estimate their reading skills

Case Study II

- CDS : English for General Science & Technology

  Developed by English teachers of NIT

  4 skills　5 items per skill

  5 point scale　1. cannot perform at all

  ⟺ 5. can perform easily　（Koyama,2008）

- Test ： Final Examination of NIT EGST course (EXAM)

  100 items　Listening-- 30

  Reading, vocabulary, grammar -- 70

  multiple choice questions

Participants

- First year students of NIT

  CDS -- 882 Exam -- 942

- Conducted in Oct. 2007 Feb. 2008
- Mark sheet

NIT CDSs　version 2

e.g.

Reading

1. I can read and understand a menu.
2. I can read a manual of an apparatus and understand how to use it.
3. I can read and understand an abridged edition of a novel.
4. I can read and understand the details of homework.
5. I can read and understand a scientific article of a newspaper.

Procedure

- CDSs and EXAM (an external evaluation test of the CDSs) were analyzed using CTT, Rasch model and NTT

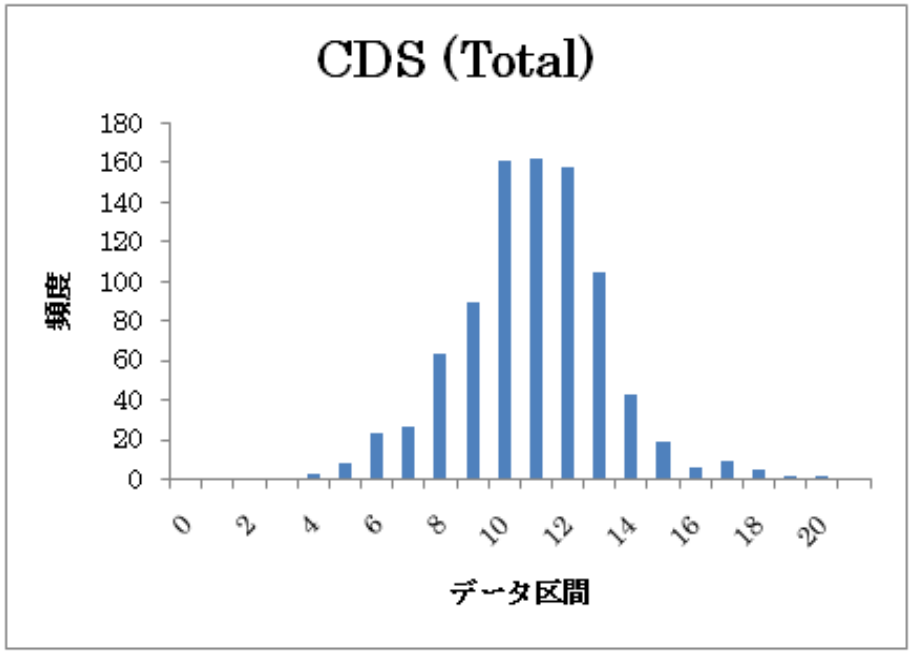
(Since the CDS data were polytomous in this case, the data were analyzed using Andrich's Rating Scale Model for Rasch model, and Graded Neural Test (GNT) model for NTT.)

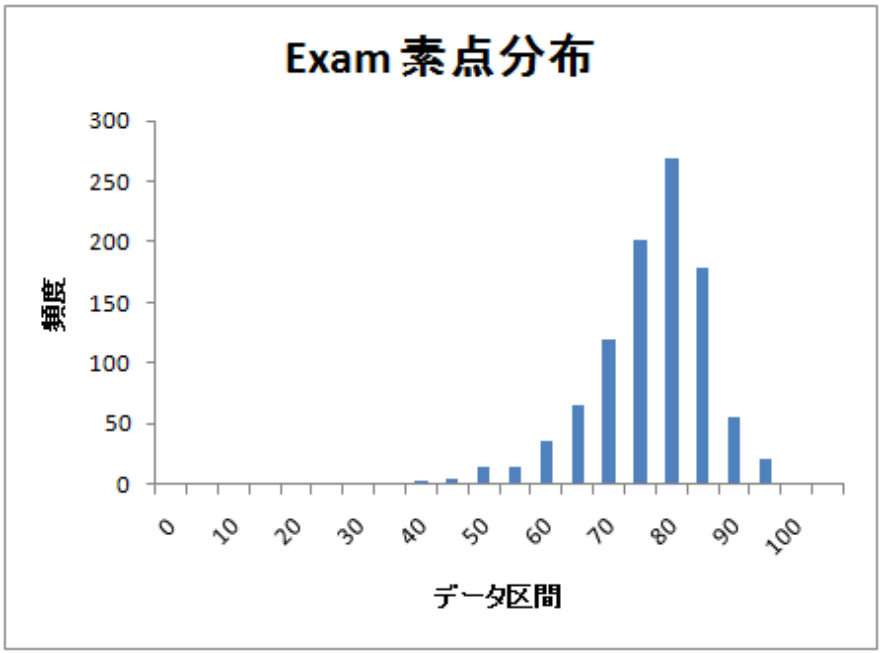
- Analysis of NTT -- Exametrika (Shojima, 2008)
- The number of ranks: 5 both for CDS and EXAM

Result 1. CTT (1)

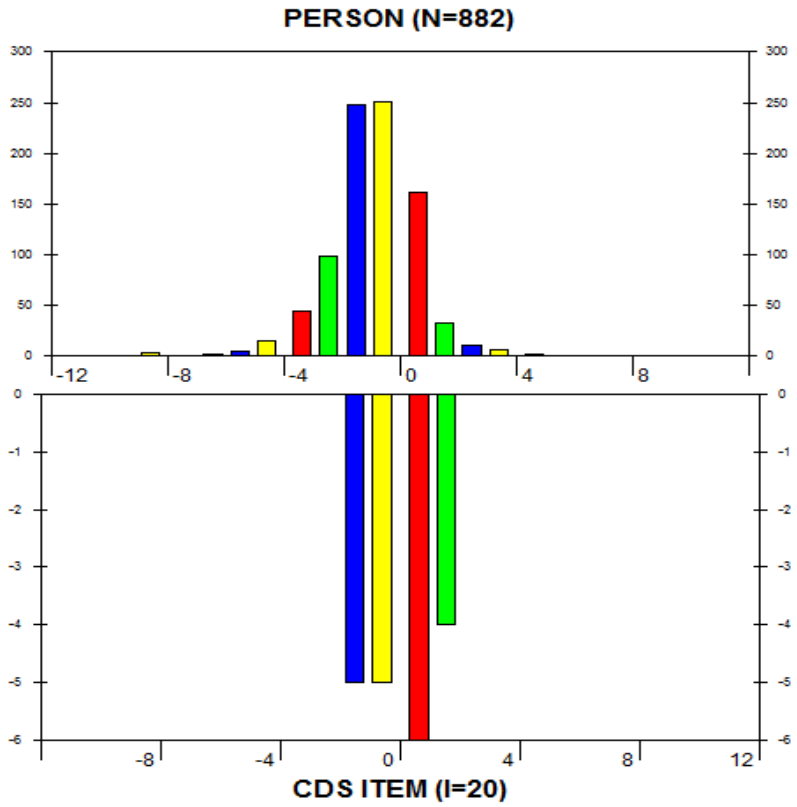
	Simple Statistics	
	CDS	EXAM
N of Examinees	882	942
N of Items	20	100
Min	20	33
Max	100	97
Median	53	76
Mean	52.375	74.606
Variance	130.532	86.685
Standard Deviation	11.425	9.310
Alpha Coefficient	0.936	0.834

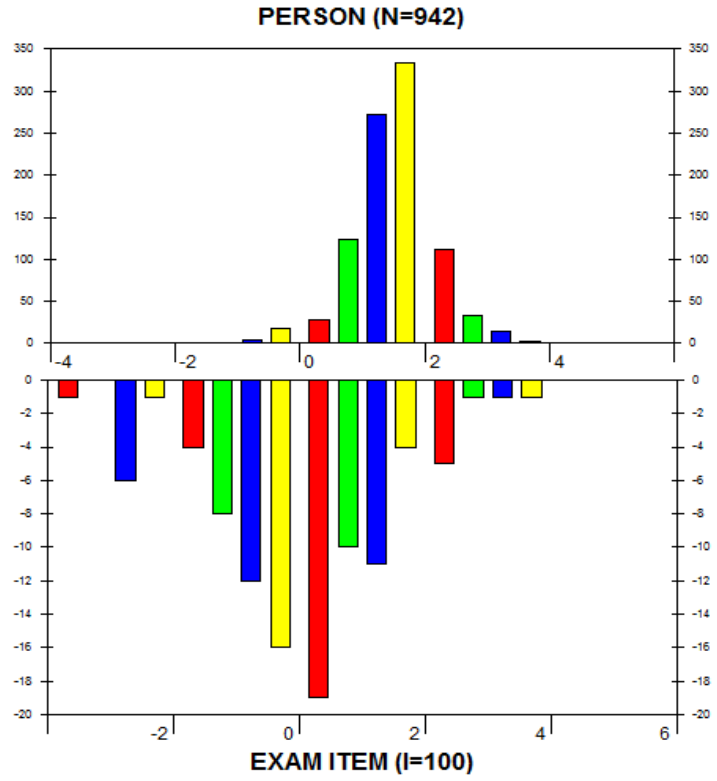
Result 2. CTT (2) Score Distribution



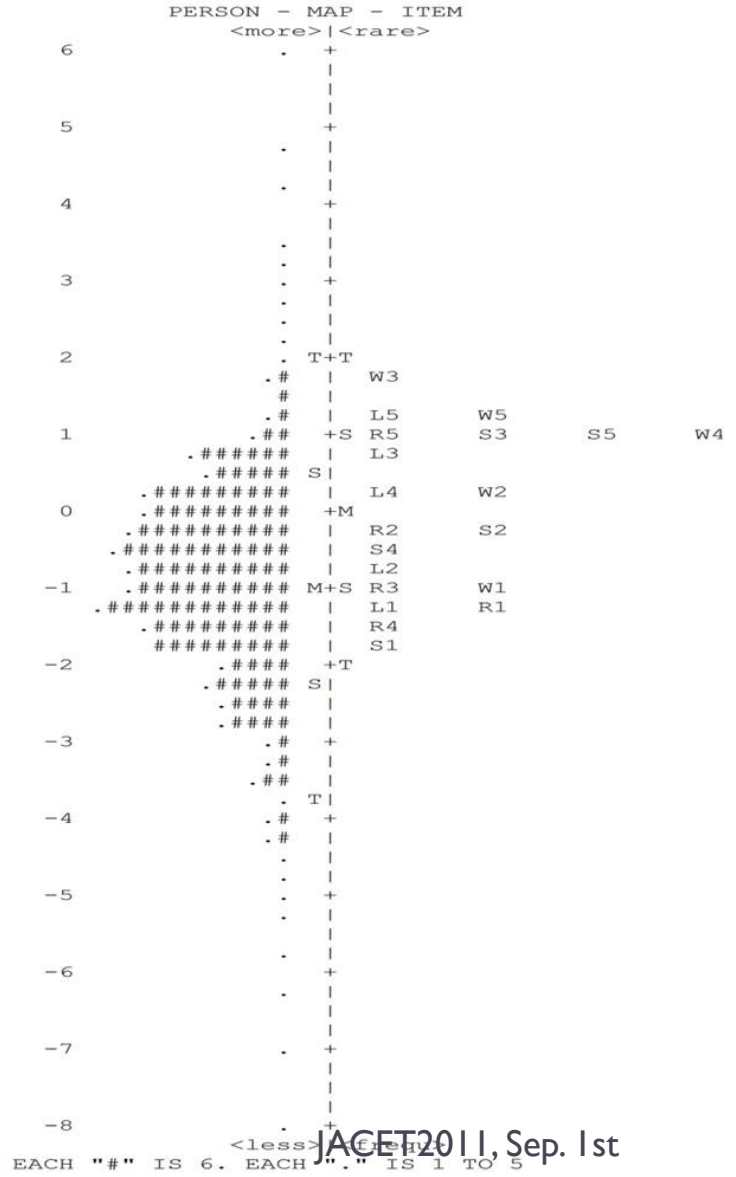


Result 3. Rasch model (1)

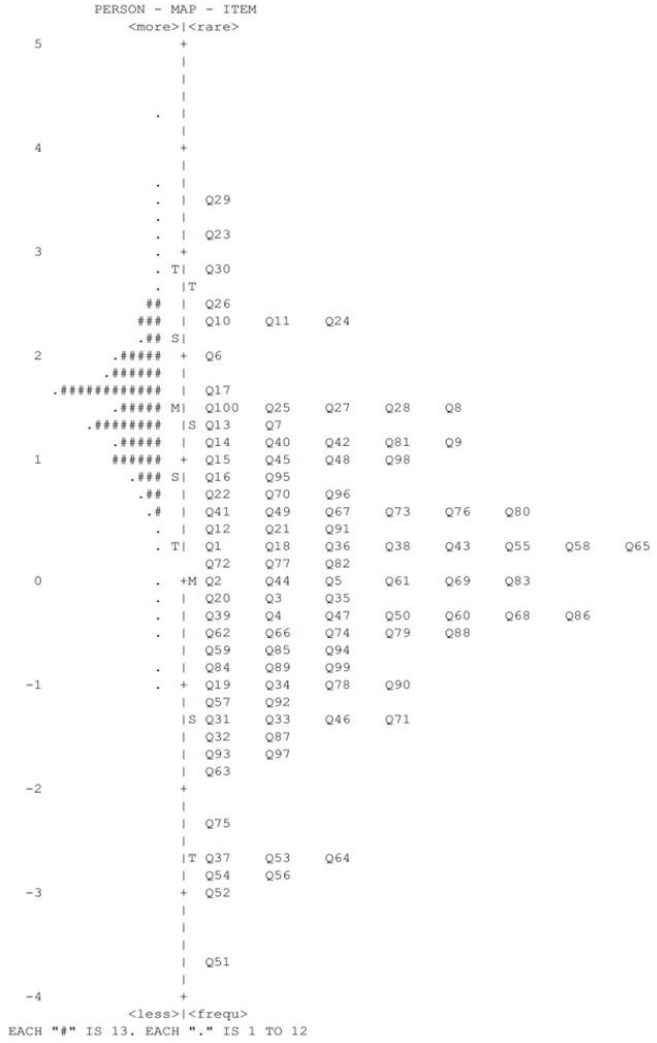




Result4. Rasch model (2)
Difficulty of CDSs & Ability of Examinees

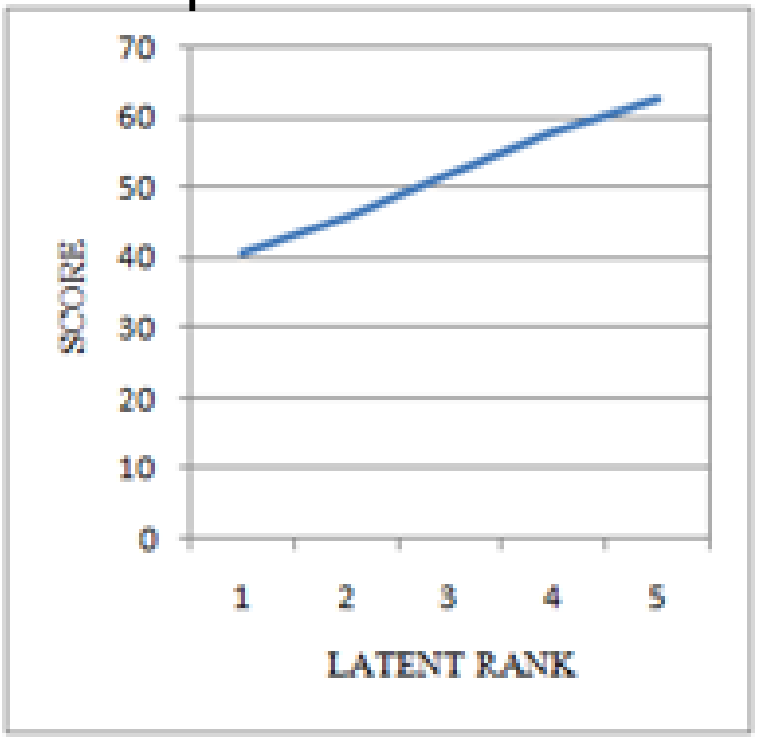


Result5. Rasch model (3)
Difficulty of EXAM & Ability of Examinees

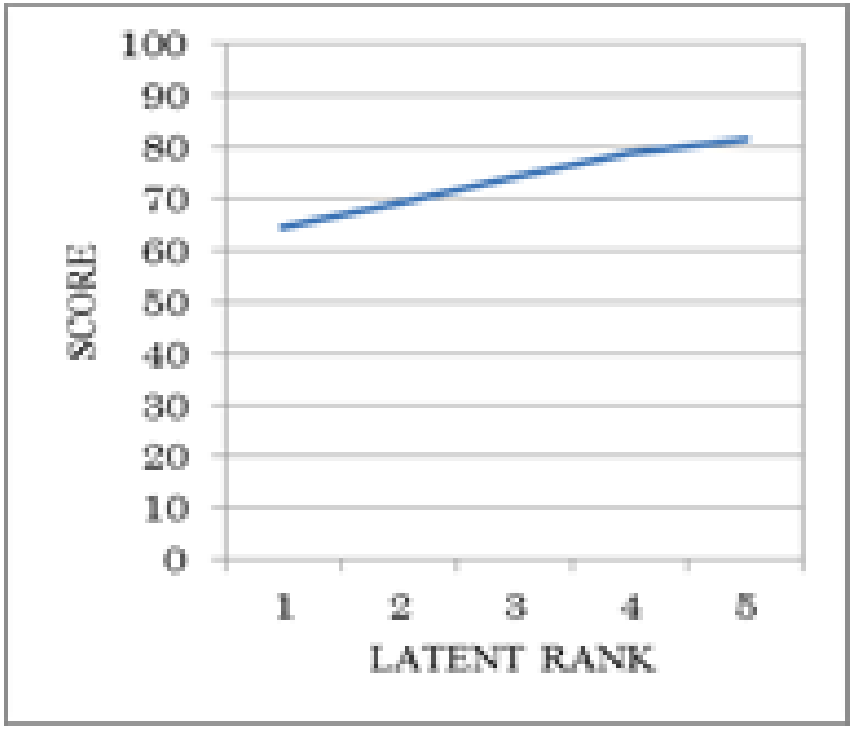


Test Reference Profile

the expected score of test-takers in each latent Rank



NTT--TRP of the CDS



NTT--TRP of the EXAM

EXAM is, on the whole, easier for the test-takers' levels.

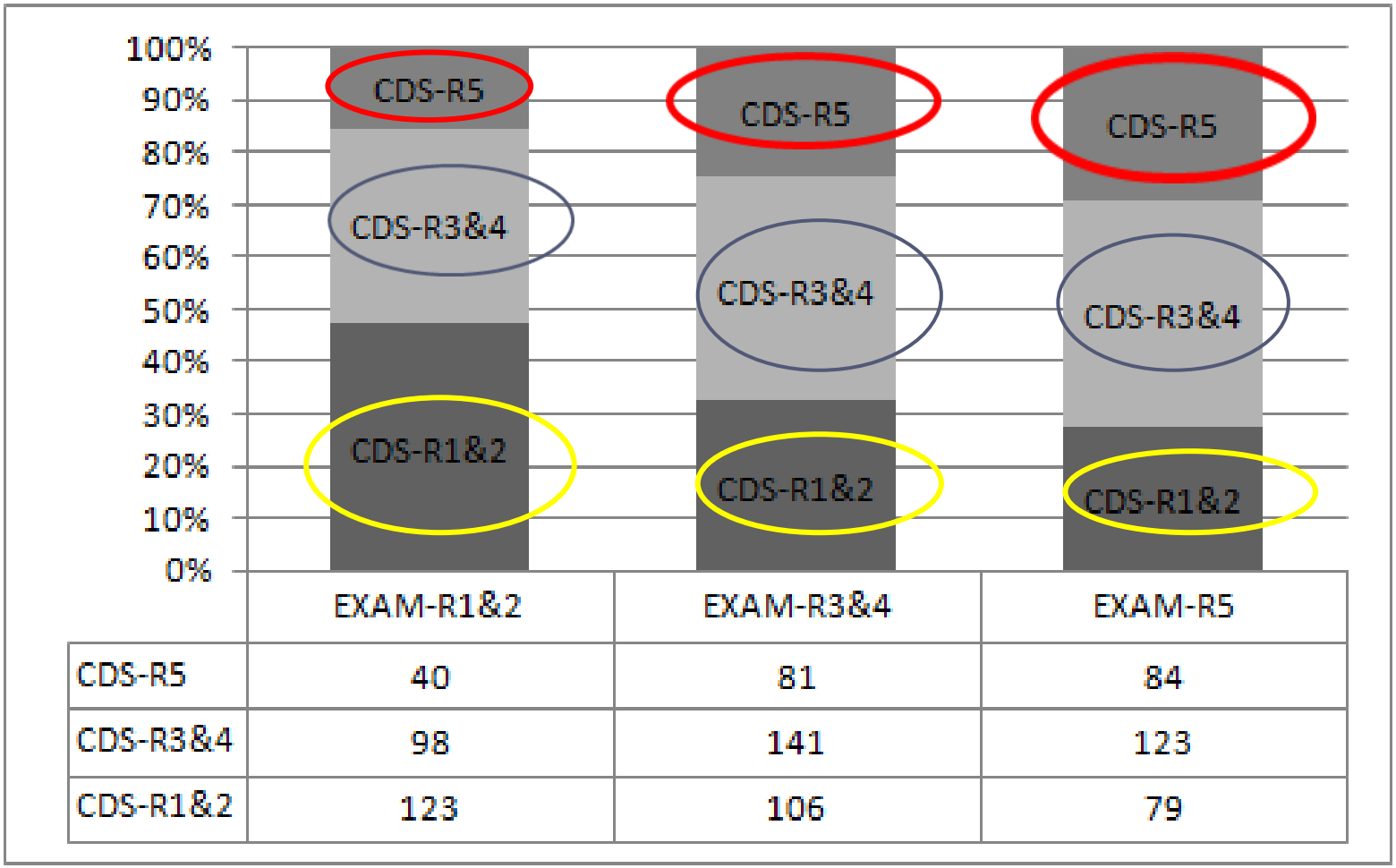
Cross-tabulation (Ranks of CDS x EXAM) 5×5

		CDS					
		R1	R2	R3	R4	R5	total
EXAM	R1	49	22	27	26	13	137
	R2	31	21	25	20	27	124
	R3	26	31	42	31	36	166
	R4	28	21	32	36	45	162
	R5	34	45	70	53	84	286
	total	168	140	196	166	205	875

Cross-tabulation (Ranks of CDS x EXAM) 3×3

		CDS			
		R1&R2	R3&R4	R5	total
EXAM	R1&R2	123	98	40	261
	R3&R4	106	141	81	328
	R5	79	123	84	286

Relation Between Ranks of CDSs and EXAM



	EXAM-R1&2	EXAM-R3&4	EXAM-R5
CDS-R5	40	81	84
CDS-R3&4	98	141	123
CDS-R1&2	123	106	79

Summary

- Compared with other test models based on the continuous scale, NTT, a test model for graded evaluation based on the ordinal scale, provides us clearer criteria for both CDSs and learners' levels/stages.
- We do not have to think about where to divide the results. Therefore, we can easily connect the learners' levels/stages to CDSs. In other words, we can easily link test results with CDSs when NTT is used for the analysis.

Limitations

- One of the limitations and disadvantages of NTT is that the methodology of item fit analysis or aberrant responses detection has not been developed sufficiently yet. In addition, cumulative of empirical data analysis based on NTT is not enough.
- However, as stated above, NTT has an essential advantage as a testing theory to analyze the test results to be classified into ranks such as those of CDSs.

References

- Alderson, J.C. (2005). *Diagnosing foreign language proficiency: the interface between learning and assessment.* New York: Continuum
- Dunlea, J. (2009). The EIKEN can-do list: improving feedback for an English proficiency test in Japan. In L. Taylor & C.J. Weir (Eds.), *Studies in language testing 31: Language testing matters,* 245-262.
- Kimura, T. (2009). Neural test riron niyoru eigo placementtest no sakusei to hyouka (Construction and evaluation of an in-house English placement test from a neural test theory perspective). *KATE Bulletin*, 23, 23-34.
- Kumagai, R. (2007, November). Neural test riron wo risanhennsuugata IRT tominasitatoki koumokutokusei wo simesu shihyou nituite (Item characteristic index and parameter estimation in NTT when considering discrete variable IRT). Paper presented at the workshop "Neural Test Theory."
- Koizumi, R. Iimura, H. (2010). Neural test riron no tokuchou: kotenteki test riron, Rasch modeltono hikaku (Characteristics of neural test theory: comparison with classical test theory and rasch modeling). *JLTA Journal, 13*, 91-109.

References

- Koyama, Y. (2008). Can-Do statements no datousei kenshou : ESP no kanten kara (Validity of can-do statements: an ESP perspective). *25th Anniversary Journal of JACET Chubu Branch 2008,* 177-187
- Linacre, J. M. (2009). WINSTEPS [Computer software]. Retrieved Februaly 16th, 2009, from http://www.winsteps.com/, originally developed by Wright, B.D., and Linacre, J. M. (1998). Chicago:MESA Press,
- Shojima, K. (2007). Neural test theory. DNC Research Note, 07-02.
- Shojima, K. (2009). Neural test theory model for graded response data. International Meeting of the Psychometric Society 2009 (Univ. of Cambridge, UK) P.112.
- Shojima, K. (2010). Exametrika (Version. 4.4) [Computer software]. Retrieved October 22, 2010, from http://www.rd.dnc.ac.jp/~shojima/exmk/
- Shojima, K. (2010). Neural test riron: gakuryoku wo dankaihyouka surutameno sennzai ranku riron (Neural test theory: latent rank theory for evaluation of academic ability). In Ueno, M. & Shojima, K. *Gakushu hyokano shin chouryu (New trend in learning evaluation)* (pp.83-111). Tokyo: Asakura Shoten.
- Society for Testing English Proficiency. (2008). *The EIKEN Can-do List.* Tokyo, Japan: The Society for Testing English Proficiency.