

第5回日本テスト学会賞受賞講演

2012.03.22.  
於：成蹊大学

# 外国語能力試験を改定するための 基礎研究

野口 裕之  
(名古屋大学)



# はじめに

- このたびは、第5回日本テスト学会賞をいただき、また、このような講演をする機会をいただき有難うございます。
- 今回の受賞は、テスト理論をテストの現場に普及し、現場で活かすことに努力したことが評価されたということです。
- 従って、この賞はこれまで一緒に「よりよいテスト」「国際標準を目指したテスト」の開発をめざして一緒に頑張ってきたみなさま方とともにいただいたものだと思っています。

- 本日は私が関係したテストの改定に関わる基礎研究を紹介しますが、そこで一緒に研究し、お名前を挙げた方々との協同作業の結果であることを特にここで申し上げておきたいと思います。
- また、お名前は挙げていませんが、テスト開発過程では多くの方々の努力の成果があります。みんなでいただいた賞であることを忘れることなく、今後もテスト理論をテストの現場に活かす努力を続けたいと思います。

# この講演の構成は・・・

- 1 外国語試験の紹介
- 2 日本語能力試験の紹介
- 3 日本語能力試験の因子分析的検討
- 4 日本語能力試験の級間共通尺度の構成
- 5 日本語能力試験における解釈規準開発の  
試み

です。3, 4, 5が改定のための基礎研究になります。

# おことわり

- 本日の話の中には、「日本語能力試験」に関わる現状や研究成果が出て来ますが、これらは日本語能力試験の開発・実施主体となる機関(国際交流基金日本語試験センター)や各委員会の正式な見解ではなく、私個人や研究に関しては連名発表者の個人研究者としての見解です。
- また、テストに関する機密保持のため、ご質問によってはお答えできないこともあります。

すみません……

# 外国語試験について・・・

## 外国語試験の例

### 英語

ケンブリッジ英検、IELTS  
TOEFL

Cambridge ESOL  
ETS

### 仏語

DELF-DALF, TCF

CIEP

### 独語

ゲーテ独語検定

Goethe-Institut

### 中国語

漢語水平考試

北京語言大学HSKセンター

### 韓国語

韓国語能力試験

国立国際教育院(NIIED)

- 最近の傾向

コミュニケーション能力を測定する

知識 から 技能 へ

IRTベースの得点化(が多い)

得点の時期間等化を実現する

測定結果の解釈規準に「～ができる」

という Can-do statements を用いる

CEFRとの関連づけを行なう

など

# 外国語試験の改定(訂)

- 応用言語学の理論的发展
- テスト理論の進化
- 情報技術の发展
- 外国語教育現場のニーズの変化

などを踏まえて常に改定される必要がある。



# 例えば、...

- TOEFLでは  
Paper based test ⇒ Computer based test ⇒  
Internet based test  
Listening, Structure, Reading, (Writing)  
⇒ Listening, Reading, Speaking, Writing
- ケンブリッジ英検では改定の根拠や過程が  
**Weir, C. and Milanovic, M. (Eds.) (2003).  
Continuity and Innovation: Revising the  
Cambridge Proficiency in English Examination  
1913-2002, Cambridge University Press. に  
まとめられている。**

# 日本語能力試験の場合は・・・

改定前

レベル

4級(初級前半)、3級(初級後半)、  
2級(中級)、1級(上級)

類と満点

文字・語彙(100点)

聴解 (100点)

読解・文法(200点)

実施回数

年1回 得点の年度間等化は実施していない

## 改定後

- レベル

N1 = 旧1級

N2 = 旧2級

N3 新設

N4 = 旧3級

N5 = 旧4級

- 得点区分と得点範囲

N1,N2,N3 言語知識(文字・語彙・文法)

読解

聴解

すべて、[ 0, 60 ]

N4,N5 言語知識(文字・語彙・文法)・読解

[ 0, 120 ]

聴解

[ 0, 60 ]

すべてに基準点が設定され、この点に達しない区分があった場合には、総合得点によらず不合格となる。

- 実施回数

年2回

得点は時期間で等化されている

## 参考までに

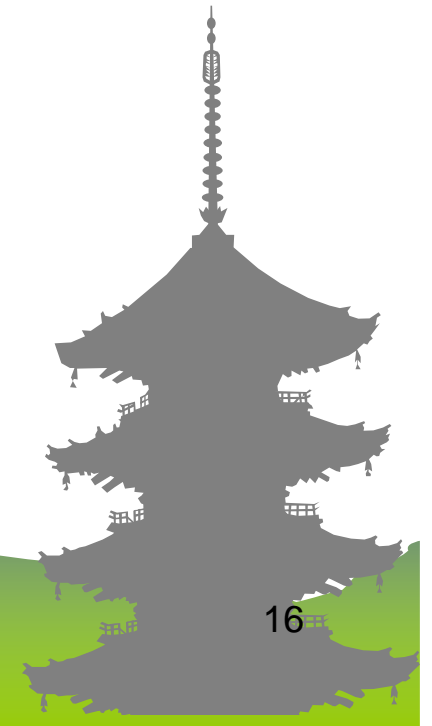
The numbers of learners of Japanese  
and examinees of the JLPT in 2009  
(Top 5)

2009	# of learners	# of examinees	%
Korea	964,014	121,675	12.6
China	827,171	327,255	39.6
Indonesia	716,353	8,211	1.1
Australia	275,710	1,143	0.4
Taiwan	247,641	17,045	6.9



Research 1

# I 日本語能力試験の 因子分析的検討





This research was presented at AILA2008

**Noguchi, H., Kumagai, R., Osumi, A., & Wakita, T.** (2008). Comparing factor structures of the Japanese Language Proficiency Test : differences in factor structure with increasing language proficiency by native language.

15th World Congress of Applied Linguistics (AILA 2008), Essen, Germany.

# この研究の目的は・・・

- 1 日本語能力試験(改定前)がどのような因子構造を持つか確認する。
- 2 日本語能力試験(改定前)の受験者を母語グループ別に分けた時に因子構造に違いがあるかどうかを確認する。

何故、母語グループを問題にするのか？

## Japanese Language notation system

Kanji (漢字, Chinese ideograms)

+

Hiragana (ひらがな, Japanese phonograms)

Katakana (カタカナ, Japanese phonograms)

- 受験者の母語における漢字表記システムの違いが、JLPTの因子構造の違いとして反映するかもしれない。
- 漢字表記システム (Kanji notation system)

Chinese	◎	almost all notation
Korean	△	rarely
Others	×	don't use
Japanese	○	

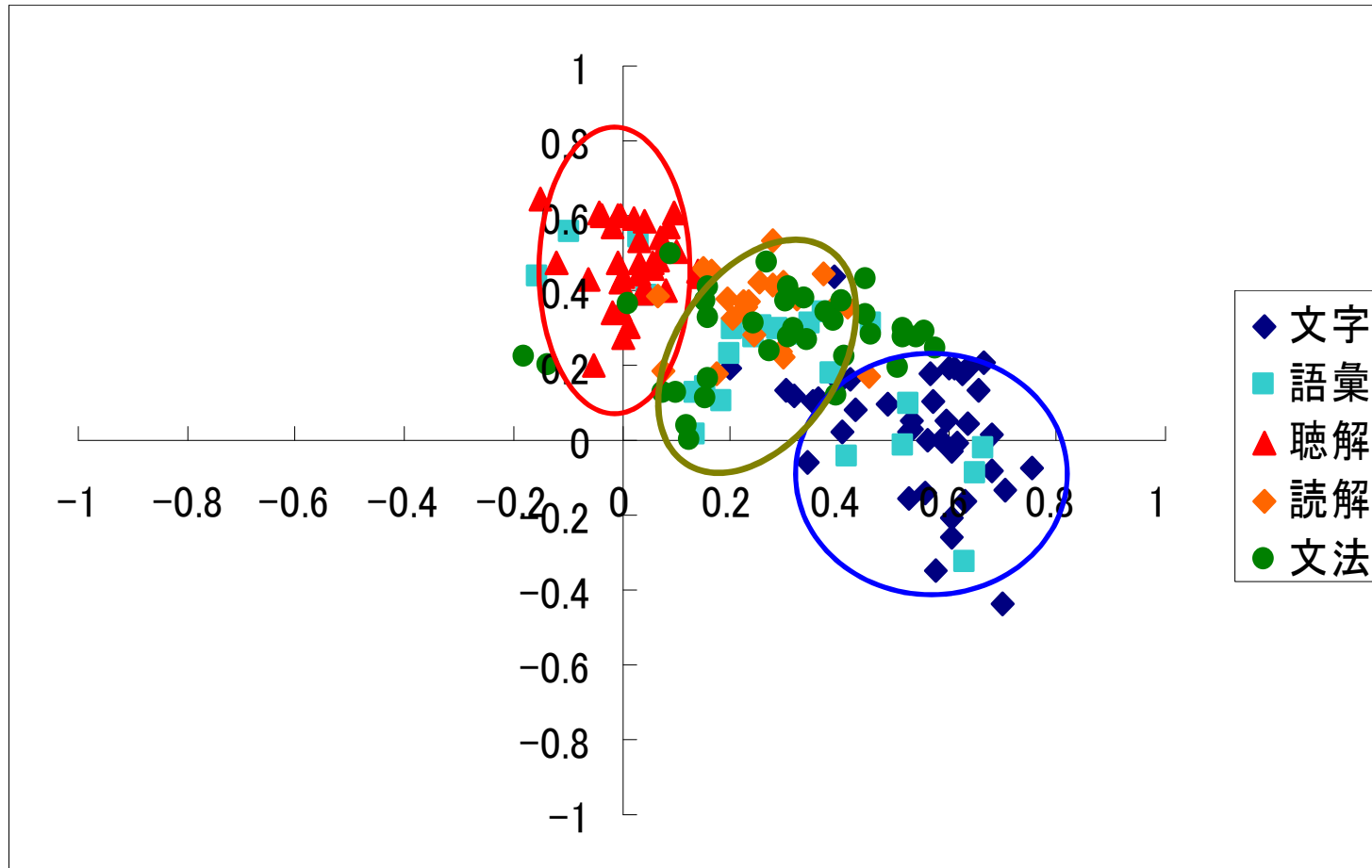
- **テスト:** JLPT 2001, 2002, 2003
- **母語グループ:** 中国語、韓国語、その他
- **分析に用いた受験者数 (JLPT2003)**

<b>レベル</b>	<b>中国語</b>	<b>韓国語</b>	<b>その他</b>
<b>1級</b>	<b>50909</b>	<b>26975</b>	<b>6666</b>
<b>2級</b>	<b>42740</b>	<b>20995</b>	<b>13423</b>
<b>3級</b>	<b>31415</b>	<b>17688</b>	<b>36596</b>
<b>4級</b>	<b>16747</b>	<b>4019</b>	<b>18497</b>

# 日本語能力試験の因子構造は・・・

- 2003年度1級受験者84550名のデータを全項目を合わせて因子分析  
→ スクリーンプロットから **2因子構造**
- 因子負荷をプロットすると、座標平面の第I象限にほぼすべての項目が存在  
**右下**部分に「**文字**」項目、  
**左上**部分に「**聴解**」項目、  
その**中間**に「**読解**」「**語彙**」「**文法**」項目  
が集まって、まとまりを形成している。

# 日本語能力試験の因子分析結果 —2003年度1級—



- 2つの因子軸に対して、問題項目は  
「文字」  
「読解・語彙・文法」  
「聴解」  
の3クラスター(まとめり)を形成している。



## 2つの因子は・・・

- 1)「漢字がもたらす情報を  
検索処理する能力」を表す因子  
(漢字情報処理因子)
- 2)「文脈を活用して理解を構築する能力」  
を表す因子  
(文脈情報処理因子)

# 漢字情報処理因子

- 日本語能力試験では、  
「文字(漢字)」の能力を単なる「形態的知識」として測定しているのではなく、「部首、偏、旁などの字形」「音」「意味」「語彙性」「文法性」その他の情報を利用して適切な漢字および語彙を検索する

高度な情報処理過程を測定していることの反映

- 言い換えると、  
「文字（漢字）」は  
言語知識という側面もあるが、  
言語処理の側面が強いと言える。

## そうだとすると・・・

- 日本語能力試験における  
「漢字力」の定義を明確にする必要がある
- CEFRなどの言語能力基準にJLPT  
を関係づける際には、欧州系の言語にはない漢字情報処理因子に注意する必要がある 補助尺度の必要性？

# 中国語との関係

- 同じく漢字を用いる中国語との関係を考えると  
中国語はすべて漢字で表記  
日本語は漢字かな混じりで表記
- この点で「文字(漢字)」の言語処理的な側面が異なり、情報処理過程を反映した異なる能力基準が想定される可能性もある。

受験者を母語グループ別に因子分析すると・・・

JLPT2001,2002,2003 に共通して

	中国語	韓国語	その他
1級	1因子	2因子	2因子
2級	1	1	2
3級	1	1	2
4級	1	1	1

# すなわち・・・

- 各母語グループは日本語能力のレベルに対して独自の因子構造の変化を示す。
- 初級前半(4級) : 各母語グループとともに共通の1因子性を示す。
- 初級後半(3級) : 「その他」グループが「中国語」「韓国語」グループと異なり、2因子性を示す。

- 中級(2級): 初級後半(3級)と同じく、「その他」グループが「中国語」「韓国語」グループと異なり、2因子性を示す。
- 上級(1級): 「その他」グループに加えて、「韓国語」グループも2因子性を示す。
- 「中国語」グループは4つのレベルを通して1因子性を示す。



## この研究で得られたことは・・・

- CEFRLなどの言語能力基準にJLPTを関係づける際には、欧州系の言語にはない漢字情報処理因子に注意する必要がある

### 補助尺度の必要性を検討する

- JLPTの改定にあたって、「中国語」「韓国語」以外の言語を母語とする日本語学習者の因子構造の発達的变化に配慮する必要がある。



## Ⅱ 日本語能力試験の 級間共通尺度の構成



この研究は下記の学会誌および  
学会で発表したものです。

野口裕之・熊谷龍一・大隅敦子 (2007).

日本語能力試験における級間共通尺度構成の試み 日本語教育, 135, 54-63.

野口裕之・熊谷龍一・大隅敦子・石毛順子

(2006). 日本語能力試験における級間共通尺度構成の試み 日本語教育学会2006年度秋期大会予稿集.

# 研究の背景

- 改定前の日本語能力試験では毎年分析報告書が刊行されて来た (1984-2009)が、4つの級が独立に分析された結果が報告されていた。
- 従って、異なる級間で項目の識別力や困難度を比較することが出来なかった。
- また、異なる級を受験した学習者の得点を相互に比較することができなかった。

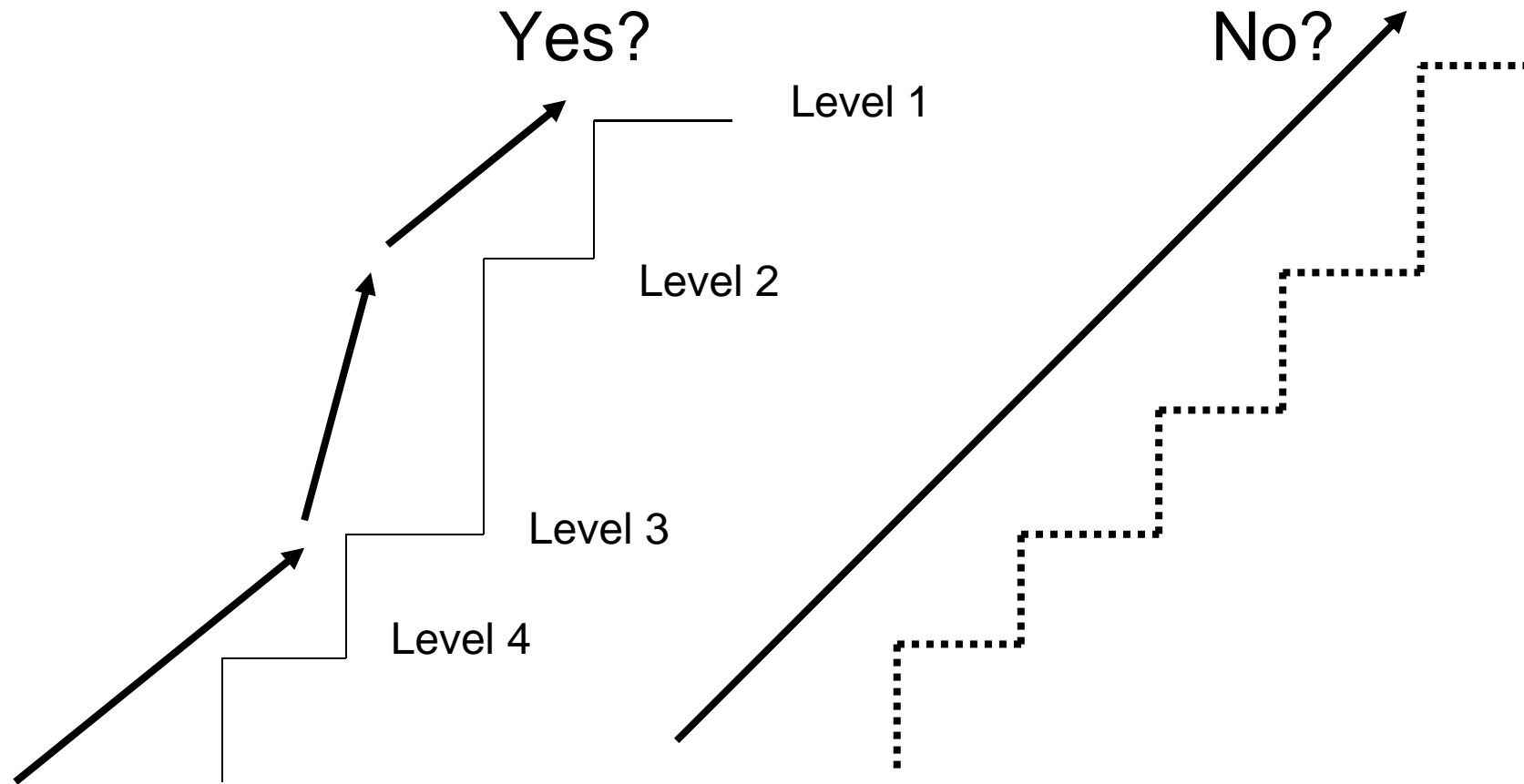
- 一方、日本語教育界で言われてきた重要な疑問

「3級合格から2級合格までの間に大きなギャップがある」

があり、

- 特に海外の非漢字圏でよく言われて来た。  
「大学学部の日本語専攻の学生でも、1年間日本へ留学しないと2級に合格できない」

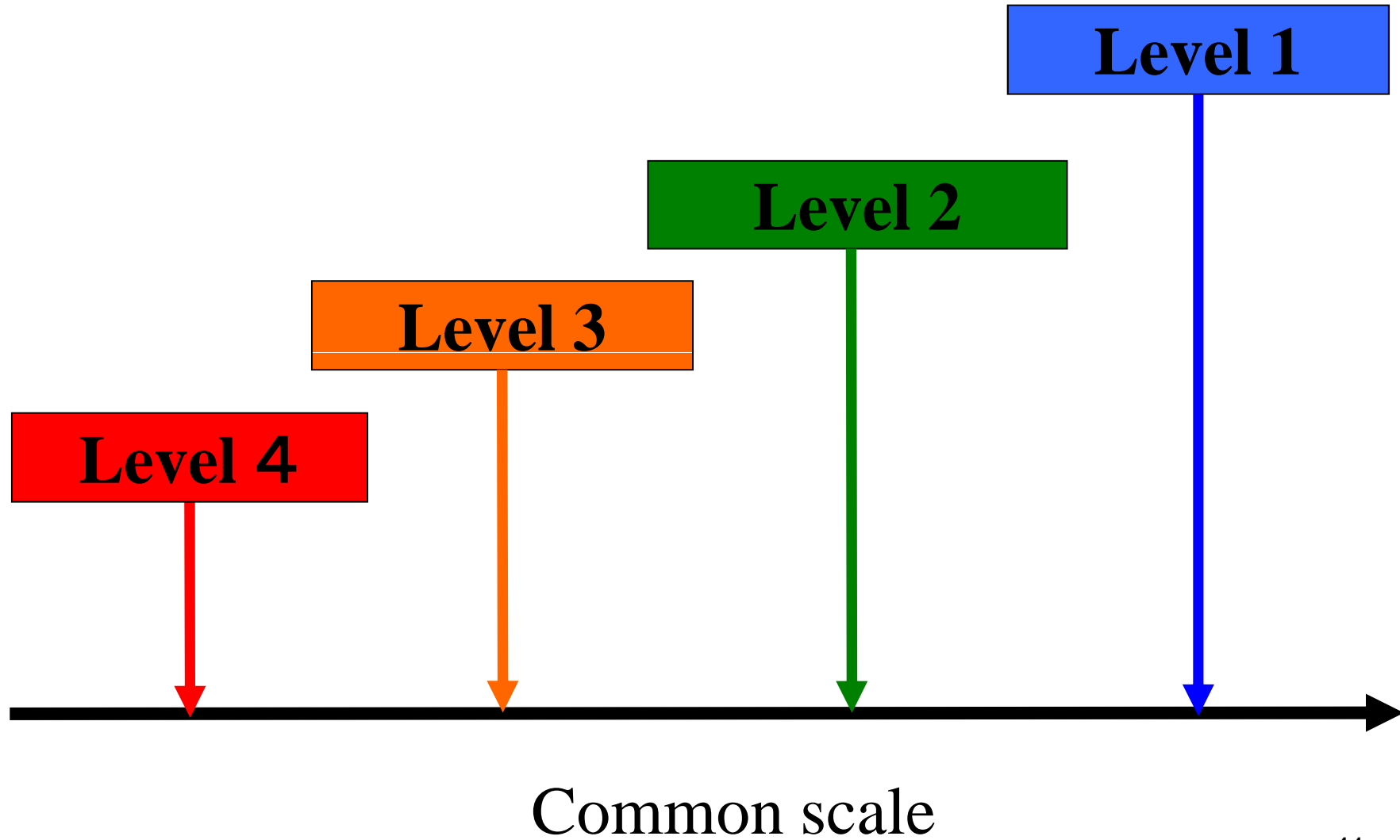
3級と2級との間に大きなギャップがあるか？



- そこで、改定後のJLPTで新しいレベルを入れるかどうか重要な決定事項になった。
- そのことに対して根拠のある(evidence based)解答を出すために、  
    **級間共通尺度を構成し、**  
    その上に4つの級の試験を位置づける  
    ことを試みた。  
    ⇒ **Vertical scaling**



# Vertical scaling



# この研究の目的は・・・

JLPTの4つの級間に共通尺度を構成することにより、

- 1) 異なる級間で項目の困難度を比較する、
- 2) 各級の合否分割点間の距離を確認する、
- 3) 3級と2級との間にギャップがあるのか否かを確認する、

そして、その結果を改定後のJLPTのレベル構成に反映させる。

# 方法 1

- 1) 共通尺度を構成するために、アンカーテスト・デザイン **anchor-test design** を採用し、3つのモニター・テスト **monitor tests** を構成した。

2001年度JLPTで実際に使われた項目を用いて

Monitor Test A ⇐ 1級および2級

B ⇐ 2級および3級

C ⇐ 3級および4級

ここで、各モニター・テストが

級間のアンカー・テストになっている。

- モニター・テストの項目を選択するに際しては項目困難度が適度に散らばるように項目の内容および形式がもとのJLPTを反映するように配慮した。  
分量はもとの60～100%  
⇒この結果から2001年度の1級から4級のすべての**問題項目の困難度**および**受験者の特性尺度値(能力)**が共通尺度上で相互に比較できるようになる。

## 方法 2

2) モニター・テストは、2006年3月に日本国内で実施した。

この実験の協力者は、2005年12月に実施されたJLPTの受験者の中から募集し、母語のバランスがJLPTとほぼ同じになるように配慮して依頼した。

## モニター試験の実施概要

- 2006年3月国内で実施
- 2005年度受験者の中から受験者を募集

モニター 試験	2005年度 受験級	人数
A	1,2級	326名
B	2,3級	380名
C	3,4級	362名

## 母語別比率

母語別	人数	比率
中国語話者	393人	36.8%
韓国語話者	132人	12.4%
インドネシア語話者	127人	11.9%
その他	416人	39.0%
合計	1068人	100.0%

## 方法 3

- 3) 共通尺度を構成するために、  
まず、各級、各類毎に2001年度JLPT受験者データから **IRT** (2-Parameter Logistic Model) **scale** を構成した。

全部で、4 水準 × 3 類 ⇒ 12 のIRT 尺度

この段階では、項目パラメタの値は相互に比較できない。



# 方法 4

4) 次に、モニター・テストの結果を用いて等化係数 **equating coefficients** を推定する。

1級 を 2級 に 等化 ⇐ Monitor test A

3級 を 2級 に 等化 ⇐ B

4級 を 3級 に 等化 ⇐ C

等化係数の推定には **the mean-sigma method** (Marco, 1977) を用いた。

# 方法 5

5) 最後に、これらの等化係数推定値を用いてJLPT2001の全ての項目を類(section)毎に共通尺度上に位置づけた。

\*共通尺度の原点と単位は2級の原点と単位に合わせた。

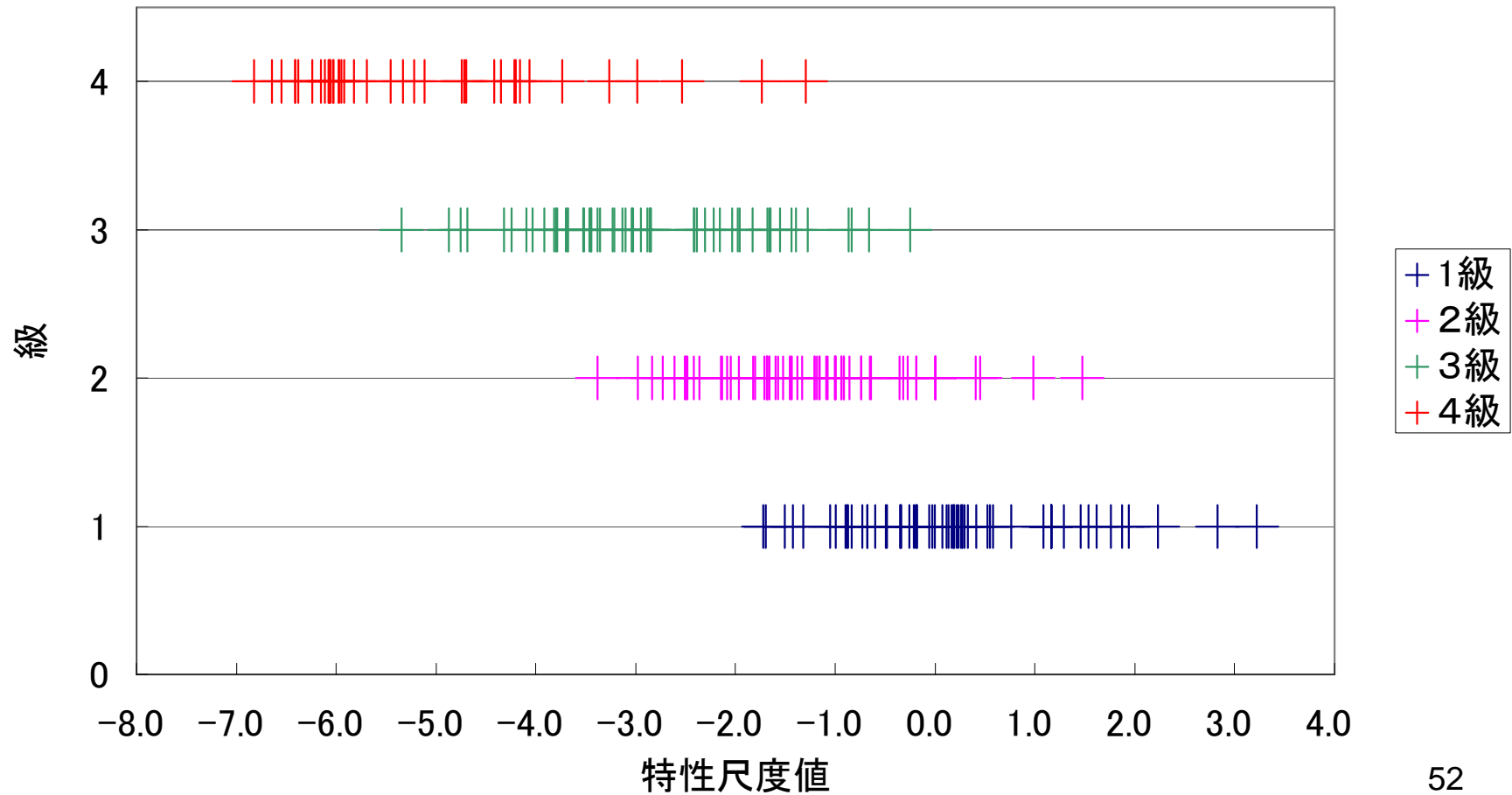
すなわち、各項目の識別力パラメタと困難度パラメタを各類毎に共通尺度上の値に変換して表わした。

# 結果 1

- 1) 項目困難度の分布状況を  
文字・語彙  
聴解  
読解・文法  
の各類毎に見ると・・・

# 文字・語彙

文字・語彙項目困難度分布



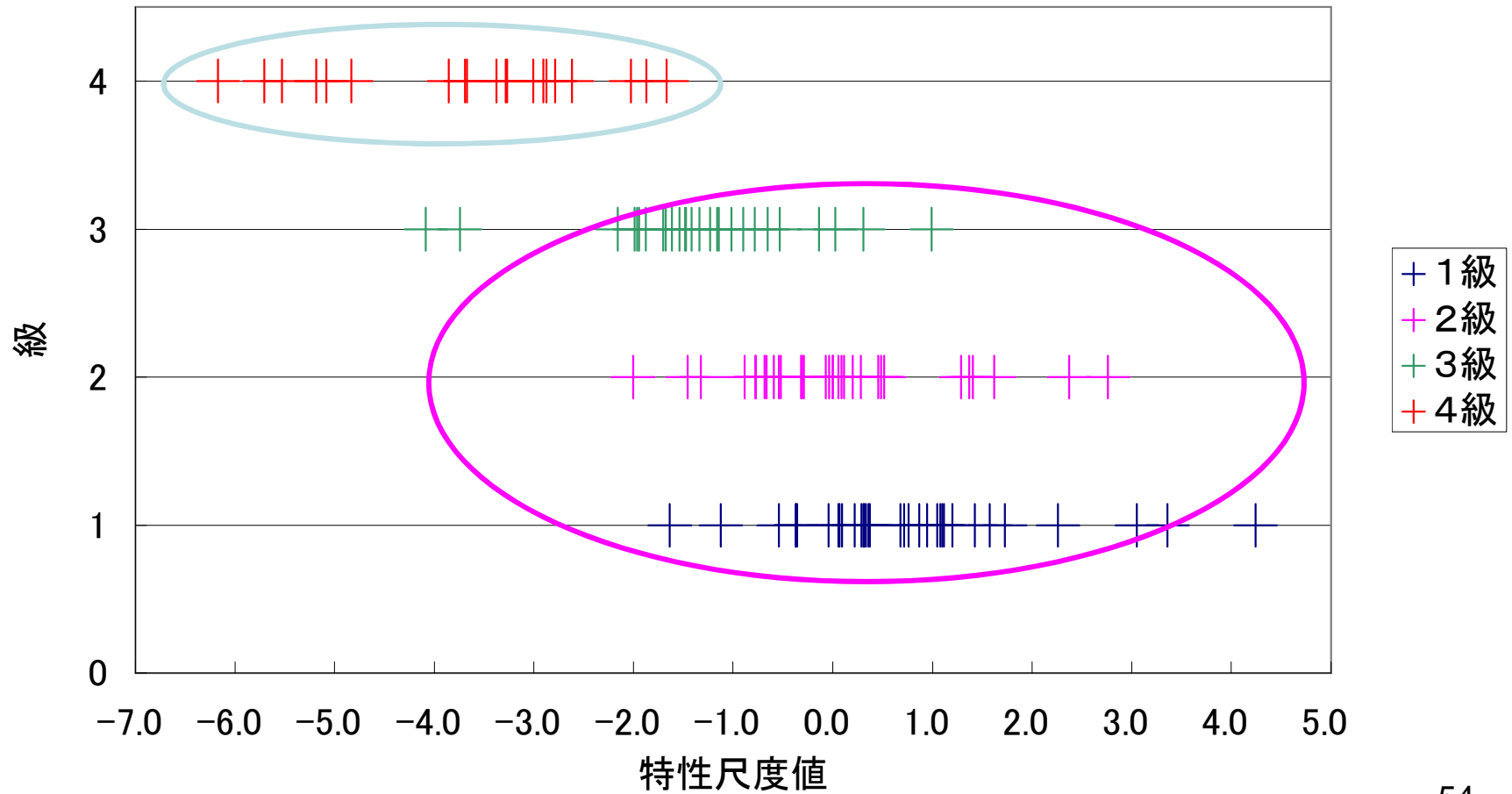


# 文字・語彙

- 級間の項目困難度分布に適度の重なりがあり、困難度が段階的に推移している

# 聽解

聽解項目困難度分布



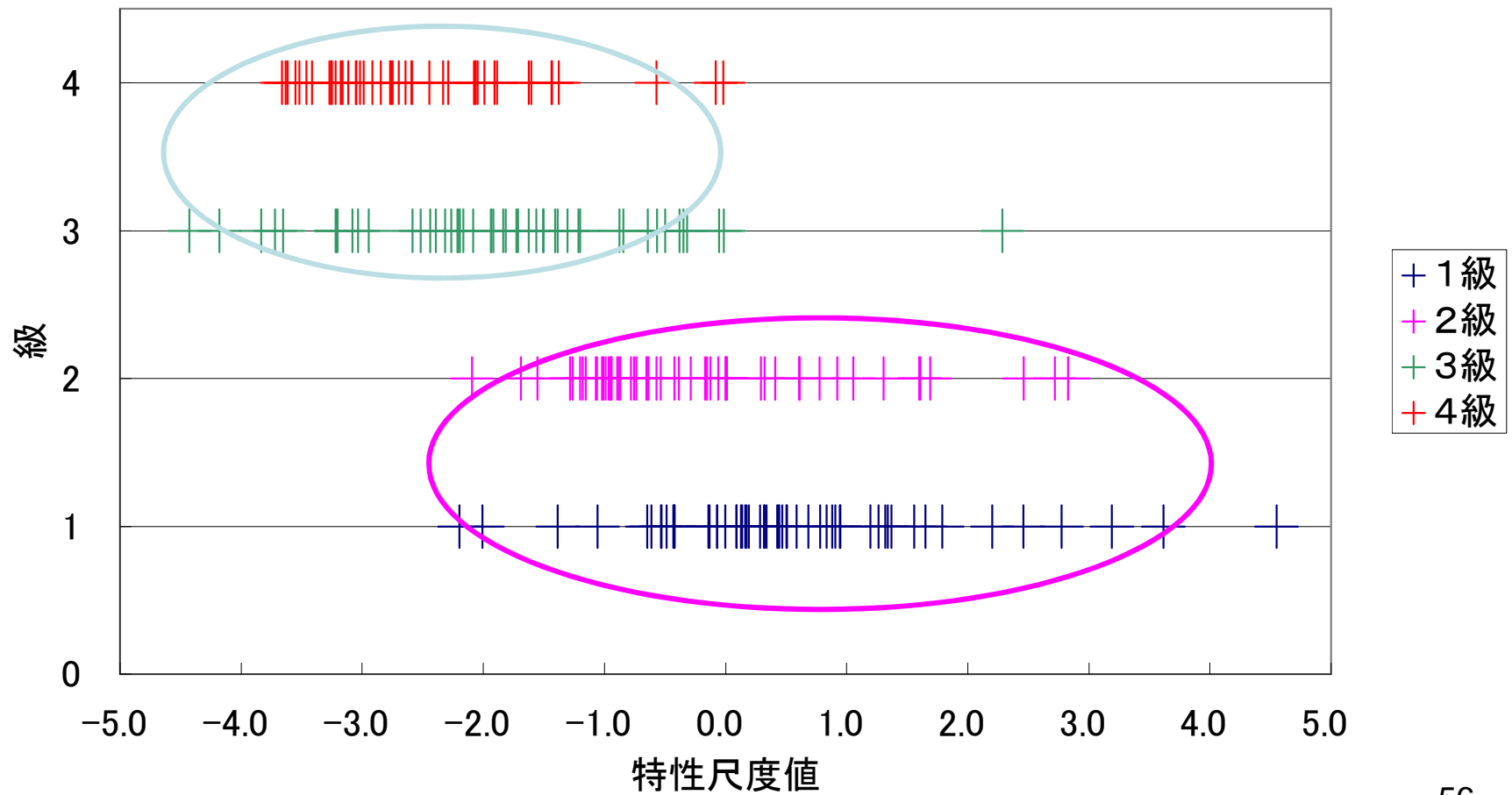
# 聴解

- 4級と3級の間での重なりが少なく、
- 3級はほとんどの項目が2級と重なっている。
- 2級と1級もほとんど重なっている。
- しかしながら、全体としては困難度が段階的に推移している傾向が見られる。



# 読解・文法

読解・文法項目困難度分布

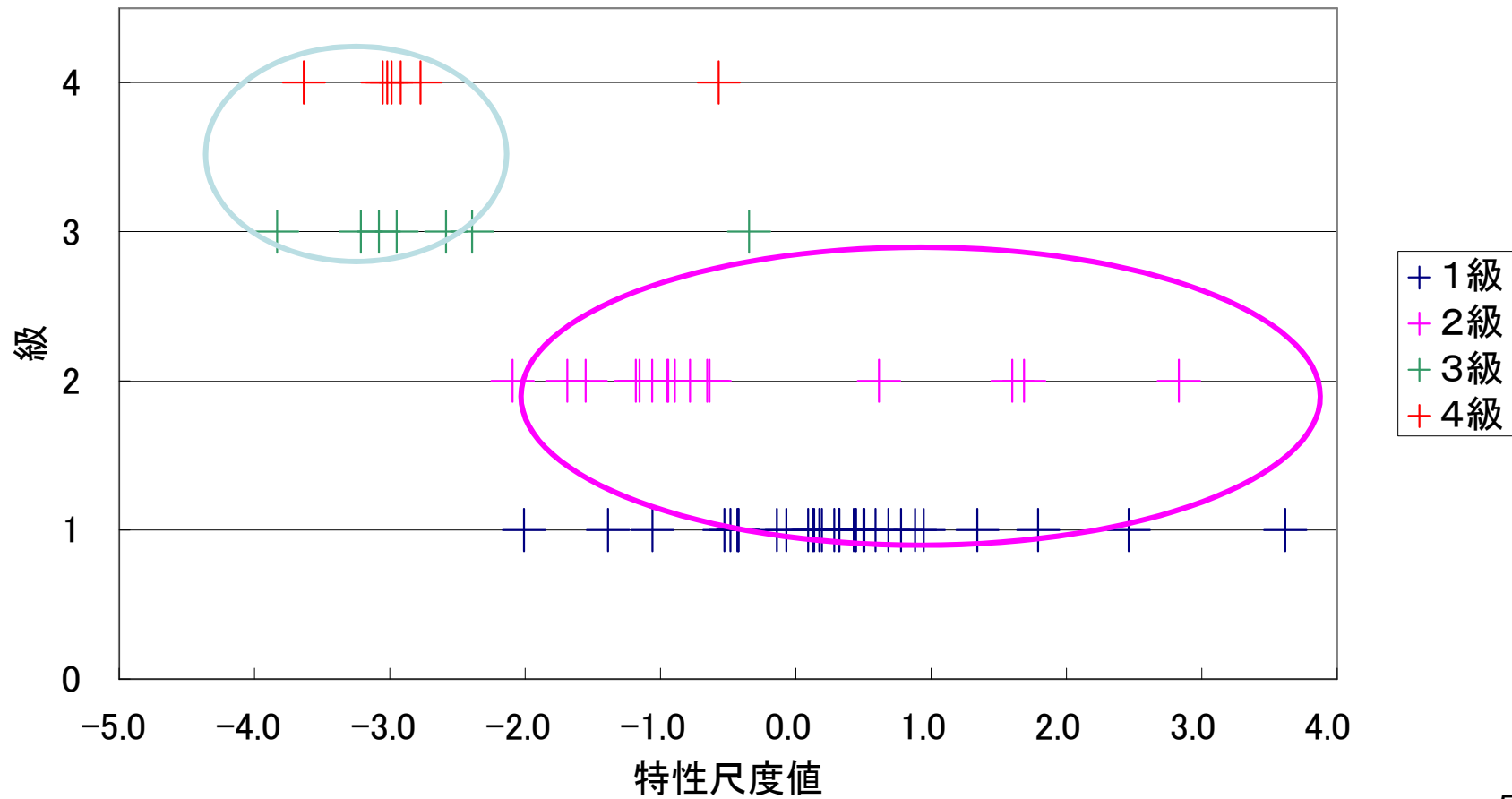


# 読解・文法

- 4級が3級に、2級が1級にそれぞれ含まれ、4, 3級と2, 1級の2群に分かれており、段階的な推移とはいいいにくい。
- また類を分けた場合、「読解」項目のみを取り出した場合(次のスライド)には、3級と2級との間で重なりがほとんど見られないのに対して、「文法」「文字」「語彙」項目のみを取り出した場合は、「文字・語彙」の場合と同様に級間の項目困難度分布に適度の重なりがあり、困難度が段階的に推移していた。

# 読解項目のみ取り出した場合

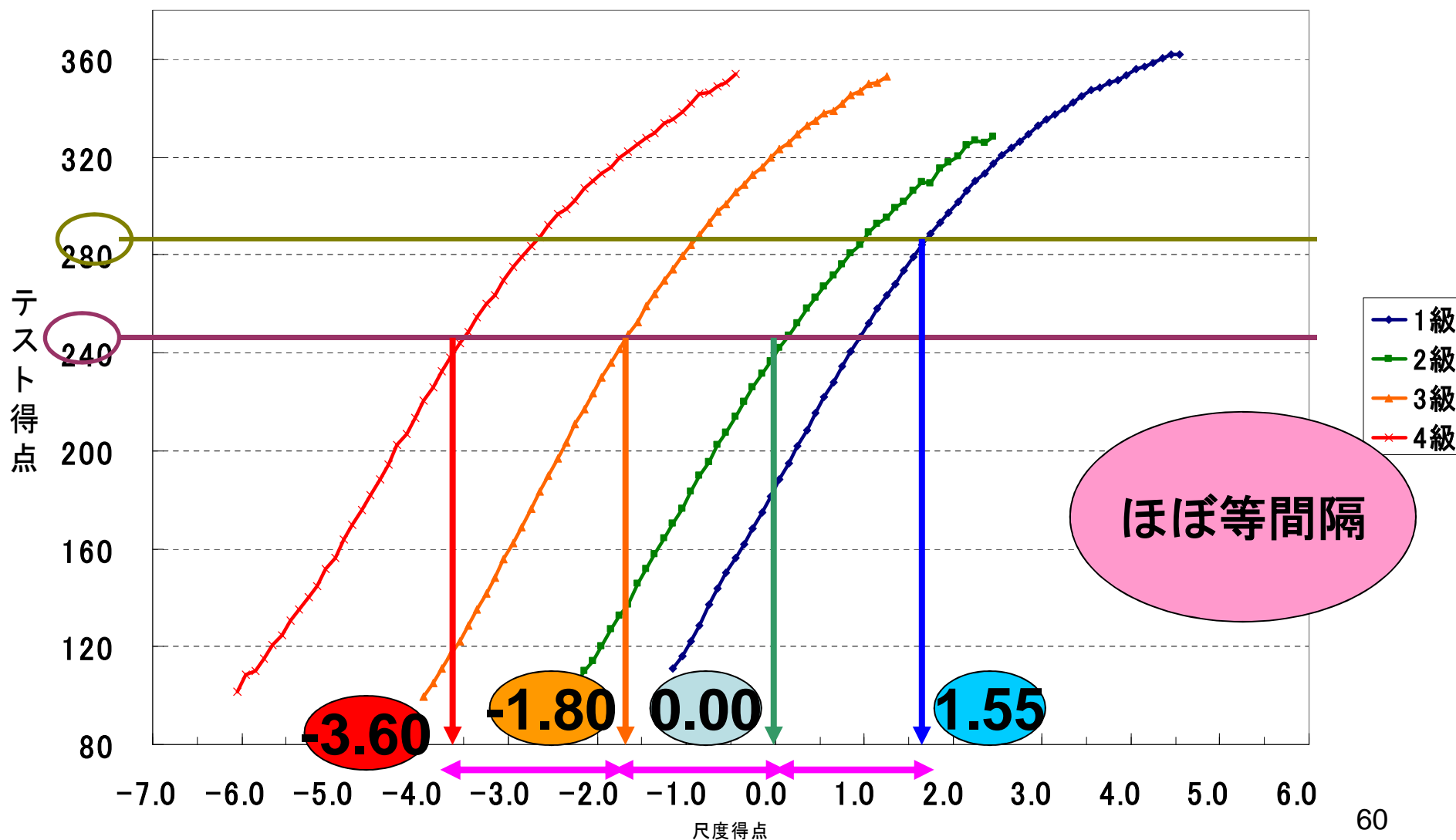
読解項目困難度分布



## 結果 2

2)総合尺度得点と、JLPTの総点との対応関係は次のスライドに示す通りである。  
図中の4本の曲線は、級別に当該総合尺度得点を示す受験者のJLPT総点の平均値を結んだものである。

# 総合尺度得点(X-axis) と JLPT総合得点(Y-axis)の関係



# 合否分割点 ⇔ 総合尺度得点

	合否分割点	総合尺度得点
1級	280	1.55
2級	240	0.00
3級	240	-1.80
4級	240	-3.60



ほぼ等間隔！

- 以上のことから、いわゆる1, 2級と3, 4級との間のギャップはJLPT受験者集団全体としては見られなかった。

## 結果 3

3) 2001年度JLPTの各類特性尺度値の分布状況を母語別(中国語・韓国語・その他)に見ると・・・



# Trait scale values by native language (文字・語彙)

native language	# of examinees			Kanji-Vocabulary			
	Level 2	Level 3		Level 2	Level 3	difference	
<b>Chinese</b>	28,718	25,506	Chinese	<b>0.53</b>	<b>-1.18</b>	<b>1.71</b>	Mean
				1.06	1.08		S.D.
<b>Korean</b>	20,268	18,264	Korean	<b>-0.34</b>	<b>-1.82</b>	<b>1.48</b>	
				1.09	1.17		
<b>Others</b>	9,058	13,699	<b>Others</b>	<b>-0.51</b>	<b>-1.51</b>	<b>1.00</b>	
				0.96	1.16		

- 文字・語彙では、  
「中国語」の差が1.71、  
「韓国語」の差が1.48に比べ、  
「その他」での差は1.00

# Trait scale values by native language (聽解)

native language	# of examinees			Listening		difference	
	Level 2	Level 3		Level 2	Level 3		
<b>Chinese</b>	28,718	25,506	Chinese	<b>-0.15</b>	<b>-1.42</b>	<b>1.27</b>	Mean
				1.06	1.00		S.D.
<b>Korean</b>	20,268	18,264	Korean	<b>-0.01</b>	<b>-1.39</b>	<b>1.38</b>	
				1.16	1.05		
<b>Others</b>	9,058	13,699	Others	<b>0.76</b>	<b>-0.56</b>	<b>1.32</b>	
				1.39	1.29		

聴解に関しては、

「中国語」で差が1.27

「韓国語」で差が1.38

「その他」で差が1.32

で、グループの能力の平均値の2-3級間の差は1.3前後でほぼ等しい。

# Trait scale values by native language (読解・文法)

native language	# of examinees			Reading and Grammar			
	Level 2	Level 3		Level 2	Level 3	difference	
Chinese	28,718	25,506	Chinese	<b>0.06</b>	<b>-1.43</b>	<b>1.49</b>	Mean
				1.03	1.27		S.D.
Korean	20,268	18,264	Korean	<b>0.15</b>	<b>-1.78</b>	<b>1.93</b>	
				1.25	1.29		
Others	9,058	13,699	Others	<b>-0.31</b>	<b>-1.19</b>	<b>0.88</b>	
				1.01	1.23		

- 読解・文法では  
「中国語」で差が1.49  
「韓国語」で差が1.93  
であるのに比べ、  
「その他」では差が0.88と相対的に  
小さい。

- 「その他」グループが代表する「非漢字圏」日本語学習者が「文字・語彙」、「読解・文法」において3級受験者の能力水準から2級受験者の能力水準に達するためには、「中国語」「韓国語」グループに比べてより長い期間を要するということが示唆された。

# まとめると・・・

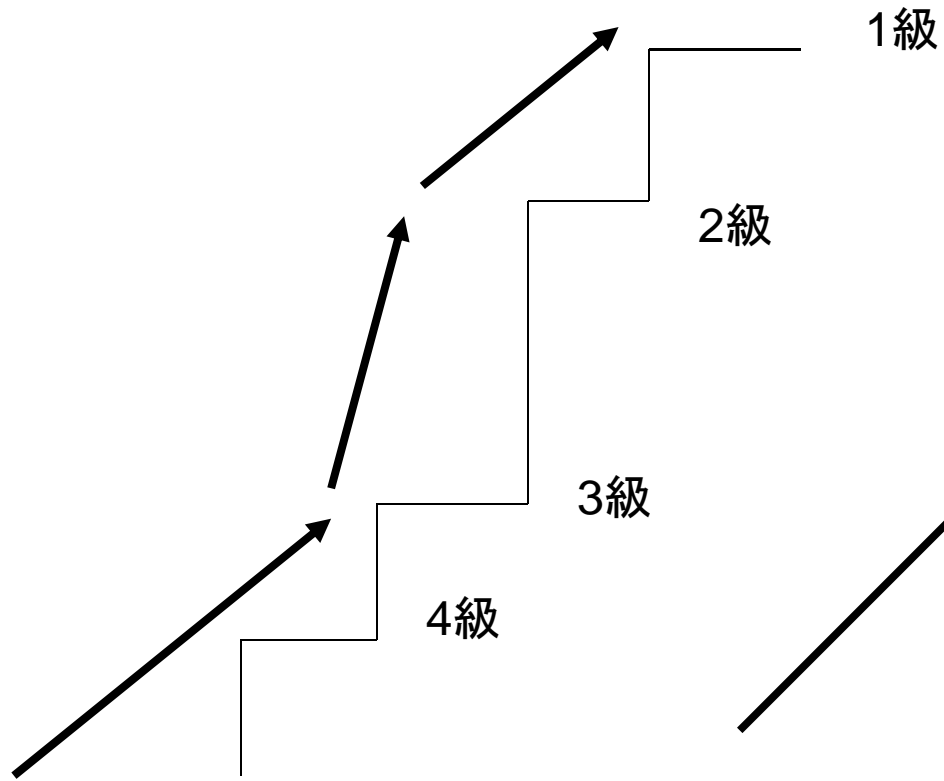
- 1) JLPTでは異なる級の間で項目の困難度が重なりを持っていること、
- 2) 級別の合否分割点の距離は共通尺度上でほぼ等間隔であること、
- 3) いわゆる「2-3級の差」に関しては、共通尺度上における合否分割点の距離が反映したものではなく、非漢字圏学習者の類別尺度得点分布における3級平均値と2級平均値との差が「文字・語彙」「読解・文法」で漢字圏学習者に比べて相対的に小さく、3級水準から2級水準へ学習が進行する速度が漢字圏学習者に比べて緩やかである可能性が反映したものであること、  
が明らかになった。



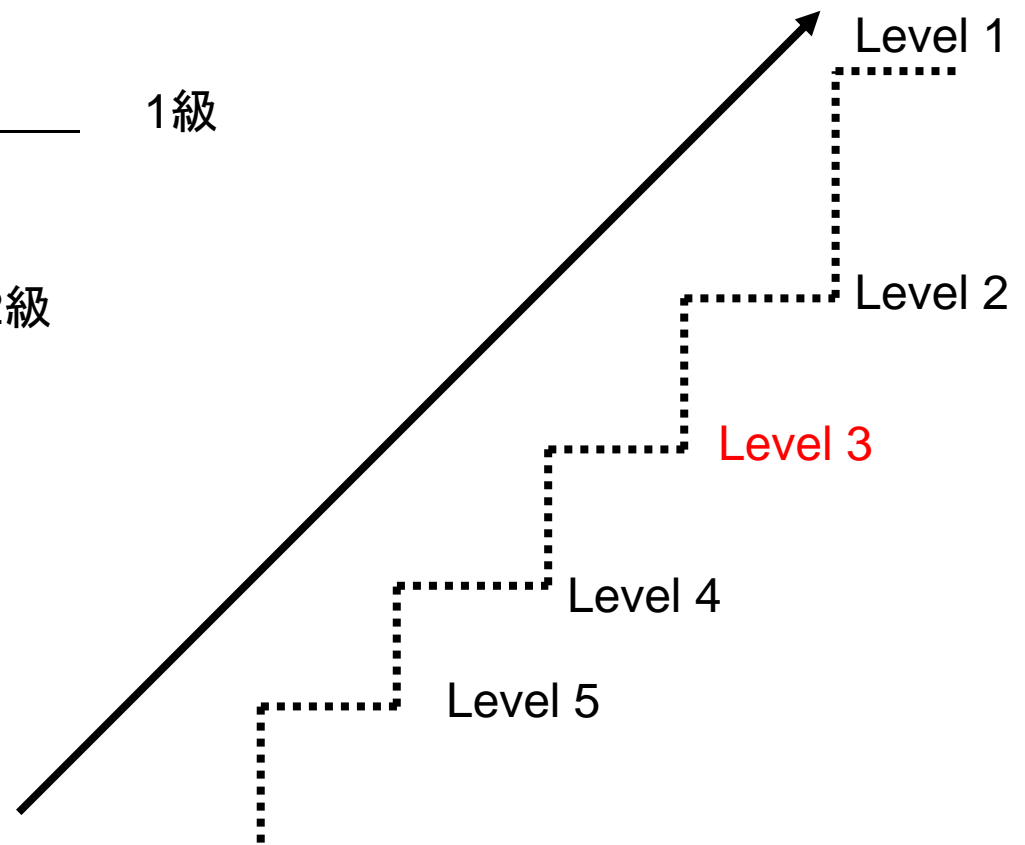
## 改定後のJLPTでは・・・

- 以上の結果を踏まえて、「その他」の言語を母語とする日本語学習者が日本語に対する学習意欲を失わないで、学習を続けられるように、新しいレベルが加えられた。

before 2009



after 2010





Research 3

# Ⅲ 日本語能力試験における 解釈規準開発の試み



この研究は下記の学会で発表したものです。

野口裕之・熊谷龍一・大隅敦子・石毛順子・長沼  
君主 2006 日本語能力試験can-do-statements  
(試行版)のIRT尺度化と日本語能力試験の得点  
段階との対応付けの試み, 5<sup>th</sup> International  
J-OPI-Symposium Berlin 2006.

大隅敦子・野口裕之・熊谷龍一・石毛順子・長沼  
君主・和田晃子・伊東祐郎 (2006), 日本語能  
力試験can-do statementsとCEFR-Dialangとの対  
応付けの試み, 5th International J-OPI-  
Symposium Berlin 2006.

- ここでは、最初に言語テストの解釈規準をめぐる最近の話題をお話しして、
- 次に、改定前の日本語能力試験に対して Can-do statements による解釈規準をつける試みについてお話しします。

# テストの解釈規準

テストの結果を受験者(学習者)や利用者が活用するためには、テスト得点やレベルに対する合否情報のみではなく、テスト結果を解釈するための規準が必要。

# 最近の大規模言語テストでは・・・

- 得点の解釈規準をCan-do statementsによる内容規準で表わしているものが増えている。
- 例えば、日本国内での受験者が多い大規模英語試験の、実用英語技能検定(略称、英検)の場合は、各級の技能毎に「英検合格者の実際の英語使用に対する自信の度合い」を具体的に表わすのに、Can-do statementsによる解釈規準が示されている。



- 日本語の場合、日本語 Can-do-statements が日本語能力試験の妥当性検討のための外的基準の1つとして用いられることを意図して開発され(島田ほか, 2006)、BJT日本語ビジネステストでは受験者の自己評価をCAN-DOレポートとして公表している。また、とよた日本語学習支援システムでは学習者の自己評価のためにCan-do-statementsを用意している(<http://www.toyota-j.com/>)。

# Can-do statementsとは・・・

- 外国語の学習者や外国語試験の受験者がその言語を実際に使用する言語行動場面で、具体的にどんなことができるかを記述した、能力記述文の集合のこと。
- Can-do statementsを利用して外国語能力のレベルを記述したり、学習者による自己評定や教師による評定により、外国語学習者の言語能力を評価したりする。

# ALTE Can-do statements

- ALTE (The Association of Language Testers in Europe) では、The ALTE levels としてALTEに加盟するテスト開発機関の外国語テストを、Can-do statements で記述される6つの水準のいずれかに位置づけて、表わしている (ALTE, 2008)。
  - ⇒ **Social Moderation ( North, 2000 )**
- このことにより、ALTE加盟各国で開発された外国語テストの測定結果の解釈規準が、利用者に分かりやすく示されている
  - ⇒ 欧州各言語テストに共通する解釈規準
- この6つの水準は、**CEFR**で設定されている言語能力の6つの水準 (Council of Europe, 2001) に対応づけて設定されている。

# CEFR (Common European Framework of Reference for Languages: Learning, teaching, assessment)

- 欧州域内で国・言語の違いを超えて、言語教育専門家(テスト開発機関や行政担当官を含む)等が言語学習、教授法、そして評価法に関する相互理解およびコミュニケーションを促進するための基盤となる枠組みを提示した文書のこと。
- 欧州評議会 (Council of Europe) が1997年に開始した言語教育プロジェクトの成果として、2001年に英語版が出版された。

- 「1つの欧州」を目指す欧州評議会の理念を受けて、CEFRには、
  - 1) 欧州市民の相互理解促進のために市民が母語以外の言語も必要に応じて使用できるようになるという「**複言語主義**」と、母語話者並みを必ずしも目標とはせず、必要な能力を身につける「**部分的な能力**」の許容、
  - 2) 学校教育終了後も自律的に学習でき、**生涯学習**を続けられる**学習者支援**、
  - 3) 欧州域内での移動に対する**言語学習の継続性確保**、
  - 4) 教師中心主義ではなく**学習者中心主義**の立場、
  - 5) **行動中心主義**（○○ができる）の言語教育観、

などの特徴がある。

- また、具体的な言語能力水準を「共通参照レベル」として、

A: 基礎段階の言語使用者 (Basic User)、

B: 独立した言語使用者 (Independent User)、

C: 熟達した言語使用者 (Proficient User)

の3レベルに設定し、更に、各レベルを2つずつ、全部で6段階に設定している。

A1: Breakthrough      A2: Waystage

B1: Threshold      B2: Vantage

C1: Effective Operational Proficiency      C2: Mastery

- この6段階のそれぞれについて、  
全体的な尺度、および  
聞くこと、読むこと、話すこと(やりとり)、  
話すこと(表現)、書くこと等の言語行動  
の諸側面を組み合わせ、格子状に配列し、  
それぞれのセルに具体的にできる言語行動  
を能力記述文で例示している。

例えば、

話すこと(やりとり)のB2レベルでは

「自分の興味関心のある分野に関連する限り、幅広い話題について、明瞭で詳細な説明をすることができる」など、

書くことのA2レベルでは

「直接必要のある領域での事柄なら簡単に短いメモやメッセージを書くことができる」など、

言語能力は通常4技能で表わされることが多いが、話すことを「話す(やりとり)」と「話す(表現)」に分けている点が特徴的である。



# JLPT改定前の解釈規準

- 得点解釈規準は用意されていない
- 認定基準は、例えば1級では、  
「高度の文法・漢字(2000字程度)・語彙(10000語程度)を習得し、社会生活をする上で必要な、総合的な日本語能力(日本語を900時間程度学習したレベル)」
- 日本語で具体的にできることがわからない。
- 試験結果の利用者にとって不便な状態にある

# 改定後の得点解釈規準の作成

- 新しい日本語能力試験では、  
「課題遂行能力と  
そのためのコミュニケーション能力」と  
それに必要な「言語構造に関する運用能力」  
を測定する。
- 得点の解釈規準（基準）は、  
「〇〇ができる」という、can-do statementsにより  
解釈規準を表示する。

# 例えば、「話す」能力では

- 学校や職場などで、専門的な話題に関して意見を述べたり、発表したりすることができる。
- 話し相手や状況に応じて、敬語を使って話することができる。
- 手順や操作、道順などについて説明することができる。

など

# 日本語能力試験can-do statements試行版 について

2006年段階では試行版・予備的研究段階  
改定前の日本語能力試験との対応づけの  
方法および、対応関係について検討した。

(野口・熊谷・大隅・石毛・長沼,2006 参照)

# JLPT-CDS 能力記述文作成(1)

2004年 CEFR-DIALANGをもとに作成開始

4技能 × 20項目 = 80項目

ほかにEurope Language Portfolio(ELP), ACTFL-OPI, TOEICなども参照

聞く	
モノローグ	説明
	ニュース
	講義
	スピーチ
ダイアログ	会話
	議論
	テレビ

話す	
モノローグ	描写
	意見
ダイアログ	通常会話
	行動会話

長沼君主氏(東京外国語大学) 作成

読む	
情報検索	日常的文章
	説明
理解	論説文
	物語文
	手紙

書く	
機能系	描写
	意見
媒体系	自己
	伝達
	客観

長沼君主氏(東京外国語大) 作成

# JLPT-CDSとCEFR-DIALANG: 能力記述文の比較

開発チームに属する複数の日本語教育専門家が、双方を比較し、総合的に同一レベルにあると見られる能力記述文を選択した

- 聞く 6項目／20項目中
- 読む 9項目／20項目中
- 書く 10項目／20項目中



## 受験者アンケート (Can-do statements) 調査 時間割

文字・語彙	聴解	読解・文法	アンケート
10 : 30 ↓ 11 : 10	11 : 30 ↓ 12 : 30	13 : 30 ↓ 14 : 40	15 : 00 ↓ 16 : 00
解答時間 30分	30分-45分	60分	60分

2006.3.19 モニター試験とアンケート調査  
受験者は「Ⅱ 日本語能力試験の級間  
共通尺度の構成」と同一

# 調査用紙 質問項目例

N O	質問	経験した ことがあります か？	どの位できますか？ 経 験したことがない人も想 像して教えてください。
2	乗換えや道順、手順や操作な どの説明を聞いて理解できる	<p>Yes <input checked="" type="radio"/></p> <p>No <input type="radio"/></p>	<p>全然できなかった ⇔ 問題なくできた</p> <p>1      2      3 <input checked="" type="radio"/> 4</p> <p>↓</p> <p>全然できないと思う ⇔ 問題なくできると 思う</p> <p>1      2      3      4      5</p>

## 受験者アンケート調査 母語別人数

母語別	人数	比率
中国語話者	393人	36.8%
韓国語話者	132人	12.4%
インドネシア語話者	127人	11.9%
その他	416人	39.0%
合計	1068人	100.0%

# CDS項目の尺度化

- 日本語能力試験can-do statementsを多値型IRTモデルのひとつである一般化部分得点モデル(Generalized Partial Credit Model)を用いてCDS各statementの困難度を推定した。  
モデルの location parameter 推定値を当該項目の困難度とした。
- その結果、Can-do statements の尺度上での順序をつけることができた。

# CEFR-DIALANGの能力記述文 との順序の比較

- 日本語能力試験Can-do statements の尺度上での順序と、CEFR-DIALANGの能力記述文の順序を両者に共通な(近似した)能力記述文については比較することができる。
- すなわち、聞く6項目、読む9項目、書く10項目について比較した。

# 聞く(6項目)の順序の比較

易

• 日常生活: 店、レストラン、郵便局、銀行、  
駅などでの日常的な場面

A2

• 電器店やデパート: 説明が長くなる

B1

• 様々な文体、話題、細かい内容、専門的討論や議論

C1

難

# 読む(9項目)の順序の比較

易

- 知人友人からきたメールやハガキ A2

- 公の標示・ポスター A2

- 学校・職場・地域の  
• 掲示 A2

- 短いメモ・  
• 年賀状 A1

- 商品・旅行のパンフB1

- 新聞には含まれている広告・チラシ A2

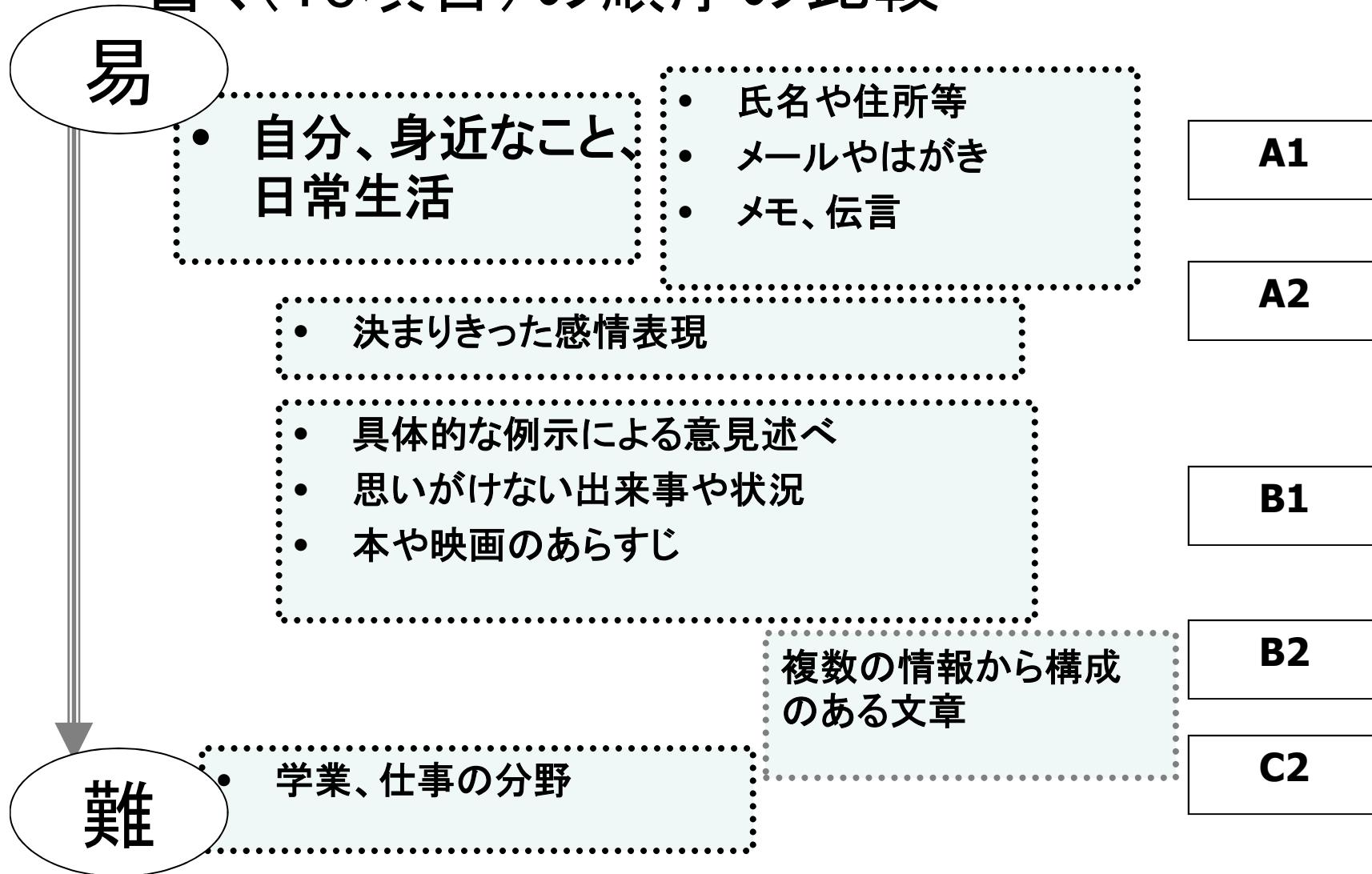
- 公衆電話や券売機 A2

- ヘアドライヤーや掃除機 B1

- 携帯電話やデジタルカメラC1

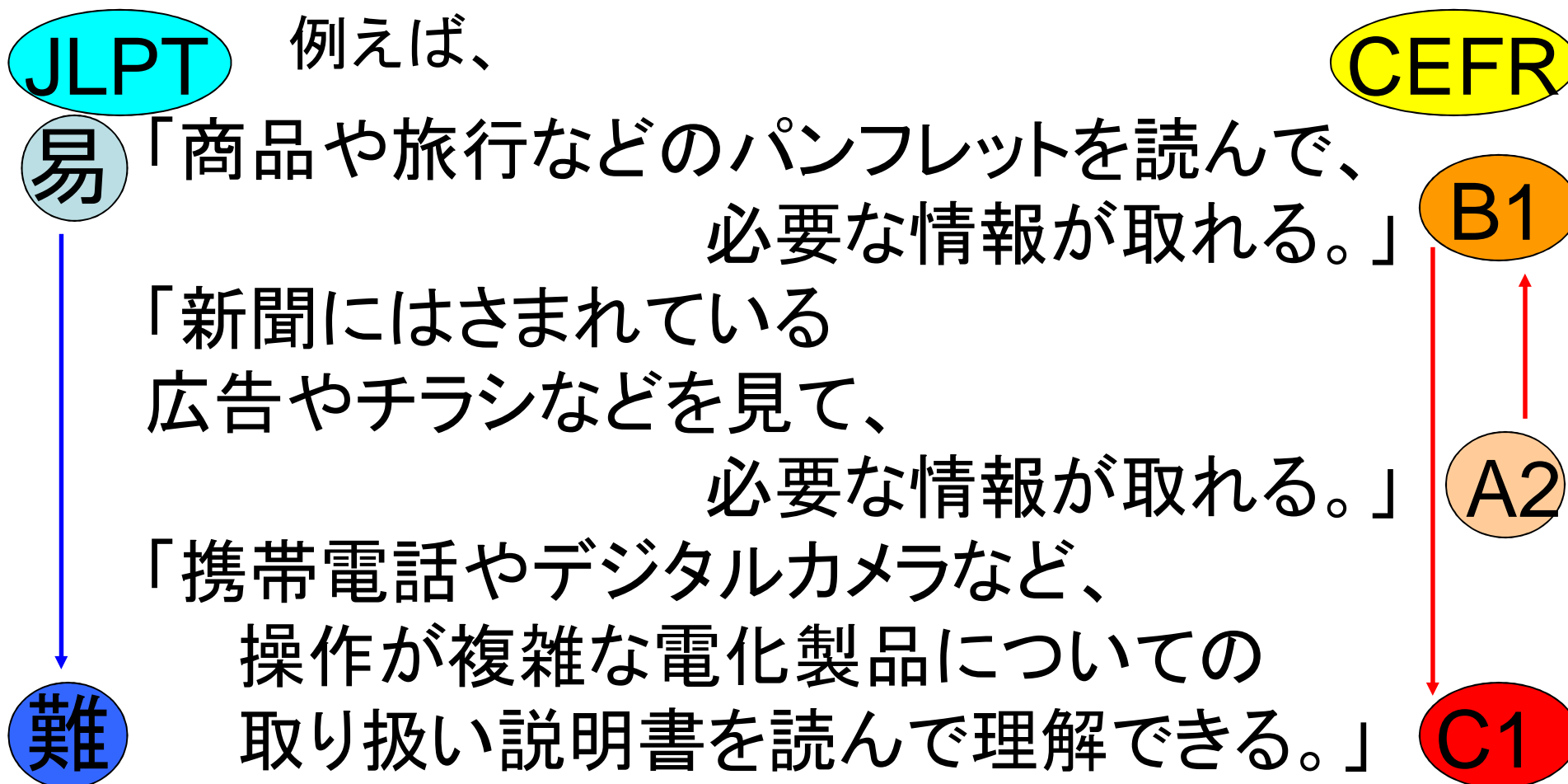
難

# 書く(10項目)の順序の比較





# 『読む』ではCEFR-DIALANGの順序性との 一貫した対応が見られない



大隅・野口・熊谷・石毛・長沼・和田・伊東,2006 参照

# JLPTとCDS能力記述文の対応付け

- 本研究における受験者1068名に関しては、  
JLPT級間共通尺度上での能力尺度値  
JLPT - CDS尺度上での特性尺度値  
とが得られている。
- 両者の同時分布から、**JLPT級間共通尺度  
推定値のJLPT - CDS特性尺度推定値に対  
する回帰直線**を求め、この直線をもってCDS  
と日本語能力試験共通尺度との**関係を表わ  
す**ことができる。

- 回帰直線を利用して、CDS各statementの困難度をJLPT級間共通尺度上に対応づけた。
- 両者の相関係数は .625と比較的高く、  
回帰直線の係数は0.943、切片は-0.08 。
- CDS各statementの日本語能力試験IRT共通尺度上に対応づけられた値を各級の合否分割点(4級が-3.6、3級が-1.8、2級が0.0、1級が1.6)と比較した。
- その結果、CDS80項目中で4級合格水準が7項目、3級が47項目、2級が26項目で1級合格水準の項目はなかった。

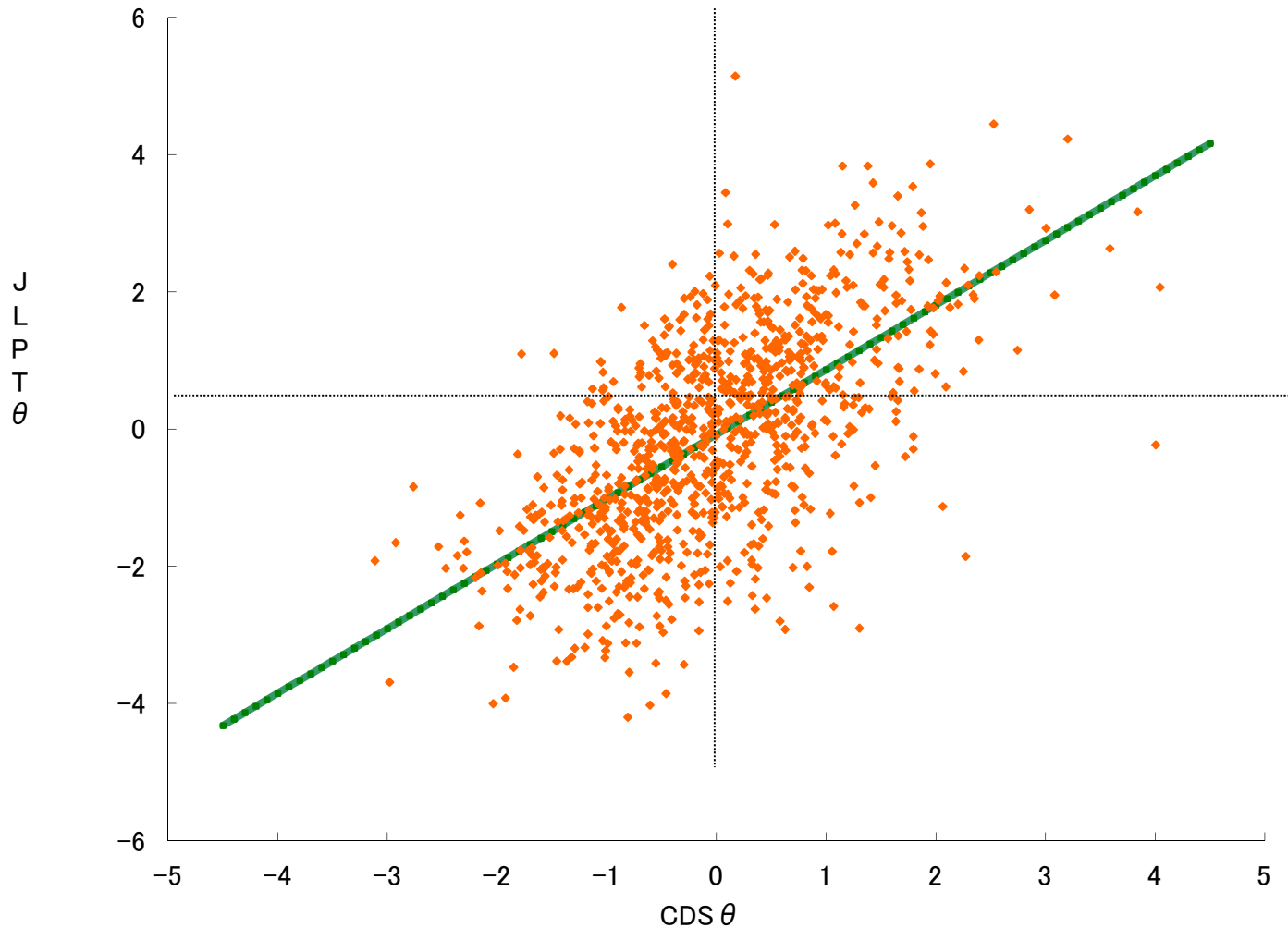


図1  
各モニター調査協力者の日本語能力試験can-do-statements尺度上の  
推定値と日本語能力試験IRT共通尺度上の推定尺度値

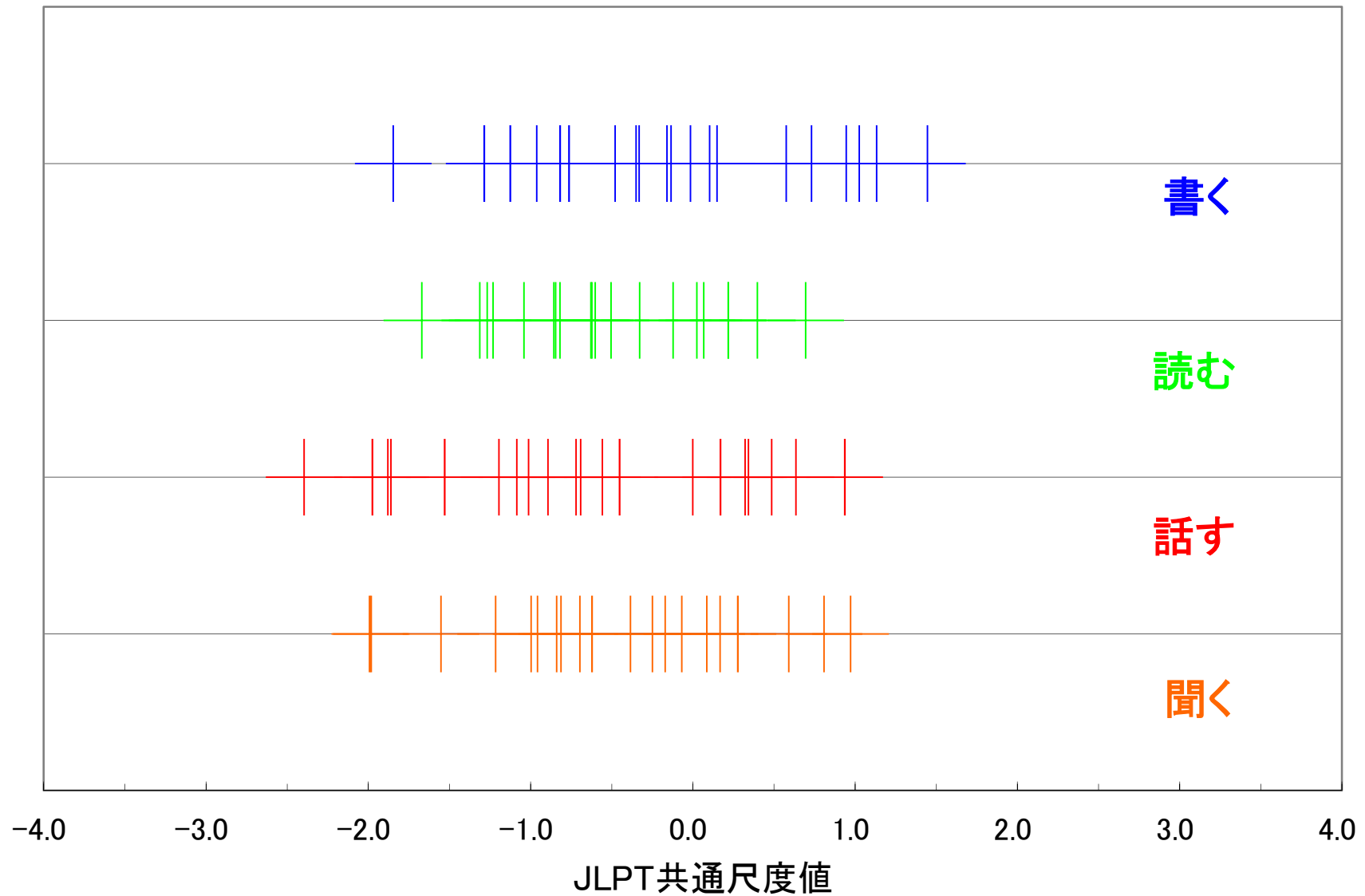


図2 日本語能力試験 can-do-statementsの各項目を  
日本語能力試験IRT共通尺度に対応付けた結果

# まとめ

1. この分析において、JLPT-CDSおよびCEFR-DIALANGは、特に「聞く」「書く」において、順序性の一致が見られた。
2. CEFR-DIALANGを外在基準として眺めた時、順序性という観点からは、「聞く」「書く」を中心に、JLPT-CDSには一定の妥当性が認められたと考えられる。
3. 日本語の場合、欧州系言語との共通性の度合いが低い。そのため、CEFRに対応づけられる能力記述とそうでない能力記述とがある。特に、「読む」「書く」の能力に関して「文字(漢字)」の能力をどう位置づけるか が問題。

4. 今後の課題として、適用したIRTモデルが適切であったか、CDS特性尺度と日本語能力試験共通尺度との対応づけの方法は適切であったか、日本語能力試験の得点段階を大きくとって、各級の合否分割点で段階を区切ったが、実際の日本語能力試験の受験者や利用者にとって、この区切りが利用しやすいものであるかどうか、などを検討する必要がある。

# 本研究の今後の課題

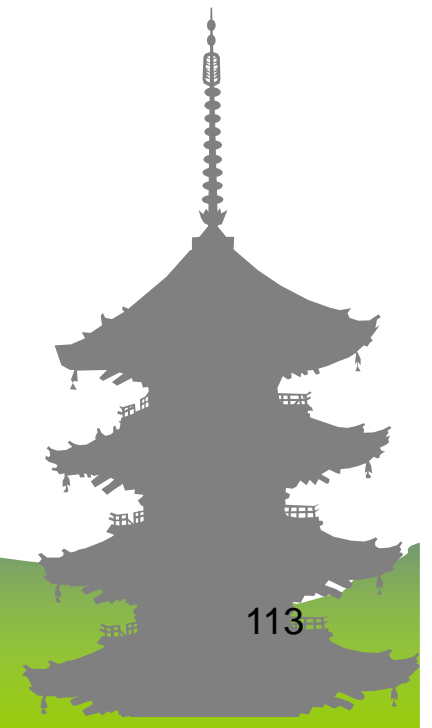
本研究は改定前の日本語能力試験の枠組みで検討を進めたものであるが、改定後の日本語能力試験の仕様 (specification) に沿った方向でCDSの改訂が進められている。

また、CDSを介して他の言語テストとの能力基準の共通化および独自性の明確化を図る必要がある。





おわりに



# おわりに

- テスト、特に言語テストは世界の潮流をきちんと見る必要がある。しかし、時流に流されてはいけない。
- テスト開発は決してテスト理論の発展だけではうまく行かない。測定対象の理論的な変化もきちんと把握しておく必要がある。しかし、現実の世界を見る眼をもつことがもっと大切なこと。

- 新しい理論・技術・道具を活用することは大切なことであるが、濫用に陥ってはいけない。
- テスト開発は極めてヒトに集約されたプロジェクトである。いいテスト開発にはいい人材の養成が不可欠。テストのことを総合的に考えられる人材を養成するシステムが必要。
- 言語テストは国家レベルの国際戦略にも位置づけられる。国際貢献のための言語政策の中に言語テストを位置づけることが必要。



どうもありがとうございます！

