

日本語能力試験を再構成したテストの実施と分析

— 古典的テスト理論と項目応答理論を用いて —

小池 康

要 旨

本稿は、日本語能力試験（JLPT）の1級から3級までの文字・語彙・文法問題を再構成したテスト（「日能テスト」）の実施結果とその考察である。テストは、2009年度の筑波大学留学生センターのJ600レベル（JLPT2級相当）の留学生を対象に実施した。

古典的テスト理論による分析の結果、3級の設問項目、中でも文法問題において識別力が低い項目が多く見られた。これらの設問項目は、受験者の能力を正しく峻別できていないと言えるので、改良の余地が残されている。また、同じく3級文法の設問項目では、正答率が5割にも満たない項目も多く見られた。項目の内容は受講生のブレースされたクラスのレベルからすると、基礎的な文法事項を問うた、簡単な項目と言えるはずのものであり、それらが正答率5割にも満たないというのは、今後当該項目の教授方法を再考しなければならないものと考えられる。

項目応答理論による分析では、漢字の書き問題でミスフィットの項目が多く今後修正していく必要がある。

【キーワード】 日本語能力試験 日能テスト 古典的テスト理論 項目応答理論

Implementation and Analysis of a Test which Rearranges Items of the JLPT: Applying Classical Test Theory and Item Response Theory

KOIKE Yasushi

[Abstract] We conducted the following test to judge the Japanese language proficiency of foreign students who attended an intermediate/advanced grammar class. The test was made of items of writing, vocabulary and grammar sections from levels 1 to 3 of the Japanese-Language Proficiency Test (JLPT, hereafter).

The results of the analysis, which uses Classical Test Theory, show that many items of level 3 were low in Item Discrimination Power. Those items cannot classify examinee's ability appropriately. Besides, there are many items of which proportion correct is less than 50%, although these items should be easy to answer for students that attended this class. Those results imply that students were lacking a basic grammatical knowledge.

From an analysis based on Item Response Theory, it was found that there were a lot of misfit items in questions of Kanji writing. We need to modify these items.

[Keywords] JLPT, the test rearranging JLPT, Classical Test Theory, Item Response Theory

1. はじめに

本稿は、2009年度に筑波大学留学生センターにて、初中級レベルの文法クラスにおいて実施した、日本語能力試験 (The Japanese Language Proficiency Test、以下JLPTと略す) を再構成したテストに関する報告である。

この再構成したテスト (以下、「日能テスト」と呼ぶ) の概略は、すでに小池 (2010) で述べた。この小池 (2010) は、2009年度春学期¹に同留学生センターにて実施した日能テストを古典的テスト理論を用いて分析したものであった。

現在、テストを分析する際に広く用いられている手法として古典的テスト理論と項目応答理論があるが、小池 (2010) では項目応答理論での分析はできなかった。その理由は、被験者の人数が62名と足りなかったためである。項目応答理論で分析する際の被験者数の目安は100~200名程度とされる²が、小池 (2010) ではその数に満たなかったため項目応答理論での分析は不相当だと考えたのである。しかし、2009年度の一年間を通して同一レベルの文法クラスを担当したことにより、結果的に119名の被験者数を得ることができた。本稿では、この被験者を対象に、古典的テスト理論および項目応答理論を用いて日能テストを分析する。

2. 旧JLPTの概要

JLPTは、主に日本語を母語としない人を対象に、公的に日本語能力を測定し、認定することを目的として、国際交流基金と日本国際教育支援協会が実施しているものである。1984年から毎年1回実施されてきたが、2010年度からはそれまでの形態を一新したものとなった。本稿で扱うJLPTは旧形態のものであるため、読者の混乱を避けるために、以後2010年度以前のJLPT (すなわち今回のテストで参考にしたJLPT) を「旧JLPT」と呼ぶことにする。

旧JLPTは、学習段階に対応する1級から4級の級別を実施され、各級は文字・語彙 (100点)、聴解 (100点)、読解・文法 (200点) の3類計400点満点で結果が表示された。設問形式は、一つの設問につき四つの選択肢があり、正答一つを選びマークシートに記入するというものであった。

2002年度以降、毎年『分析結果に関する報告書』 (以下『報告書』と略す) が国際交流基金より刊行され、試験の全設問項目の結果とその分析・評価が記載されている。

3. 日能テストの概要

現在の大学の中には、日本語能力試験2級相当の日本語能力を条件として留学生の入学を認めているところがある。筆者もこれまでにそのような大学で教えた経験があるが、実際に講義を行なってみると2級レベルには至っていないと思われる学生も多く在籍する大

学があった。

そこで、当の学生のレベルが実際にどれくらいのレベルなのかを判定する目的で、『報告書』に記載されている旧JLPTの設問項目およびその分析結果を基に、新たにテストを再構成した。それが「日能テスト」である。

日能テストに関する詳細は小池(2010)を参照していただくこととして、ここでは必要最小限の部分に限り説明をしておく。

3.1 テストの構成

旧JLPTは文字・語彙、聴解、読解・文法の三つの類のテストから構成されていたが、日能テストではこの内、文字、語彙、文法より設問を選定し作成した。

この三つに限った理由としては、時間的な制約などが挙げられる。

3.2 設問項目の選定

日能テストは、旧JLPTの3級以上の設問項目より選出した。その理由は、2級相当のクラスにプレースされていても、すべての受講生が2級までの日本語の知識を持っているとは限らないと考えられたため、一つ下位級の3級の設問項目も含めたのである。

日能テストの設問項目は、『報告書』に記載されている、当該年度に出題された設問項目

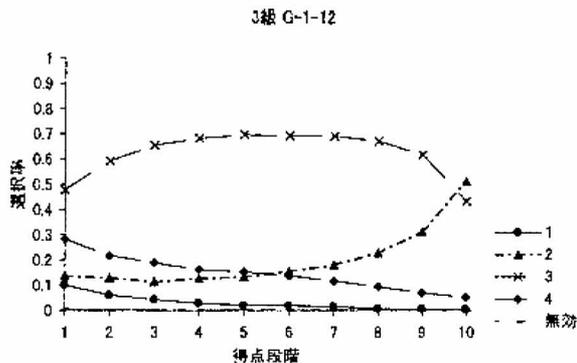


図1 得点段階と選択率のグラフ

目とその正答率と識別力、および当該設問項目の回答パターンとも言える、得点段階と選択率のグラフなどの分析結果(図1は2004年度文法3級問題I-12の結果で、正答は2)を基に選定した。

特に、得点段階と選択率の現れ方を大きくAからEの5つに分類し、さらに複数の選択肢間で交差があった場合に応じて、下記のように下位区分した。

A Aタイプー得点段階10(高得点者群)で正答が1位にきた設問項目

: 相対的に得点の高かった、得点段階8以上の受験者でも間違えた設問項目
(図1参照)

Aタイプー正答の選択率が得点段階の8以上で1位にきた設問項目

: 大半の受験者が誤答の設問項目=できる受験者が正解した設問項目

2. 二つの選択肢間で交差(近接)があった設問項目

3. 三つの選択肢間で交差（近接）があった設問項目
 4. 四つの選択肢間で交差（近接）があった設問項目
- Bタイプ—正答の選択率が得点段階4～7で1位にきた設問項目
- ：AタイプとCタイプの中間
2. 二つの選択肢間で交差（近接）があった設問項目
 3. 三つの選択肢間で交差（近接）があった設問項目
 4. 四つの選択肢間で交差（近接）があった設問項目
- Cタイプ—正答の選択率が得点段階の3以下で1位にきた設問項目
- ：得点の低かった受験者が間違えた設問項目
2. 二つの選択肢間で交差（近接）があった設問項目
 3. 三つの選択肢間で交差（近接）があった設問項目
 4. 四つの選択肢間で交差（近接）があった設問項目
- Dタイプ—正答の選択率が得点段階1（低得点者群）で1位にきた設問項目
- ：大半の受験者が正解した設問項目
- Eタイプ—どの得点段階においても、正答と誤答の差が少ない設問項目

Aタイプの設問項目は、総合得点が高かった受験者でも誤答した者の多かった、難易度の高い項目であり、逆にDタイプは、総合得点が低かった受験者でも正答した、難易度の低い項目ということである。また、Eタイプは、受験者の得点の高低にかかわらず正答者の傾向が見られない項目で、結果的に識別力が低いものばかりであった。

このように得点段階別のグラフを比較することで、同じ級の設問項目でも難易度に差があることがわかり、さらには当該の設問項目が当該級の能力を適切に反映しているものなのかどうかもわかるのである。

以上のような、設問項目とその分析結果（正答率・識別力、得点段階と選択率のグラフ）を参考に、文字、語彙、文法の過去問題より、日能テスト用の設問項目を選定した。なお、このテストを実施する段階までに入手できた『報告書』が、2004年度および2005年度の二年分のみだった³ので、これより設問項目を選定した。

設問項目は、1～3級の設問を上記A～Eに分類し、そこから識別力や正答率などで適切な難易度の項目がバランスよく含まれるよう配慮した。ただし、このテストは2級相当以上の日本語能力を持つ学生の能力を測ることを念頭に置いているため、各級の項目数の割合は1・2級が同数程度、3級はそれより少ない項目数とした。また、特に3級の項目は、3級の中でも相対的に難易度の高かった項目（AやBレベルの項目）を中心に構成した。その結果、各級の設問数は1級が25問、2級が25問、3級が16問の合計66問となった。

問題構成は、大問を全部で5つ設け、問1は漢字の読み問題、問2は漢字の書き問題、

問3は語彙問題、問4は文法問題、問5は語の使用法の問題となっている。各問とも、設問項目は3級レベルから1級レベルになるようにし、また同一級内では基本的に上述のD（大半の受験者が正解した設問項目）からA（大半の受験者が誤答の設問項目＝能力の高い受験者が正解した設問項目）になるよう配列した。

以上を踏まえ、日能テストを作成した。具体的な設問項目は本稿末資料を参照のこと。

4. テスト実施の概要

日能テストは、筑波大学留学生センターにて筆者の担当する文法クラスで実施した。このクラスは、クラス名を「補講J600文法」といい、レベルとしては中上級に該当した⁴。この「補講J600」レベルはおおよそ旧JLPTの2級に相当し、日能テストの出題レベルと同程度と想定されたため、このクラスでテストを実施したのである。

テストは、2009年度の各学期の初回講義日に実施した。全被験者数は142名であったが、同じ学生が複数回本クラスを受講した場合もあった⁵ので、そのような学生のデータは初回受験時のデータのみを使用した。その結果、以下のように異なり受験者数は119名であった。

| | | | |
|-----|----------|---------|--|
| 春学期 | 62名 | | |
| 秋学期 | 30名（異なり） | 40名（延べ） | |
| 冬学期 | 27名（異なり） | 40名（延べ） | |

なお、3つの学期すべてを受講した者が2名いたほか、春学期と秋学期の重複者は8名、以下春と冬は2名、秋と冬は9名であった。

テストは、全学期同一のものであり、設問項目の変更や修正は加えていない。また、テストの制限時間は、旧JLPTの文字・語彙の実施時間と日能テストの設問項目数より40分と定めた⁶。

5. 分析

5.1 古典的テスト理論による分析

5.1.1 設問項目から見た特徴

表1（次頁）は、各級別に受験者の正答数・正答率・識別力（点双列相関係数）を表わしたものである。「10正答数／10正答率／10識別力」は小池（2010）での結果である。これは2009年度春学期のみを分析したデータなので、表1では春学期のみの結果と通年を通しての結果の比較ができるようになっている。また、識別力のうち太枠で囲まれた数値は、同一級内における識別力が0.3以下であることを示している。

表1より、識別力の低かった項目数を小池（2010）と比較すると、以下のようになる（“／”の右項は小池2010の結果）。

表1 各級別設問項目の正答数・正答率・識別力

| 3級 | Q 1 | | | | | | Q 2 | | | Q 3 | | | |
|-------|-------|-------|--|--|--|--|-------|--|--|-------|-------|-------|-------|
| | 1 | 2 | | | | | 16 | | | 23 | 24 | 25 | 26 |
| | B2 | B4 | | | | | C2 | | | B4 | A4 | A4 | A4 |
| 正答数 | 100 | 82 | | | | | 106 | | | 109 | 89 | 82 | 65 |
| 正答率 | 84.0% | 68.9% | | | | | 89.1% | | | 91.6% | 74.8% | 68.9% | 54.6% |
| 識別力 | 0.210 | 0.433 | | | | | 0.159 | | | 0.235 | 0.260 | 0.339 | 0.328 |
| 10正答数 | 52 | 41 | | | | | 56 | | | 59 | 45 | 34 | 32 |
| 10正答率 | 83.9% | 66.1% | | | | | 90.3% | | | 95.2% | 72.6% | 54.8% | 51.6% |
| 10識別力 | 0.434 | 0.604 | | | | | 0.252 | | | 0.290 | 0.283 | 0.419 | 0.347 |

| 2級 | Q 1 | | | | | | Q 2 | | | Q 3 | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 3 | 4 | 5 | 6 | 7 | 8 | 17 | 18 | 19 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |
| | C3 | C3 | C4 | C4 | B2 | B2 | C A | C2 | C4 | C4 | B4 | B3 | B3 | A2 | A3 | A4 |
| 正答数 | 87 | 86 | 84 | 88 | 58 | 63 | 115 | 102 | 93 | 76 | 88 | 58 | 81 | 70 | 34 | 68 |
| 正答率 | 73.1% | 72.3% | 70.6% | 73.9% | 48.7% | 52.9% | 96.6% | 85.7% | 78.2% | 63.9% | 73.9% | 48.7% | 68.1% | 58.8% | 28.6% | 57.1% |
| 識別力 | 0.560 | 0.320 | 0.648 | 0.323 | 0.524 | 0.410 | 0.171 | 0.434 | 0.461 | 0.460 | 0.171 | 0.402 | 0.398 | 0.585 | 0.213 | 0.475 |
| 10正答数 | 38 | 39 | 36 | 47 | 26 | 28 | 60 | 51 | 48 | 38 | 45 | 26 | 37 | 26 | 18 | 30 |
| 10正答率 | 61.3% | 62.9% | 58.1% | 75.8% | 41.9% | 45.2% | 96.8% | 82.3% | 77.4% | 61.3% | 72.6% | 41.9% | 59.7% | 41.9% | 29.0% | 48.4% |
| 10識別力 | 0.578 | 0.243 | 0.686 | 0.316 | 0.397 | 0.353 | 0.112 | 0.435 | 0.461 | 0.502 | 0.340 | 0.478 | 0.437 | 0.608 | 0.169 | 0.433 |

| 1級 | Q 1 | | | | | | | Q 2 | | | Q 3 | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 20 | 21 | 22 | 34 | 35 | 36 | 37 | 38 | 39 |
| | C2 | B2 | B3 | B4 | B4 | A3 | A2 | C A | C2 | A A | C3 | C2 | B2 | B2 | B4 | A2 |
| 正答数 | 64 | 45 | 43 | 54 | 32 | 23 | 35 | 93 | 55 | 45 | 99 | 43 | 35 | 41 | 45 | 94 |
| 正答率 | 53.8% | 37.8% | 36.1% | 45.4% | 26.9% | 19.3% | 29.4% | 78.2% | 46.2% | 37.8% | 83.2% | 36.1% | 29.4% | 34.5% | 37.8% | 79.0% |
| 識別力 | 0.374 | 0.320 | 0.305 | 0.223 | 0.331 | 0.346 | 0.369 | 0.409 | 0.407 | 0.235 | 0.388 | 0.655 | 0.395 | 0.443 | 0.479 | 0.348 |
| 10正答数 | 25 | 21 | 19 | 26 | 15 | 7 | 19 | 48 | 24 | 18 | 51 | 14 | 13 | 17 | 20 | 47 |
| 10正答率 | 40.3% | 33.9% | 30.6% | 41.9% | 24.2% | 11.3% | 30.6% | 77.4% | 38.7% | 29.0% | 82.3% | 22.6% | 21.0% | 27.4% | 32.3% | 75.8% |
| 10識別力 | 0.332 | 0.418 | 0.274 | 0.293 | 0.399 | 0.135 | 0.381 | 0.396 | 0.302 | 0.319 | 0.443 | 0.453 | 0.208 | 0.311 | 0.367 | 0.361 |

| 3級 | Q 4 | | | | | | Q 5 | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 60 | 61 |
| | A A | B4 | B4 | B4 | A4 | A4 | A2 | B3 | B4 |
| 正答数 | 74 | 81 | 104 | 77 | 53 | 52 | 56 | 98 | 103 |
| 正答率 | 62.2% | 68.1% | 87.4% | 64.7% | 44.5% | 43.7% | 47.1% | 82.4% | 86.6% |
| 識別力 | 0.300 | 0.345 | 0.272 | 0.327 | 0.227 | 0.084 | 0.090 | 0.369 | 0.158 |
| 10正答数 | 35 | 37 | 51 | 35 | 29 | 29 | 24 | 49 | 53 |
| 10正答率 | 56.5% | 59.7% | 82.3% | 56.5% | 46.8% | 46.8% | 38.7% | 79.0% | 85.5% |
| 10識別力 | 0.463 | 0.413 | 0.314 | 0.300 | 0.335 | 0.322 | 0.372 | 0.483 | 0.264 |

| 2級 | Q 4 | | | | | | Q 5 | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 62 | 63 |
| | C3 | C4 | B2 | B2 | B2 | A2 | A4 | B3 | A2 |
| 正答数 | 79 | 97 | 76 | 76 | 45 | 36 | 22 | 57 | 86 |
| 正答率 | 66.4% | 81.5% | 63.9% | 63.9% | 37.8% | 30.3% | 18.5% | 47.9% | 72.3% |
| 識別力 | 0.435 | 0.428 | 0.327 | 0.477 | 0.318 | 0.350 | 0.254 | 0.298 | 0.255 |
| 10正答数 | 37 | 47 | 38 | 29 | 24 | 13 | 13 | 27 | 44 |
| 10正答率 | 59.7% | 75.8% | 61.3% | 46.8% | 38.7% | 21.0% | 21.0% | 43.5% | 71.0% |
| 10識別力 | 0.540 | 0.379 | 0.599 | 0.481 | 0.353 | 0.278 | 0.500 | 0.348 | 0.260 |

| 1級 | Q 4 | | | | | | Q 5 | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 54 | 55 | 56 | 57 | 58 | 59 | 64 | 65 | 66 |
| | C2 | C2 | C2 | B2 | B4 | A A | C2 | B4 | A A |
| 正答数 | 40 | 47 | 43 | 59 | 25 | 28 | 63 | 35 | 40 |
| 正答率 | 33.6% | 39.5% | 36.1% | 49.6% | 21.0% | 23.5% | 52.9% | 29.4% | 33.6% |
| 識別力 | 0.508 | 0.475 | 0.309 | 0.262 | 0.342 | 0.337 | 0.367 | 0.492 | 0.177 |
| 10正答数 | 13 | 19 | 16 | 28 | 10 | 8 | 30 | 13 | 22 |
| 10正答率 | 21.0% | 30.6% | 25.8% | 45.2% | 16.1% | 12.9% | 48.4% | 21.0% | 35.5% |
| 10識別力 | 0.292 | 0.220 | 0.224 | 0.127 | 0.274 | 0.102 | 0.417 | 0.135 | 0.220 |

3級：9項目／4項目 2級：6項目／5項目 1級：4項目／12項目

これより、まず以下のことが言える。

1. 1級では識別力の高くなった項目が3倍に増加しているのに対して、2級ではほぼ同数、逆に3級では識別力が低くなった項目が倍以上に増えている。

さらに、表1から以下の点が指摘できよう。

2. 識別力が0.3以下で、小池(2010)と同じ傾向だった項目は、以下の10項目である。
3級：16、23、24、61 2級：17、32、63 1級：12、57、66
3. 小池(2010)では識別力が0.3以下だったが、今回は0.3以上になった項目は以下の11項目であった。
3級：なし 2級：4、52 1級：11、14、36、54、55、56、58、59、65
4. 3級において、小池(2010)で識別力が低かった4項目(16, 23, 24, 61)は、いずれも今回の識別力も低かった。また、文法問題(Q4)では、小池(2010)ではすべての項目で識別力は0.3以上であったが、今回は4つの項目(42, 44, 45, 46)で0.3以下となり、中でも設問45や46は0に近い値となっている。
5. 1級の文法問題(Q4)において、小池(2010)では全項目が識別力0.3以下であったが、今回は設問57のみであった。

以上より、今回の結果の最大の特徴としては、<1級の設問項目の識別力は上昇したが、3級の設問項目のそれは下降した>という点であろう。

識別力は、正答率が高すぎたり低すぎたりした場合に低くなり、また正答率が50%に近い場合は高くなる傾向があるとされる(日本語教育学会編 1991)が、今回3級で識別力の低い設問項目の正答率を見ると、必ずしも正答率が高い項目だけではなく、設問44, 45, 46などは正答率が5割程度でも識別力が低かった。なお、設問44, 45, 46は、自他動詞+テイル、アスペクト、理由を表わすデスの問題であった。

5.1.2 受験者の結果

表2(次頁)は、異なり受験者の全体の正答率および各級ごとの正答率を百分率で示したものである。全体の平均点は37点(56.1%)で、級別の平均点は3級[16点満点]:11点(68.8%)、2級[25点満点]:15点(60.0%)、1級[25点満点]:10点(40.0%)であった。

表中の太枠で囲まれた斜体太字の数値は、当該級内で7割以上の得点であったことを示している⁷。「認定級」欄の「2/08」(第1位の受験者、以下同)とは、当該受験者が2008年度の旧JLPTにて2級を認定されたことを表わす。同じく「1不/09」(第2位)は2009年度に1級を受験したが不合格であったことを示している。「1/?」は1級の認定を受けてはいるが、それが何年のことなのか回答が記入されていなかったものである。

表2 受験者の得点傾向

| 順位 | 国籍 | 認定級 | 正答率 | | | |
|-----|--------|-------|-----|-----------|-----------|-----------|
| | | | 66問 | 3級 16問 | 2級 25問 | 1級 25問 |
| 1 | ベトナム | 2/08 | 94 | 88 | 100 | 92 |
| 2 | 中国 | 1不/09 | 85 | 88 | 84 | 84 |
| 3 | 台湾 | 1/06 | 85 | 81 | 92 | 80 |
| 4 | 中国 | | 83 | 94 | 72 | 88 |
| 5 | 中国 | 1/07 | 82 | 94 | 84 | 72 |
| 6 | 中国 | | 82 | 94 | 76 | 80 |
| 7 | 中国 | 1/05 | 82 | 88 | 92 | 68 |
| 8 | 中国 | 1/06 | 80 | 81 | 84 | 76 |
| 9 | 中国 | | 79 | 88 | 80 | 72 |
| 10 | 中国 | 1/07 | 79 | 75 | 80 | 80 |
| 11 | 中国 | 1/09 | 79 | 75 | 76 | 84 |
| 12 | 台湾 | 1/08 | 77 | 100 | 80 | 60 |
| 13 | 中国 | 1/04 | 77 | 88 | 84 | 64 |
| 14 | 台湾 | 1不/09 | 77 | 81 | 68 | 84 |
| 15 | 中国 | 2/? | 76 | 94 | 88 | 52 |
| 16 | ベトナム | 2/08 | 76 | 88 | 92 | 52 |
| 17 | 中国 | 1/? | 76 | 88 | 80 | 64 |
| 18 | 韓国 | | 76 | 81 | 84 | 64 |
| 19 | 中国 | 1/? | 76 | 63 | 92 | 68 |
| 20 | 台湾 | | 74 | 88 | 76 | 64 |
| 21 | 中国 | 1/08 | 74 | 69 | 80 | 72 |
| 22 | 中国 | 1/06 | 73 | 81 | 76 | 64 |
| 23 | 中国 | 1不/07 | 73 | 69 | 84 | 64 |
| 24 | 韓国 | 1/? | 71 | 75 | 80 | 60 |
| 25 | 中国 | 1/08 | 71 | 75 | 80 | 60 |
| 26 | 中国 | 1/08 | 70 | 81 | 72 | 60 |
| 27 | 台湾 | 1/07 | 70 | 69 | 72 | 68 |
| 28 | 韓国 | 2/? | 70 | 63 | 92 | 52 |
| 29 | カザフスタン | 2/? | 68 | 94 | 80 | 40 |
| 30 | 中国 | 1/? | 68 | 88 | 84 | 40 |
| 31 | 中国 | 1/06 | 68 | 81 | 68 | 60 |
| 32 | エジプト | 3/? | 67 | 100 | 80 | 32 |
| 33 | 中国 | 2/? | 67 | 81 | 72 | 52 |
| 34 | 中国 | 1不/? | 67 | 75 | 80 | 48 |
| 35 | 中国 | 2/? | 65 | 81 | 80 | 40 |
| 36 | 中国 | 2/08 | 65 | 81 | 76 | 44 |
| 37 | ベトナム | 2/08 | 65 | 69 | 80 | 48 |
| 38 | 中国 | 1/08 | 65 | 69 | 76 | 52 |
| 39 | 中国 | 1/? | 64 | 88 | 68 | 44 |
| 40 | マレーシア | 2/08 | 64 | 75 | 76 | 44 |
| 41 | 中国 | 1/? | 62 | 81 | 72 | 40 |
| 42 | 中国 | | 62 | 75 | 76 | 40 |
| 43 | 中国 | 1不/07 | 62 | 63 | 76 | 48 |
| 44 | 中国 | 1/08 | 62 | 44 | 84 | 52 |
| 45 | エジプト | 3/? | 61 | 94 | 68 | 32 |
| 46 | 中国 | | 61 | 81 | 64 | 44 |
| 47 | 中国 | 2/08 | 61 | 75 | 68 | 44 |
| 48 | インドネシア | 1不/? | 61 | 75 | 60 | 52 |
| 49 | 中国 | 1不/07 | 61 | 69 | 72 | 44 |
| 50 | 中国 | 1/09 | 61 | 63 | 76 | 44 |
| 51 | 中国 | 1不/? | 59 | 81 | 72 | 32 |
| 52 | ルーマニア | | 58 | 88 | 72 | 24 |
| 53 | 中国 | 2/? | 58 | 63 | 72 | 40 |
| 54 | 中国 | 2/? | 58 | 63 | 64 | 48 |
| 55 | カザフスタン | 2不/? | 56 | 81 | 72 | 24 |
| 56 | 韓国 | | 56 | 69 | 72 | 32 |
| 57 | カザフスタン | 3/06 | 55 | 81 | 56 | 36 |
| 58 | 中国 | | 55 | 75 | 52 | 44 |
| 59 | 韓国 | | 55 | 69 | 64 | 36 |
| 60 | ミャンマー | 2/? | 55 | 69 | 56 | 44 |
| 61 | 韓国 | 3/08 | 55 | 69 | 56 | 44 |
| 62 | 中国 | | 55 | 63 | 60 | 44 |
| 63 | 中国 | 1不/? | 53 | 69 | 60 | 36 |
| 64 | 中国 | 2/? | 53 | 44 | 68 | 44 |
| 65 | エジプト | 3/07 | 52 | 88 | 60 | 20 |
| 66 | セルビア | | 52 | 81 | 68 | 16 |
| 67 | 中国 | | 52 | 75 | 52 | 36 |
| 68 | スロバキア | 3/05 | 52 | 69 | 68 | 24 |
| 69 | 中国 | | 52 | 63 | 68 | 28 |
| 70 | タジキスタン | | 52 | 63 | 56 | 40 |
| 71 | 中国 | 2/? | 52 | 63 | 56 | 40 |
| 72 | ウクライナ | 2/? | 50 | 81 | 52 | 28 |
| 73 | 中国 | | 50 | 50 | 72 | 28 |
| 74 | 中国 | | 50 | 44 | 56 | 48 |
| 75 | ブラジル | 2/? | 48 | 94 | 44 | 24 |
| 76 | エストニア | | 48 | 88 | 60 | 12 |
| 77 | ウクライナ | 2/07 | 48 | 75 | 60 | 20 |
| 78 | 中国 | | 48 | 69 | 48 | 36 |
| 79 | 韓国 | 2不/07 | 48 | 63 | 64 | 24 |
| 80 | 中国 | | 48 | 63 | 56 | 32 |
| 81 | 中国 | | 48 | 56 | 52 | 40 |
| 82 | 韓国 | | 48 | 56 | 48 | 44 |
| 83 | 中国 | | 47 | 88 | 36 | 32 |
| 84 | 台湾 | | 47 | 56 | 60 | 28 |
| 85 | 中国 | 1不/06 | 45 | 63 | 52 | 28 |
| 86 | 韓国 | 3/08 | 45 | 63 | 52 | 28 |
| 87 | リトアニア | | 44 | 75 | 44 | 24 |
| 88 | 中国 | 1/? | 44 | 69 | 40 | 32 |
| 89 | 韓国 | | 44 | 63 | 52 | 24 |
| 90 | カザフスタン | 3/07 | 44 | 63 | 44 | 32 |
| 91 | 中国 | | 44 | 56 | 36 | 44 |
| 92 | リトアニア | | 42 | 75 | 40 | 24 |
| 93 | タイ | 3/07 | 42 | 75 | 36 | 28 |
| 94 | 韓国 | | 42 | 63 | 44 | 28 |
| 95 | 中国 | | 42 | 56 | 52 | 24 |
| 96 | 中国 | 3/? | 41 | 50 | 56 | 20 |
| 97 | イタリア | | 39 | 81 | 40 | 12 |
| 98 | クロアチア | | 39 | 56 | 52 | 16 |
| 99 | チュニジア | | 39 | 56 | 44 | 24 |
| 100 | スロベニア | | 39 | 56 | 32 | 36 |
| 101 | カザフスタン | 2/06 | 39 | 50 | 44 | 28 |
| 102 | カザフスタン | 3/? | 38 | 75 | 40 | 12 |
| 103 | 韓国 | 2/09 | 38 | 50 | 40 | 28 |
| 104 | 台湾 | 3/? | 38 | 44 | 32 | 40 |
| 105 | ドイツ | 3/08 | 36 | 81 | 28 | 16 |
| 106 | カザフスタン | | 36 | 69 | 40 | 12 |
| 107 | 韓国 | | 36 | 50 | 36 | 28 |
| 108 | アメリカ | | 35 | 63 | 36 | 16 |
| 109 | 韓国 | | 35 | 50 | 44 | 16 |
| 110 | カナダ | | 35 | 50 | 24 | 36 |
| 111 | 中国 | | 33 | 50 | 32 | 24 |
| 112 | 中国 | 2/? | 33 | 44 | 28 | 32 |
| 113 | 中国 | 1/05 | 33 | 19 | 52 | 24 |
| 114 | スペイン | 2不/? | 32 | 63 | 24 | 20 |
| 115 | エストニア | | 30 | 44 | 36 | 16 |
| 116 | ラトビア | | 26 | 44 | 16 | 24 |
| 117 | タイ | | 26 | 31 | 24 | 24 |
| 118 | 中国 | | 23 | 38 | 20 | 16 |
| 119 | タイ | 2不/08 | 18 | 31 | 20 | 8 |

順位は、正答率を基に順次3級・2級・1級での正答率順で行なった。これは、本稿の分析が、5.2で述べるRasch Modelの前提と同じ立場を取るためである。

さて、表2より、まず次の二つのことがわかる。

1. 3級の正答率が70%以上だった者は58名、以下2級では49名、1級では12名であった。このうち、上位級では7割以上の正答率であるのに、下位級では7割に満たないといった、いわゆる「ねじれ」を起こしている者が15名いた。
2. 3級の正答率が60%以下だった者が27名(22.7%)と、4分の1弱を占める。

さらには、次の点も指摘できよう。

- 3-1. 上位10%の者(第11位まで)のうち、第7位を除き、3級から1級のすべてについて70%以上の正答率であった。
- 3-2. 3-1より、上位10%の者(第11位まで)は事実上1級相当(以上)の日本語能力を有している可能性がある。
- 4-1. 上位10%を除く、その他の上位半数(第12位から第56位まで)の者は、2級相当の日本語能力は有していると思わせる。
- 4-2. 第19位から第58位の間には、2級は70%以上の正答率だが、3級では70%に至らない者が14名いた。ただし、そのうち正答率が69%(16問中11問正答)だった者が7名おり、他も60%代がほとんどであった。しかし、第44位のように、2級は84%の正答率だが3級は44%(16問中7問正答)と、大きな開きがある者もいた。
5. 第61位以下では、ところどころ3級で7割以上の正答率であった者がいるが、概して各級の正答率は低い。
6. 上位級を有しているからと言って、下位級の者より高い正答率になるとは限らない。たとえば、第113位や第88位は1級認定者だが、各級とも低い正答率であった。もっとも認定級については自己申告であり、筆者も厳密に確認は取っていないので、本人の誤記入ということも考えられないこともない。

以上より、今回の日能テストの結果、筆者が担当していたクラス(J600レベル)の受講生は、等しくJ600と位置づけられてはいたが、旧JLPT的に見ると、上位10%は1級相当の、続く40%は2級相当、続く25%は3級相当、残りの25%は3級も危ういレベルの受講生であったと言える。

2級相当とされたJ600レベルであったが、3級レベルもしくはそれ以下の受講生が半数近くもいたことは、今後のプレイズメント・テストに何らかの修正を求める必要があろう。

また、5.1.1の結果と関連づけるならば、表1において、1級で識別力の低い設問数が少なかったことより、今回の日能テストは能力の高い受講生を妥当に判定できる設問であったと言える。その反面、3級では識別力の低い設問数が増加していた。3級レベルの設問は、2級レベル相当の本クラスの受講生からすればやさしい設問であり、それゆえ正答率

も極端に高くなれば識別力が低くなることが考えられるが、今回の結果（表1）を見るとそうとは言えない。特に、文法問題（Q4）では3級レベルであるのに正答率が50%の設問が多かった（設問45～47）。これらの結果は、受講生が学習を進めていくうちに、基礎的な部分を欠落させている可能性があることを示唆しているように考えられる。このような基礎事項の欠落をどのように測り、そのような学生をどのように位置づけるかは今後の課題となろう。

5.2 項目応答理論による分析

項目応答理論 (Item Response Theory : IRT) による分析を行なった。分析には、大友賢二氏によって開発されたTDAP (Test Data Analysis Program) の修正版であるTDAP Ver.2.02 を用いた (伊東2008所収)。

項目応答理論では、ロジット得点 (logit score) を基に、「その設問項目はどのような特性を持っているか」「受験者の能力はどのくらいか」について推定を行ない、パラメータ (parameter) と呼ぶ推定値を算出する。パラメータは、絶対的な原点を持つ間隔尺度 (等間隔の尺度) である (中村・大友 2002)。

項目応答理論は、どのようなテストでも同一の尺度に沿って受験者能力の推定ができ、同一の能力を持つ受験者ならば、異なったテストを受験しても同一の受験者能力パラメータが与えられる。つまり、テストに依存しない受験者能力の推定ができるのである。

本稿では、項目応答理論の分析のうち、設問項目の最終項目困難度パラメータ (Final Calibration) と標準誤差 (Standard Error)、およびモデルとの適合度検討 (Analysis of Fit) の結果を示す。

設問項目の最終項目困難度パラメータと標準誤差の分析では、各設問項目ごとに項目困難度パラメータおよびその標準誤差が出力される。最終項目困難度パラメータは「平均的な困難度」の絶対的な原点0を持ち、負の値は「平均より易しい」度合い、正の値は「平均より難しい」度合いを示す。項目困難度は、通常-3.000から+3.000の間の値を取る。

モデルとの適合度検討は、それぞれの受験者がそれぞれの設問項目に対してどのくらいの確率で正答できるかというパターンを予測するモデル (これは項目応答理論において設定される) を用い、そのモデルと実際の結果 (受験者の解答傾向) との適合度 t を見るものである。

このモデル (Rasch Model) は、基本的に、

①能力の高い受験者の設問項目に正解する確率は、どのような項目においても、能力の低い受験者よりも高い。

②どのような受験者でも、難しい項目よりはやさしい項目に正解する確率は高い。という前提に成り立っている。そして、ここから導き出される正答確率と実際の解答とを

比較し、そのズレを適合度 t として示すのである。

t が +2 以上の場合は「ミスフィット (misfit)」と呼ばれ、モデルと実際の解答のズレが想定よりも大きすぎていることを示す (たとえば、能力の高い受験者の誤答が多く、逆に能力の低い受験者の正答が多い場合など)。また、 t が -2 以下の場合は「オーバーフィット (overfit)」と呼ばれ、モデルと実際の解答のズレが想定よりも小さすぎること示している。ある項目の解答パターンが、能力の高い受験者がすべて正答し、低い受験者がすべて誤答するといったような、ある一点ではっきりと分かれてしまうような場合にオーバーフィットになりやすい。なお、以下の表 3～7 中の「Sigma (z^2)」は t の計算途中の数値である。

以下、問 1 から問 5 の大問ごとに、設問項目の最終項目困難度パラメータと標準誤差およびモデルとの適合度検討の結果を示す。

5. 2. 1 「漢字の読み問題」の分析

表 3 漢字の読み問題 (問 1) の結果

| Item No. | Final Calib. | Standard Error | Sigma (Z^2) | t |
|----------|--------------|----------------|-----------------|-------------------------|
| 1 | -1.889 | 0.278 | 178.888 | <u>3.647</u> ← misfit |
| 2 | -0.823 | 0.220 | 87.450 | <u>-2.079</u> ← overfit |
| 3 | -1.075 | 0.230 | 90.527 | -1.846 |
| 4 | -1.023 | 0.227 | 127.164 | 0.651 |
| 5 | -0.921 | 0.223 | 75.660 | <u>-3.018</u> ← overfit |
| 6 | -1.128 | 0.232 | 92.172 | -1.724 |
| 7 | 0.223 | 0.204 | 96.150 | -1.432 |
| 8 | 0.015 | 0.204 | 115.627 | -0.090 |
| 9 | -0.027 | 0.205 | 133.229 | 1.027 |
| 10 | 0.779 | 0.211 | 133.229 | 1.027 |
| 11 | 0.869 | 0.213 | 131.208 | 0.903 |
| 12 | 0.391 | 0.205 | 146.841 | 1.844 |
| 13 | 1.408 | 0.232 | 119.086 | 0.136 |
| 14 | 1.951 | 0.262 | 124.424 | 0.478 |
| 15 | 1.252 | 0.225 | 107.044 | 0.666 |

表 3 を見ると、漢字の読み問題は特に難しくもやさしくもなかったようである。モデル適合度 t では、設問 1 がミスフィット、設問 2 と 5 がオーバーフィットであった。

設問 1 は比較的やさしい設問だったにもかかわらず、成績上・中位の受験者で誤答があったため、ミスフィットになったものと考えられる。

1. 私は北区にある学校に通っています。〔3級〕

通って 1 かゆって 2 かよって 3 とうって 4 とおって
 (0.0) (84.0) (2.5) (12.6) (NA0.8)

| No. 1 | Ss | NA | 1 | (2) | 3 | 4 |
|-------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 0 | 31 | 0 | 1 |
| MID | 55 | 1 | 0 | 44 | 2 | 8 |
| LOW | 32 | 0 | 0 | 25 | 1 | 6 |
| TOT | 119 | 1 | 0 | 100 | 3 | 15 |
| p | | 0.008 | 0.000 | 0.840 | 0.025 | 0.126 |

注) () で括られた数字は正答を示す。また、「Ss (Number of examinees)」の「UPP (Upper group)」の“32”は成績上位 (27%) の者が32名いたことを意味する。以下、「MID (Middle group)」は中位 (46%) の者55名、「LOW (Lower)」は下位 (27%) の者32名を意味する。「NA (No answer)」は無回答者数、「TOT (Total)」は合計数、「P (Item difficulty index)」は選択比率である (以下同)。

これに対し、設問 2 や 5 では、成績上位者と下位者の成績の違いが明確に出ていることがオーバーフィットの判定につながったのであろう。ここでは設問 5 の結果を示す。

5. その小さなミスが悲劇を招く原因となった。〔2級〕

招く 1 はぶく 2 まねく 3 いだく 4 のぞく
 (10.1) (70.6) (11.8) (5.0) (NA2.5)

| No. 5 | Ss | NA | 1 | (2) | 3 | 4 |
|-------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 0 | 32 | 0 | 0 |
| MID | 55 | 0 | 6 | 43 | 5 | 1 |
| LOW | 32 | 3 | 6 | 9 | 9 | 5 |
| TOT | 119 | 3 | 12 | 84 | 14 | 6 |
| p | | 0.025 | 0.101 | 0.706 | 0.118 | 0.050 |

5. 2. 2 「漢字の書き問題」の分析

表 4 漢字の書き問題 (問 2) の結果

| No. | F.C. | S.E. | Sigma (Z ²) | t |
|-----|---------------|------------|-------------------------|-------------------------|
| 16 | -1.130 | 0.335 | 153.658 | <u>3.636</u> ← misfit |
| 17 | <u>-2.744</u> | 0.573 ← 易問 | 262.333 | <u>9.399</u> ← misfit |
| 18 | -0.731 | 0.300 | 49.306 | <u>-4.068</u> ← overfit |
| 19 | -0.046 | 0.257 | 56.733 | <u>-3.313</u> ← overfit |
| 20 | -0.046 | 0.257 | 55.833 | <u>-3.401</u> ← overfit |
| 21 | <u>2.037</u> | 0.238 ← 難問 | 97.656 | 0.047 |
| 22 | <u>2.660</u> | 0.264 ← 難問 | 162.653 | <u>4.157</u> ← misfit |

漢字の書き問題は、7問中6問がミスフィットもしくはオーバーフィットの判定であった。ミスフィットは設問16, 17, 22、オーバーフィットは設問18, 19, 20であった。中でも、設問17の t が極端に大きい。

17. 夫婦はお互いにりかいしあうことが必要だ。〔2級〕

りかい 1 理解 (1.7) 2 埋触 (0.8) 3 理触 (0.8) 4 理解 (96.6)

| No. 17 | Ss | NA | 1 | 2 | 3 | (4) |
|--------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 1 | 0 | 0 | 31 |
| MID | 55 | 0 | 0 | 0 | 0 | 55 |
| LOW | 32 | 0 | 1 | 1 | 1 | 29 |
| TOT | 119 | 0 | 2 | 1 | 1 | 115 |
| p | | 0.000 | 0.017 | 0.008 | 0.008 | 0.966 |

この項目は、全体の正答率が96.6%で、誤答者は4名だけであった。下位者でもほとんどの者が正答しているということで、ミスフィットと判定されたと考えられる。

オーバーフィットの判定だった設問18, 19, 20では、上位者はすべての設問で全員正答であった。最終項目困難度パラメータ (F.C) は0に近い数値であることから考えると、問題の難易度は難しくもなくやさしくもないものであったが、上位者の正答数が100%であったということで、オーバーフィットの判定となったのであろう。

漢字の書き問題ではミスフィットになった項目が多く、今後の修正が必要である。

5.2.3 「語彙問題」の分析

表5 語彙問題 (問3) の結果

| No. | F.C. | S.E. | Sigma (Z^2) | t |
|-----|---------------|------------|-----------------|-------------------------|
| 23 | <u>-2.426</u> | 0.365 ← 易問 | 75.065 | <u>-3.068</u> ← overfit |
| 24 | -0.840 | 0.233 | 137.076 | 1.262 |
| 25 | -0.484 | 0.219 | 129.293 | 0.784 |
| 26 | 0.263 | 0.204 | 146.546 | 1.827 |
| 27 | -0.207 | 0.211 | 119.795 | 0.182 |
| 28 | -0.786 | 0.231 | 185.292 | <u>3.990</u> ← misfit |
| 29 | 0.552 | 0.203 | 106.617 | -0.695 |
| 30 | -0.436 | 0.217 | 110.688 | -0.418 |
| 31 | 0.053 | 0.206 | 88.359 | <u>-2.010</u> ← overfit |
| 32 | 1.621 | 0.226 | 161.588 | <u>2.692</u> ← misfit |
| 33 | 0.138 | 0.205 | 124.198 | 0.464 |
| 34 | -1.464 | 0.271 | 84.979 | <u>-2.270</u> ← overfit |
| 35 | 1.191 | 0.212 | 73.243 | <u>-3.222</u> ← overfit |
| 36 | 1.570 | 0.224 | 109.956 | -0.468 |
| 37 | 1.282 | 0.215 | 110.132 | -0.456 |
| 38 | 1.102 | 0.210 | 105.980 | -0.739 |
| 39 | -1.129 | 0.249 | 114.449 | -0.168 |

語彙問題では、17問中6問にミスフィットもしくはオーバーフィットの判定が出た。ミスフィットは設問28と32、オーバーフィットは設問23, 31, 34, 35であった。

ミスフィットの具体的な設問は以下の通りである。

28. A 「お茶のおかわり、いかがですか？」

B 「あ、もう、()。そろそろ帰りますので。」 [2級]

- 1 ごえんりょなく (17.6) 2 おまたせしました (3.4)
3 かしこまりました (4.2) 4 おかまいなく (73.9) (NA0.8)

| No. 28 | Ss | NA | 1 | 2 | 3 | (4) |
|--------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 5 | 0 | 0 | 27 |
| MID | 55 | 0 | 9 | 3 | 2 | 41 |
| LOW | 32 | 1 | 7 | 1 | 3 | 20 |
| TOT | 119 | 1 | 21 | 4 | 5 | 88 |
| p | | 0.008 | 0.176 | 0.034 | 0.042 | 0.739 |

32. A 「あれ、小林くんは？」

B 「小林くんなら、() 帰りましたよ。」 [2級]

- 1 さらに (19.3) 2 どこかに (14.3) 3 いまに (37.0) 4 とつくに (28.6) (NA0.8)

| No. 32 | Ss | NA | 1 | 2 | 3 | (4) |
|--------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 3 | 1 | 15 | 13 |
| MID | 55 | 0 | 8 | 8 | 22 | 17 |
| LOW | 32 | 1 | 12 | 8 | 7 | 4 |
| TOT | 119 | 1 | 23 | 17 | 44 | 34 |
| p | | 0.008 | 0.193 | 0.143 | 0.370 | 0.286 |

この二つの設問項目は、表1でも識別力が低かった。このことと併せて考えると、成績上位者では誤答の多さが、成績下位者では正答の多さが、適合度tを逸脱したために、このような判定となったのであろう。

オーバーフィットの判定であった設問23, 31, 34, 35のうち、設問23と34は成績上位者の正答率が100%であったため、また設問35では下位者の正答者数が0であったためと考えられる。設問31は、上位者32名中31名が正答であり、下位者は32名中7名が正答で、16名が誤答選択肢3を選んでいたので、成績の上下により正誤がはっきりと分かれたためであろう。

5.2.4 「文法問題」の分析

文法問題ではおおむねモデルに適合していた設問が多かったが、3級レベルの設問45でミスフィットが見られた。

表6 文法問題（問4）の結果

| No. | F.C. | S.E. | Sigma (Z^2) | t |
|-----|---------------|------------|-----------------|-----------------------|
| 40 | -0.596 | 0.202 | 118.438 | 0.028 |
| 41 | -0.891 | 0.210 | 108.910 | -0.604 |
| 42 | <u>-2.235</u> | 0.295 ← 易問 | 97.083 | -1.430 |
| 43 | -0.719 | 0.205 | 114.189 | -0.250 |
| 44 | 0.221 | 0.197 | 130.577 | 0.798 |
| 45 | 0.260 | 0.197 | 167.551 | <u>2.959</u> ← misfit |
| 46 | 0.106 | 0.196 | 138.145 | 1.261 |
| 47 | -0.804 | 0.207 | 104.218 | -0.926 |
| 48 | -1.719 | 0.252 | 90.400 | -1.922 |
| 49 | -0.678 | 0.204 | 124.815 | 0.437 |
| 50 | -0.678 | 0.204 | 104.925 | -0.877 |
| 51 | 0.538 | 0.202 | 115.599 | -0.157 |
| 52 | 0.923 | 0.213 | 110.736 | -0.480 |
| 53 | 1.662 | 0.252 | 116.826 | -0.077 |
| 54 | 0.747 | 0.207 | 102.873 | -1.019 |
| 55 | 0.457 | 0.200 | 103.327 | -0.988 |
| 56 | 0.620 | 0.204 | 125.338 | 0.471 |
| 57 | -0.010 | 0.196 | 114.416 | -0.235 |
| 58 | 1.481 | 0.240 | 115.490 | -0.164 |
| 59 | 1.315 | 0.231 | 105.583 | -0.831 |

45. あ、お金が（ ）。〔3級〕

- 1 落としています (10.9) 2 落としてあります (18.5)
 3 落ちています (43.7) 4 落ちてあります (25.2) (NA1.7)

| No. 45 | Ss | NA | 1 | 2 | (3) | 4 |
|--------|-----|-------|-------|-------|-------|-------|
| UPP | 32 | 0 | 2 | 6 | 20 | 4 |
| MID | 55 | 2 | 7 | 7 | 22 | 17 |
| LOW | 32 | 0 | 4 | 9 | 10 | 9 |
| TOT | 119 | 2 | 13 | 22 | 52 | 30 |
| p | | 0.017 | 0.109 | 0.185 | 0.437 | 0.252 |

設問45は上位者でも誤答者の比率が高いことから、いわゆる「状態の継続」を表わすテイルが上位者でも定着が不安定なことがわかる。3級レベルの設問であり、また日本語学習上重要な項目であるテイルについて、上位者でも誤答が多いという今回の結果は、基礎的な項目が成績上位学習者であっても未定着でもあり不安定でもあるということを示している。

5.2.5 「語の使用法」の分析

語が正しく用いられているものを選ぶ問題では、最終項目困難度パラメータもモデル適合度も特に問題のある設問項目はなかった。

表7 語の使用法の問題（問5）の結果

| No. | F.C. | S.E. | Sigma (Z^2) | t |
|-----|--------|-------|-----------------|--------|
| 60 | -1.436 | 0.280 | 118.168 | 0.340 |
| 61 | -1.882 | 0.321 | 115.215 | 0.147 |
| 62 | 0.630 | 0.204 | 108.819 | -0.281 |
| 63 | -0.673 | 0.232 | 105.932 | -0.478 |
| 64 | 0.381 | 0.204 | 110.395 | -0.174 |
| 65 | 1.611 | 0.225 | 90.229 | -1.603 |
| 66 | 1.369 | 0.216 | 122.194 | 0.600 |

5.3 低識別力とミスフィットが重なった項目

5.1の古典的テスト理論による分析において識別力が低かった項目と、5.2の項目応答理論によるミスフィット項目について触れておく。

表1より、今回の調査で識別力が0.3以下だった19項目のうち、表3～7でミスフィットと判定された項目は7項目、すなわち問1の1、問2の16, 17, 22、問3の28, 32、問4の45であった。

ある設問項目の識別力が低くなるということとミスフィットになるということは、共に高得点者が誤答した設問項目を低得点者が正答したり、あるいはモデルの期待値以上に低得点者が高いレベルの設問項目に正答した場合に起こると言えるので、識別力の低かった項目はミスフィットになると考えられるが、今回の結果では必ずしもそうにはならなかった。この違いが何に起因しているのかについては、今後の検討課題としたい。

6. おわりに—今後の課題—

本稿では、旧JLPTの『報告書』の1級から3級までの文字・語彙・文法のデータを基に、日本語能力が不明もしくは曖昧な留学生を対象とするテストを試験的に作成し、実施したものの結果の一部を報告した。

テストは、2009年度の筑波大学留学生センターのJ600レベル（旧JLPT2級相当）の留学生を対象にしたものであったが、古典的テスト理論と項目応答理論の手法を用いて分析した結果、以下のような点が見いだされた。

まず、古典的テスト理論による分析では、3級レベルの設問項目で識別力の低いものが多数見られた。特に文法問題では7問中4問で識別力が0.3以下であった。正答率が50%前後の場合に識別力は高くなるとされるが、3級の文法問題である設問44, 45, 46では正答率

が40～50%であったのに識別力は低かった。

受験者の結果を見ると、1級相当の能力があると見なされる者から3級にも満たない可能性のある学生もいることがわかった。予想以上にばらつきが大きかったクラスであったと言えよう。

項目応答理論による分析では、語の使用法の問題（問5）以外ではミスフィット／オーバーフィットの判定となった項目が見られた。特に、漢字の書き問題（問2）と語彙問題（問3）では多く見られたので、今後の修正が必要であろう。また、文法問題（問4）ではミスフィット／オーバーフィットの数は少なかったとは言え、3級問題で、しかも重要な文法項目であるテイルの誤答が成績上位者に多く見られたことは、今後の文法教育（今回ではテイル）の定着を再考しなければならないと言える。

今後も継続してデータを収集し、他の能力を測る試験（SPOTなど）との比較検討も行うことにより、テストの安定性を図っていきたい。

注

- 1 筑波大学は春学期（4月から7月）・秋学期（9月から11月）・冬学期（12月から2月）の3学期制である。
- 2 1パラメータ・モデルの場合。大友（1996）参照。
- 3 2011年1月現在、2008年度版まで出版されている。2007年度版よりCD-ROMとなっている。
- 4 日能テスト実施時（2009年度）の補講クラスは、J100からJ800の八つのレベルであったが、2010年度よりカリキュラム変更により、補講クラスはJ100からJ900の九つのレベルに分けられ、J600は「中級前期」相当のレベルとなった。
- 5 2009年度における当クラスは、学期ごとに異なる内容を講義していたため、学期ごとに受講することが可能であった。しかし、2010年度より、毎学期同じ講義内容をするようにカリキュラムが変わり、学期末テストに合格すれば再び当クラスを受講する必要はなくなったため、再履修する学生はいなくなった。
- 6 旧JLPTの文字・語彙の設問項目数と試験時間は、1・2級は65問／45分、3級は55問／35分であった。
- 7 旧JLPT 1級の認定基準は正答率が7割以上、2・3級では6割以上が目安となっていた。

参考文献

伊東祐郎（2008）『留学生の日本語能力測定のためのテスト項目プールの構築』、平成16年度～平成19年度科学研究費補助金基盤研究（A）研究成果報告書（課題番号16202008）研

究代表者：伊東祐郎

大友賢二 (1996) 『項目応答理論入門』、大修館書店

小池康 (2010) 「日本語能力試験を援用した複数級間にわたる能力テストの試案—中上級レベルの学習者を対象に—」『筑波大学留学生センター日本語教育論集』25：87-105

中村洋一 (著)・大友賢二 (監修) (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門～』、桐原書店

日本語教育学会編 (1991) 『日本語テストハンドブック』、大修館書店

24. この肉はかたいので、よく() 食べてください。
 1 こんで 1.7
 2 のんで 13.4
 3 やんで 74.8
 4 かんて NA1.7
25. 庭にきれいな花を()。
 1 とりましよう 12.6
 2 うえましよう 68.9
 3 かえましよう 9.2
 4 はりましよう 5.0
26. 強い風で大きな木の枝が() しまった。
 1 やぶれて 12.6
 2 おれて 54.6
 3 たおれて 26.9
 4 こわれて 5.9
27. レポートは手書きでも可。() それに。
 1 だって 16.0
 2 それに 8.4
 3 そのうえ 11.8
 4 ただし 63.9
28. A 「お茶のおかわり、いかがですか?」
 B 「お、もう()。そろそろ帰りますので。」
 1 ごえんりよなく 17.6
 2 おまたせしました 3.4
 3 かしこまりました 4.2
 4 おかまいなく 73.9
29. 税金を()のは、国民の義務である。
 1 あずける 29.4
 2 かぞえる 48.7
 3 おさめる 4
 4 すませる 10.1
30. あの人は、こちらが何度だめだと言っても、またたのみに来る。本当に()人だ。
 1 こまかい 5.9
 2 すまない 10.9
 3 しつこい 68.1
 4 おもたい 13.4
31. こわい人かと思っていたら、() いい人だった。
 1 少々 3.4
 2 案内 58.8
 3 事実 31.1
 4 当然 5.9
32. A 「あれ、小林くんは?」
 B 「小林くんなら、() 帰りましたよ。」
 1 すでに 19.3
 2 どこかに 14.3
 3 いまに 37.0
 4 とつくに 28.6
33. このスneeプを作るには()も時間もかかります。
 1 手間 57.1
 2 手紙 10.9
 3 手入れ 23.5
 4 手段 6.7
34. 同じ英語()の国といっても、そこで使われている英語はさまざまだ。
 1 園 83.2
 2 産 1.7
 3 園 0.8
 4 籍 14.3
35. 彼女はいつもここにきていて()がいい。
 1 感情 34.5
 2 愛想 36.1
 3 愛情 10.9
 4 感想 16.8
36. 彼女は自分には才能がある() いる。
 1 おもむいて 27.7
 2 ひやかして 23.5
 3 うぬぼれて 29.4
 4 あつらえて NA8.4
37. このカメラは使い方が() ので、評判が悪い。
 1 あつかましい 18.5
 2 たくましい 11.8
 3 ややこしい 34.5
 4 ふさわしい NA7.6
38. 今後とも、() よろしくお願ひ申し上げます。
 1 何より 20.2
 2 何だか 12.6
 3 何でも 26.9
 4 何とぞ 37.8
39. レポートのために図書館から借りていた本を() した。
 1 返却 79.0
 2 返済 7.6
 3 返還 11.8
 4 返品 1.7
- Q 4 ()のところにどのことばを入れますか。1・2・3・4から一番いいものを選びなさい。**
40. 私は母()泣かせるようなことはしたくない。
 1 に 37.8
 2 を 62.2
 3 へ 0.0
 4 の 0.0
41. 私が弟のシャツを洗って()。
 1 くださった 2.5
 2 いただいた 0.0
 3 やった 68.1
 4 くれた 28.6
42. 荷物は私が来週の月曜日にお届け()。
 1 くださいます 5.0
 2 なさいます 4.2
 3 いたします 87.4
 4 ございます 3.4
43. ここに車を()。じやまだ。
 1 止まるな 27.7
 2 止めるな 64.7
 3 止めろな 4.2
 4 止まれな 3.4
44. これからパンを() ところです。
 1 焼ける 24.4
 2 焼く 44.5
 3 焼けている 12.6
 4 焼いている 18.5
45. あ、お盆が()。
 1 落ちています 10.9
 2 落ちてあります 18.5
 3 落ちています 43.7
 4 落ちてあります 25.2
46. A 「あした、いつしよに出かけますか。」
 B 「あしたは友だちがうちに来る予定()。」
 1 なんですから 44.5
 2 ので 7.6
 3 なんです 47.1
 4 からです 0.0
47. 通勤に車を使っていると、運動不足に()。
 1 しそうだ 5.0
 2 なりがちだ 66.4
 3 するべきだ 10.1
 4 なるせいで NA3.4

