

# 日本語基礎能力テストの項目プールの作成

平 直樹\*  
 前川 眞一\*\*  
 小野 博\*\*\*  
 林部 英雄\*\*\*\*  
 内田 照久\*\*\*

## 要 約

児童・生徒に対する基礎的な日本語の単語の理解力を測定するための適用範囲の広い日本語基礎能力テスト項目が開発された。本テストは2つの下位テストから成る。基本的な日本語の単語についての471項目の多枝選択形式の基本語彙テスト(語彙テスト)と977項目の自由記述形式の基本的な漢字の読みのテスト(漢字テスト)である。双方の下位テストとも、6歳程度から18歳程度までが適用範囲である。

それぞれの下位テストは11の版からなっており、小学校1年生から高校3年生までの児童・生徒に実施された。項目パラメタは約3,000人分のデータを用いて版毎に別々に尺度化され、その後共通尺度に等化された。項目プールの共通尺度においては、語彙テストの場合、各学年ごとの尺度値の平均は小学校1年生の80から高校3年生の120、漢字テストでは67から116の範囲に分布している。標準偏差は、語彙テストでは下の学年で約4、上の学年では10程度、漢字テストの場合は、下の学年で約13、上の学年で約4であった。

キーワード：項目プール、テストの等化、日本語能力テスト、語彙テスト、漢字テスト

## 1 問題と目的<sup>1</sup>

ことばの能力の測定は、それ自体も教育や研究のテーマと成り得るものである上に、他の能力や特性との関連の中で基礎的で重要な指標となり得るものである(平他, 1995)。しかしながら、日本語能力を簡便に測定するための標準化されたテストは入手しにくい状況にある。

本研究では、現在まで我々が一連の研究を通じ

て開発してきた小中学生用と高校生用の日本語基礎能力テストを等化して、共通の尺度上に表すことにより、広い能力範囲に適用可能な項目プールを作成した。

小中学生用項目は、助詞、文型などの理解を含む8領域の下位テストの一部である(小野他, 1989)。8領域のうち、高校生程度以上の問題が作成可能であると考えられたのは語彙と漢字の2領域であった。高校生用項目は、その結果を受け、語彙と漢字の領域に限って作成された(平他, 1992;

\* 大学入試センター研究開発部 試験方法研究部門  
 〒153-8501 東京都目黒区駒場2-19-23  
 Telefax: 03-5478-1297 e-mail: taira@rd.dnc.ac.jp

\*\* 大学入試センター研究開発部 進学適性研究部門

\*\*\* 大学入試センター研究開発部 特別試験研究部門

\*\*\*\* 横浜国立大学 教育人間科学部

平他, 1993a). 小中学生用の語彙と漢字, 高校生用の語彙と漢字という4つのテストは, いずれも項目反応理論に基づいて尺度化されている(小野他, 1989; 平他, 1995). 本研究で作成した項目プールに含まれる項目を「小野式日本語基礎能力テスト項目」と呼ぶこととする。

ことばの能力の定義は難しい問題であるが, 本研究は今までの研究(小野他, 1989; 平他, 1992; 平他, 1993a; 小野他, 1993; 平他, 1993b; 平他, 1995)と同じように, 紙筆テスト等の形式で簡単に実施できる基本的な要素である「語彙理解力」と「漢字の読み取り能力」を「日本語基礎能力」として代表させることとする。したがって, 基礎的なことばの能力の中でも, 会話力やコミュニケーション能力等の日本語の文字の解読を前提にしな

い能力を測定することはできない。しかしながら, 本研究で測定可能な能力も日本語能力の基礎として重要であることは言うまでもないところである。

小野式語彙テスト項目は, 全て五枝選択形式である。幹に示されたことばに最も近い意味の選択肢を選ぶ項目(図1参照)を中心に, 一部に絵を用いた項目(図2参照), また, 穴埋め問題形式の項目(図3参照)を含む。

また, 小野式漢字テスト項目は, 問題文の中の一部の漢字にふりがなを振る形式で項目が構成されている。漢字読み取りテストの項目の例は, 図4に示す通りである。

小野式日本語基礎能力テスト項目は項目反応理論(Lord & Novick, 1968; Lord, 1980; Ham-

(7) <small>じゅうじつ</small> 充実した生活	1. <small>まんぞく</small> 満足していること	2. <small>よく</small> 良くできていること	3. <small>りっぱ</small> 立派なこと
	4. <small>おんが</small> お金がたくさんあること	5. <small>こ</small> 内容が濃く豊かなこと	
(8) <small>けいぞく</small> 継続する	1. <small>あと</small> 後をつぐ	2. <small>あいだ</small> 間をうめる	3. つなげる
	4. やめる	5. <small>とぎ</small> 途切れないで続ける	
(9) <small>じんざい</small> 人材	1. 体力のある人	2. <small>のうりよく</small> 能力があり役に立つ人	3. <small>しんにゅうしゃいん</small> 新入社員
	4. <small>せいかく</small> <small>ちが</small> 性格の違う人たち	5. <small>りんぎょう</small> <small>かんけい</small> 林業に関係する人々	
(10) <small>かねがね</small> うかがっております	1. <small>いぜん</small> 以前から	2. <small>じゅうぶん</small> 十分	3. だいたい
	4. テレビで見て	5. うわさで聞いて	

図1 語彙テスト項目例

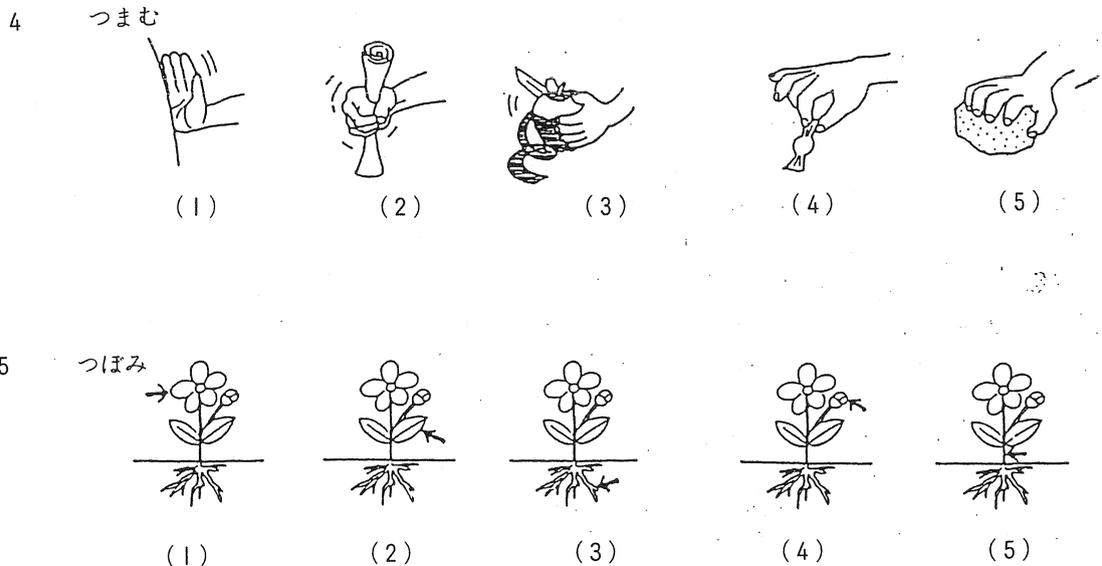


図2 語彙テスト項目例(絵を用いた項目)

- 48 人を呪わば ( ) 二つ  
 1. 瓜 (うり) 2. 山 (やま) 3. 足 (あし)  
 4. 穴 (あな) 5. 家 (いえ)
- 49 無い ( ) は振れぬ  
 1. 袂 (たもと) 2. 袖 (そで) 3. しっぽ  
 4. 旗 (はた) 5. 腕 (うで)

図3 語彙テスト項目例 (穴埋め項目)

- |          |          |             |             |        |            |        |            |
|----------|----------|-------------|-------------|--------|------------|--------|------------|
| (82)     | (81)     | (80)        | (79)        | (78)   | (77)       | (76)   | (75)       |
| 何回もかぞえる。 | 野原であそぼう。 | 五月五日は子どもの日。 | 同じかたちをあつめる。 | かわいい妹。 | ぬる色をかんがえた。 | 本を読もう。 | 字をていねいに書く。 |

図4 漢字テスト項目例

bleton & Swaminathan, 1985; 大友, 1996) に基づき, 困難度  $b$  と識別力  $a$  の2つの項目パラメータを持つ2パラメータ・ロジスティックモデルで尺度化されている。現在までに, 語彙理解力に関して項目反応理論を用いて作成されたテスト, および, 項目プールには, 芝 (1978), 服部 (1990) がある。小野式語彙テスト項目は, これらのものとは測定の内容やテスト項目の形式という側面で共通する。その上で, 本研究の項目プールは, より近年になって作成されたものであり, 尺度化に用いられた被験者の規模 (表1参照) やサンプリングの全国性, 高校生用版においては8名の高校国語教師が問題作成スタッフに加わるという項目作成者の専門性 (平他, 1995) の面で優れているという特徴を持っている。また, 漢字テスト項目に関しては, 現在まで類似のテストは作成されていないと思われる。

尺度化に項目反応理論を用いることにより, 様々な利点が得られることになる。まず第1に, 小学生程度の項目から高校生程度の項目まで同じ尺度上に表現されることで, 広い範囲でのテスト結

果の比較が可能になるということが挙げられる。このことは, 被験者の能力を表すパラメータ  $\theta$  についてだけではなく, 項目の困難度を表すパラメータ  $b$  についても同様である。さらに, 項目と被験者のパラメータが同一の尺度で表されていることから, 項目困難度の意味も把握しやすい。すなわち, ある項目の困難度  $b$  と同じ値の  $\theta$  を持つ被験者が当該項目に正答する確率がモデル上0.5であるので, 困難度の値は, その項目のレベルに相当する被験者の能力レベルを表すと考えればよい。

テストの利用の面では, 異なる項目を用いて測定された結果を直接比較できることが大きな利点である。すなわち, 被験者の特性に合わせて自由に適応形テスト (野口他, 1983; 平他, 1989) を作成することが可能であり, 個人個人が解答した項目が同一でなくともその結果を直接比較することが可能となるのである。

本研究で作成された項目プールから構成されるテストは, 例えば, ことばの能力の発達研究の測定用具として用いることが考えられる。同時に, 基礎的な日本語能力以外に焦点を当てた研究にとっても, 研究に関連する一つの重要な要因として被験者の言語能力を知っておきたい場合には, 十分利用可能な測定尺度と考えられる。例えば, 平 (1995) においては, 中学1年生と高校3年生の作文能力の評価を行った際, より基礎的な日本語能力の指標として, 小中学生用項目と等化される以前の高校生レベルの語彙テスト・漢字読み取りテストの項目プール (平他, 1995) から, 独自のテストを作成して用いている。その際, 事前の主観的な予測から, 中学1年生と高校3年生とで別々の項目を用いたテスト冊子を作成し, それぞれ精度の高い測定を行った上でその結果を比較している。

このように, 本研究の項目プールは, 様々な用途に応じたテストを構成することが可能であり, 幅広い応用範囲を持っている。

## 2 項目プールの作成の手続き

本研究の目的は, 前述したように, 小野他 (1989) の小中学生用の項目と, 平他 (1995) の高校生用の項目を等化することにより, 小学校1年生程度から高校生程度までの日本語基礎能力を測定する

テスト項目の項目プールを完成させることである。

項目プールの作成は、具体的には以下のような手続きに基づいて行った。

## 2.1 項目の再分析

本研究を行う以前にも、本研究と同様の方法で小中学生用項目と高校生用の項目との等化を試みたが、必ずしも望ましい結果が得られなかった(平他, 1993b)。その原因を検討するためにデータの再吟味を行った結果、語彙理解テスト、漢字読み取りテストともに、困難度が被験者集団の能力と合致せずに極端な正答率を持つ項目が多数含まれていたことが分かった。特に、小中学生用の項目に易しすぎる項目(正答率が極端に高い項目)が多かった。そこで、再び正答率に基づく古典的な項目分析を行い、極端な正答率の項目を尺度化から排除することとした。項目の内容との兼ね合いもあり、排除の基準は一律ではないが、易しすぎる項目の排除については、概ね正答率95%程度以上を目安とした。また、複数の学年にまたがるアンカー項目については、下の学年での正答率が85%程度以上の項目については内容を一つひとつ吟味し、ほとんどの項目について、項目プールから排除するか、一部の学年のデータを用いない等の措置を講じることとした。

なお、各版ごとの分析に用いたデータ数は、表1に示す通りである。

## 2.2 項目パラメタの再推定

小中学生用の項目については、一部の項目を取り除いた後、項目パラメタの再推定を行った。項目パラメタの推定のプログラムは、周辺最尤推定法によるプログラム BILOG (Mislevy & Bock, 1989) を用いた。

この手続きにおいては、それぞれの学年ごとに、能力パラメタ  $\theta$  の平均が0、分散が1となることを制約条件とした。

なお、中学生については、実施した項目は全年共通であるが、項目パラメタの再推定は学年毎に行った。

## 2.3 等化

版毎、学年毎の項目パラメタの再推定の後、小

表1 データ数

版	語彙	漢字	
小学校1年生	2,179人	2,694人	
小学校2年生	2,315人	2,906人	
小学校3年生	2,317人	3,012人	
小学校4年生	2,420人	3,140人	
小学校5年生	2,792人	3,232人	
小学校6年生	2,727人	3,347人	
中学生	合計 (1年生) (2年生) (3年生)	6,974人 (2,050人) (2,411人) (2,513人)	9,983人 (3,270人) (3,455人) (3,258人)
高校生 91年版	合計 (1年生) (2年生) (3年生)	18,293人* (5,784人) (6,273人) (6,236人)	12,487人* (3,324人) (4,347人) (4,816人)
高校生 92年A	合計 (1年生) (2年生) (3年生)	2,978人 (986人) (992人) (1,000人)	2,978人 (986人) (992人) (1,000人)
高校生 92年B	合計 (1年生) (2年生) (3年生)	2,822人 (953人) (955人) (914人)	2,822人 (953人) (955人) (914人)
高校生 92年C	合計 (1年生) (2年生) (3年生)	2,885人 (979人) (971人) (935人)	2,885人 (979人) (971人) (935人)

\*: 尺度化に際しては、それぞれ4,000名を無作為抽出した。

学生から高校生までの推定パラメタとアンカー項目の情報を元に、尺度の等化を行った。等化の方法の詳細については、次章で述べる。

## 2.4 項目の最終チェック

全ての手続きを終えたあと、項目パラメタのリストから、推定された項目パラメタの値が不自然な項目を洗い出し、一つ一つについて吟味した。その結果、項目自体の妥当性が疑問であったり、項目として不適切であると思われるものを削除し、最終的な項目プールが完成した。

### 3 等化の方法

テストを作成する場合に項目反応理論を用いる利点としては、似たような性質を持つテストの作成が容易であること、被験者の能力のレベルに合わせてきめ細やかなテストを作成することが可能であること、等があげられるが、いずれの場合にしても、各テストに含まれるべき項目の項目パラメタの値が既知であることが前提となる。すなわち、項目プール（項目バンク）を作成しなくてはならない。ここでいう項目プールとは、同一の尺度を用いて計測された、項目パラメタが既知の項目のセットのことである。

項目反応理論においては、能力値が  $\theta$  の被験者が第  $j$  番目の項目に正答する確率が、項目  $j$  の項目パラメタ ( $a_j, b_j$ ) を用いて、

$$\begin{aligned} \text{prob}(\text{項目 } j \text{ に正答} | \theta) \\ &= P(\theta | a_j, b_j) \\ &= \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))}, \quad a_j > 0 \quad (1) \end{aligned}$$

で与えられることを仮定しているが、この能力値  $\theta$  は潜在変数であり、実際に観測されるものではない。また上式の形から、能力値の原点と単位は任意であることが知られている。

項目プールを作成するためには大量の項目パラメタを推定しなければならないが、通常は、まず全ての項目をいくつかのテストに分配し、それぞれを、能力値の分布がかならずしも同一であるとは限らない集団に実施するという手続きを踏む場合が多い。しかし、上記の原点と単位の任意性から、各テストに含まれる項目のパラメタをテストごとに独立に推定すれば、テストごとに異なる尺度が形成されてしまう。したがって、項目プールを作成するためには、何らかの方法でこの原点と単位の不定性を取り除く作業、すなわち、項目パラメタの等化が必要となる。

表2と表3に示すように、小野式語彙テスト項目、漢字テスト項目においては、小学1年生用から高校生用まで合計11のテストが用いられているが、研究の初期の段階から、各テストの中にアンカーとなる項目を配置し、各テストごとに項目

パラメタを独立に推定したとしても、このアンカー項目の情報を用いて各テストのリンクを取ることが可能となるようなデザインを用いた。このようなアンカー項目を用いながら、独立に推定された項目パラメタを共通尺度に乗せる方法は共通項目を用いた項目パラメタの等化法と呼ばれている。

本研究で用いられた等化法は、前川(1991)に示されているものであるが、第  $g$  番目のテストのデータのみを用いて推定された項目  $j$  の項目パラメタを ( $a_{jg}, b_{jg}$ )、第  $g$  番目のテストを受けた被験者の能力分布の密度関数（本研究の場合には基準正規分布の密度関数）を  $h_g(\theta)$  とすれば、

$$ERSS = \sum_{\text{項目 } j \text{ がテスト } g \text{ に含まれる場合}} \int_{-\infty}^{+\infty} [P(\theta | a_{jg}, b_{jg}) - P(q_g + r_g\theta | a_j, b_j)]^2 h_g(\theta) d\theta \quad (2)$$

を最小とするように、全ての項目の共通尺度上での項目パラメタ ( $a_j, b_j$ ) と、各テストごとの尺度を共通尺度へ変換する係数 ( $q_g, r_g$ ) を推定するものである。なお、本研究のように、各テストごとに項目パラメタを推定する際に、その集団における能力値の平均をゼロ、分散を1とおいた場合には、 $q_g$  と  $r_g$  はそれぞれ共通尺度上での各集団の能力の平均と標準偏差を表すパラメタとなっている。また、このようにして得られる共通尺度の原点と単位も任意であるため、本研究では、被験者全体の平均がほぼ100、標準偏差が15となるように標準化を行った。

## 4 結 果

### 4.1 項目プール

本研究では、最終的に、語彙テスト項目471項目、漢字テスト項目977項目からなる項目プールが完成した。各版に含まれる項目数は、表2に示す通りである。また、1つの版のみに含まれる項目と2つ以上の版に含まれるアンカー項目との区別は、表3に示す。例えば、表2から、小学校2年生用の版に当初含まれていた語彙テストの項目は全部で50項目であったが、不良項目7項目が削除され、最終的に項目プールに採用されたのは43項目であったことが分かる。また、表3から

表2 各版に含まれる項目数

版	語彙	漢字
小学校1年生	30 (30) 項目	86 (100) 項目
小学校2年生	43 (50) 項目	102 (128) 項目
小学校3年生	59 (60) 項目	114 (143) 項目
小学校4年生	59 (60) 項目	120 (145) 項目
小学校5年生	59 (60) 項目	124 (145) 項目
小学校6年生	57 (60) 項目	107 (145) 項目
中学生	72 (80) 項目	170 (198) 項目
高校生91年版	62 (73) 項目	139 (140) 項目
高校生92年A	73 (75) 項目	122 (125) 項目
高校生92年B	73 (75) 項目	105 (105) 項目
高校生92年C	75 (75) 項目	103 (105) 項目

注: ( ) 内は当初数

は、その43項目のうち、小学校2年生の版のみに含まれていたのは僅かに1項目であり、23項目が1年生の版と共通、19項目が3年生の版と共通の項目であったことが分かる。

#### 4.2 標準化

等化を行った後、語彙理解テスト、漢字読み取りテストの双方について、表1に示した被験者全体の能力値がほぼ平均100、標準偏差15となるように線型変換した。なお、これは能力パラメタ $\theta$ の推定値に負の値が見られないようにすることを目的としたものである。例えば、よく用いられるウェクスラー法の「偏差IQ」は年齢ごとの平均、標準偏差が本研究の能力値の平均、標準偏差と同じ値となっている(東他, 1988)が、本研究の場合は、それとは定義が異なることに注意して欲しい。

各版の項目パラメタの平均と標準偏差は、表4に示すとおりである。

学年ごとの能力推定値の平均と標準偏差を表5に示す。さらに、中学生と高校生は版と学年が対応していないので、版ごとの値も示すこととする表4、表5を利用することによって、本研究のテ

表3 各版に含まれる項目数とアンカー項目数

版	語彙	漢字
小学校1年生のみ	7項目	57項目
小学校2年生のみ	1項目	48項目
小学校3年生のみ	20項目	48項目
小学校4年生のみ	20項目	57項目
小学校5年生のみ	21項目	66項目
小学校6年生のみ	10項目	63項目
中学生のみ	28項目	91項目
高校生91年版のみ	34項目	63項目
高校生92年Aのみ	58項目	102項目
高校生92年Bのみ	58項目	85項目
高校生92年Cのみ	60項目	83項目
小1, 小2	23項目	15項目
小2, 小3	19項目	10項目
小1, 小2, 小3	—	14項目
小3, 小4	20項目	13項目
小2, 小3, 小4	—	15項目
小4, 小5	19項目	12項目
小3, 小4, 小5	—	14項目
小5, 小6	19項目	16項目
小4, 小5, 小6	—	8項目
小4, 中学	—	1項目
小6, 中学	26項目	12項目
小6, 中学, 高91	2項目	—
小5, 小6, 中学	—	8項目
中学, 高91	11項目	56項目
中学, 高校全般	5項目	2項目
高校全般	10項目	18項目
合計	471項目	977項目

トを利用して能力の推定値を得た場合、学年標準を目安とした結果の判断が可能になる。

ちなみに、表6は前述の平(1995)の結果について、本研究で得られた項目パラメタを用いて再分析した結果である。平(1995)では、中学生以下のデータが等化されていなかったため、全て高校生用の項目が用いられた。また、被験者とした中学生の集団の能力については、「標準的な中学1年生としては非常に能力の水準が高い」と記述するに止まっていたが、同学年の被験者との比較、また、学年標準(平均)を元にした対応関係から、これらの被験者の日本語基礎能力についてより分かりやすい情報が得られている。すなわち、この集団の語彙理解力の平均は「高校1~2年生程度」、漢字読み取り能力の平均は「高校1年生程度」、同学年集団との比較を「偏差値(平均50, 標準偏差10)」に換算すると、語彙理解力の平均は中学1年生の標準で77.0, 漢字読み取り能力は75.9という集団であったことが分かる。

なお、表3から分かるとおり、漢字読み取りテストについては、版独自の項目数に比べて、アンカー項目の数が相対的に少なかった。小学生と中

学生をつなぐアンカー項目は全て合わせて21項目であったが、結果的に不足していたかもしれない。そのためか、「漢字読み取り能力」については、小学校6年生までと中学1年生以上で能力パラメタ推定値のちらばりが極端に違っている(表5参照)。これは実質的な能力分布の反映というより、アンカー項目を用いて等化を行った本研究の技術的な限界であろう。この結果は本質的に好ましいことではないので、項目プールを実際を利用する上では特に注意が必要と思われる。具体的には、漢字読み取りテストの尺度を利用する場合、小学生用の項目を用いて測定した結果を中学・高校生集団の標準で表したり、その逆を行ったりすることはできるだけ控えることが必要だということである。また、小学校6年生と中学1年生の境界程度の能力の被験者集団については、小学生用、中学生用の冊子のいずれか一方のみに含まれる項目での能力の測定を行わずに、小学生用として作成された項目と中学生用として作成された項目を適宜混ぜ合わせた上で、テストを構成することが望ましい。

なお、表7, 表8は、語彙理解テスト, 漢字読み

表4 各版の項目パラメタの平均と標準偏差

版	語 彙		漢 字	
	識別力 a 平均 (sd)	困難度 b 平均 (sd)	識別力 a 平均 (sd)	困難度 b 平均 (sd)
小学校1年生	0.177 (0.073)	79.2 ( 5.2)	0.169 (0.068)	65.9 (9.2)
小学校2年生	0.176 (0.064)	81.1 ( 3.5)	0.262 (0.083)	79.3 (5.3)
小学校3年生	0.164 (0.049)	83.4 ( 2.7)	0.340 (0.104)	85.5 (4.1)
小学校4年生	0.150 (0.051)	88.0 ( 3.5)	0.382 (0.096)	89.8 (2.3)
小学校5年生	0.119 (0.044)	92.2 ( 4.8)	0.392 (0.097)	93.8 (2.4)
小学校6年生	0.088 (0.037)	97.2 ( 4.4)	0.339 (0.087)	96.9 (2.6)
中学生	0.073 (0.024)	103.6 ( 6.7)	0.278 (0.065)	104.7 (5.6)
高校生91年版	0.066 (0.024)	111.9 (10.1)	0.286 (0.070)	111.9 (4.6)
高校生92年A	0.068 (0.021)	120.0 ( 9.5)	0.293 (0.087)	116.2 (3.7)
高校生92年B	0.069 (0.021)	114.1 ( 6.8)	0.291 (0.082)	114.1 (3.5)
高校生92年C	0.090 (0.029)	104.5 ( 5.9)	0.302 (0.079)	111.6 (3.5)

表5 各学年、版ごとの被験者の能力推定値の平均と標準偏差

版	語彙 ( $\hat{\theta}_v$ )	漢字 ( $\hat{\theta}_k$ )
	平均 (sd)	平均 (sd)
小学校1年生	79.7 (4.2)	66.5 (13.1)
小学校2年生	82.7 (4.1)	83.5 (10.0)
小学校3年生	85.9 (3.7)	90.4 (11.4)
小学校4年生	89.0 (3.7)	94.2 (8.8)
小学校5年生	92.5 (4.8)	97.6 (8.0)
小学校6年生	96.4 (5.9)	100.9 (7.9)
中学1年生	98.4 (6.1)	102.9 (3.9)
中学2年生	102.2 (7.2)	105.6 (3.9)
中学3年生	105.9 (8.4)	108.1 (4.7)
高校1年生	113.7 (9.2)	113.0 (3.5)
高校2年生	117.4 (10.1)	114.4 (3.9)
高校3年生	120.4 (10.5)	115.6 (4.3)
中学生版	102.4 (8.0)	105.5 (4.7)
高校生91年版	117.1 (10.1)	114.4 (4.1)
高校生92年A版	123.2 (10.3)	116.3 (3.1)
高校生92年B版	116.9 (10.0)	114.5 (3.3)
高校生92年C版	112.0 (8.9)	112.7 (4.5)

取りテストの項目プールから一部(各50項目)を抜粋し、項目パラメタの推定値を示したものである<sup>2</sup>。

## 5 まとめと課題

本研究で作成された日本語基礎能力テストの項目プールは、通常の言語環境で日本語を第一言語として習得した者には、簡便で有効なテストとして十分機能すると考えられる。また、項目反応理論を用いているので、被験者の能力の程度や調査の目的に応じて適切な項目を抜粋し、適応形テストを作成することも可能である。今後、本研究で作成された項目プールが、様々な研究で有効に活用されることを期待するものである。

しかしながら、本研究で作成した項目プールが、更により利用しやすいものとなるためには、いくつかの課題が残されている。

まず、簡易得点化の問題である。本研究の項目プールは、項目反応理論に熟達した利用者にとっては、十分その利点を生かして便利に活用することが可能と思われる。確かに、日本でも項目反応理論に関する研究は非常に進んでおり、それに従事する研究者の数も少なくはない。しかし、現時点で一般に広く浸透しているとは言いがたい。したがって、現状のままでは利用者は測定の結果から能力推定値を計算し、それを適切に解釈できる者に限られてくると思われる。

そこで、より多くの研究者、教育者に利用されるためには、語彙テスト項目、漢字テスト項目ともに、一連の研究の中で利用されてきた11種類の版をそのまま活用し、簡易得点化するためのフォーマットを示すことが必要であると考えられる。確かに、目的に沿って自由にテストを構成できる、という本研究の項目プールの利点は制約されるが、簡易得点化によって、テスト理論やコンピュータに精通していない者にとっても、小野式

表6 平(1995)の再分析結果

	平均 (sd)	学年標準 (平均)	同一学年内偏差値
中学1年生語彙	114.9 (5.6)	高校1~2年程度	77.0
中学1年生漢字	113.0 (2.6)	高校1年生程度	75.9
高校3年生語彙	126.0 (9.1)	高校3年生程度以上	55.3
高校3年生漢字	118.7 (5.1)	高校3年生程度以上	57.2

表7 日本語語彙理解尺度項目プール (抜粋)

項目 (採用版)	a	b
1 つまむ (小1, 絵項目)	0.242	73.8
2 降参した (小2, 小3)	0.156	78.8
3 世話 (小3)	0.085	79.2
4 一斉に (小2, 小3)	0.186	79.9
5 すね (小1, 小2, 絵項目)	0.090	81.3
6 仮病 (小2, 小3)	0.212	83.4
7 つらい (小3, 小4)	0.103	84.4
8 懲りる (小3)	0.186	84.8
9 期待 (小4, 小5)	0.125	85.5
10 身辺 (小5)	0.101	86.3
11 一応 (小3, 小4)	0.199	87.0
12 空き (小4)	0.158	88.1
13 断る (小4)	0.100	88.4
14 誤解する (小5)	0.106	90.3
15 欠点 (小5, 小6)	0.192	92.6
16 肝心 (小5)	0.097	92.7
17 最期 (高B)	0.053	93.2
18 即座に (小5, 小6)	0.117	95.5
19 腕がなる (高C)	0.134	95.9
20 あえぐ (小6, 中学)	0.095	96.5
21 つまはじき (小6, 中学)	0.066	97.4
22 つつましい (小6)	0.118	97.6
23 たそがれ (小6)	0.042	98.0
24 不如意 (高C)	0.066	98.8
25 不平 (小5, 小6)	0.050	100.8
26 重視 (小6)	0.113	101.2
27 余儀ない (高C)	0.071	101.4
28 茶飯事 (高C)	0.126	103.0
29 匙を投げる (高C)	0.113	105.0
30 やましい (中学, 高91)	0.077	105.1
31 助言 (中学, 高91)	0.099	105.6
32 待遇 (中学)	0.070	106.1
33 皆無 (高91)	0.052	106.2
34 ~のきらいがある (高C)	0.087	107.2
35 へつらう (小6, 中学)	0.028	109.6
36 手だれ (高A)	0.059	111.8
37 青菜に塩 (高B)	0.060	111.8
38 鶴の一声 (高B)	0.076	112.1
39 冒瀆 (高91)	0.070	112.5
40 贅を尽くす (高校生共通)	0.080	114.3
41 饒舌 (高91)	0.059	115.1
42 ~にまがう (高A)	0.058	115.6
43 堪能 (高B)	0.083	116.4
44 矢の催促 (高B)	0.067	118.0
45 不逞 (高A)	0.057	118.9
46 辛辣 (高A)	0.072	121.0
47 無常 (高91)	0.054	122.5
48 下馬評 (高B)	0.038	122.8
49 人身御供 (高A)	0.047	124.6
50 実るほど頭を垂れる稲穂かな (高A)	0.037	138.9

表8 漢字読み取り尺度項目プール (抜粋)

項目 (採用版)	a	b
1 赤ちゃん (小1)	0.163	55.3
2 青い (小1)	0.169	62.2
3 立つ (小1)	0.134	66.1
4 冬 (小2)	0.283	75.5
5 楽しく (小2)	0.251	77.2
6 新しい (小2, 小3)	0.177	79.3
7 用 (小2)	0.349	81.7
8 集まれ (小1~小3)	0.214	84.4
9 安全 (小3)	0.476	85.1
10 港 (小3)	0.260	86.6
11 薬局 (小3, 小4)	0.258	87.8
12 満月 (小4)	0.476	88.7
13 事故 (小3~小5)	0.466	89.5
14 堂々 (小4, 小5)	0.565	90.3
15 害虫 (小4, 小5)	0.385	91.1
16 報告 (小5)	0.466	92.5
17 情熱 (小5)	0.469	93.3
18 天候 (小4)	0.277	93.7
19 高価 (小5)	0.446	94.6
20 展覧会 (小6)	0.356	95.8
21 異常 (小6)	0.433	96.6
22 余って (小5, 小6)	0.295	97.1
23 撮影 (小6, 中学)	0.326	97.8
24 刃物 (中学)	0.269	98.8
25 誤り (小6)	0.325	99.5
26 批評 (小6)	0.281	100.5
27 丹念 (中学)	0.277	102.5
28 錠剤 (中学, 高91)	0.295	103.6
29 主催 (中学)	0.327	104.0
30 窒息 (高C)	0.235	104.7
31 復讐 (高91)	0.334	107.8
32 巨匠 (中学)	0.304	107.9
33 中枢 (高C)	0.183	107.9
34 語弊 (高B)	0.274	108.8
35 優劣 (中学)	0.234	109.0
36 会釈 (高C)	0.335	109.8
37 喝采 (高91)	0.391	110.0
38 貴賓 (中学, 高91)	0.244	111.2
39 発心 (高A)	0.212	111.4
40 償う (高C)	0.314	111.7
41 不朽 (中学, 高91)	0.267	112.5
42 癒す (高91)	0.243	112.6
43 野放図 (高B)	0.126	113.0
44 強靱 (高B)	0.381	113.4
45 団塊 (高A)	0.191	114.7
46 培う (高B)	0.240	116.4
47 概ね (高A)	0.289	116.8
48 髣髴 (高A)	0.250	117.1
49 躓く (高91)	0.230	118.7
50 軋轢 (高A)	0.367	119.9

日本語基礎能力テストが格段に利用しやすくなると思われるからである。

次の課題は調査対象者の拡大である。本章の冒頭でも述べたとおり、本研究の日本語基礎能力テストの尺度がそのまま適用可能なのは、あくまでも通常の言語環境で日本語を第一言語として習得した者である。第二言語として日本語を習得した者や特殊な言語環境で日本語を習得した者に対しても適応可能であるかどうかを診断するために、本研究で得られた項目パラメタを所与として、本研究のテストを一般項目反応理論 (GIRT: Generalized Item Response Theory) (孫・芝, 1990; 前川他, 1993) 等のモデルに拡張したり, Rule Space Methodology (Tatsuoka, 1995) を適用してその認知的特徴を検討する考慮する必要があると思われる。

最後に、コンピュータを用いた適応形テストの開発が課題として残っている。特に、漢字読み取りテストは回答を自筆で記入する形式であるため、技術的に工夫すべき余地が残されていると思われる。

以上のような課題を逐次解決していくことにより、今後、項目プールをより使い易くしていくことがこれからの課題である。

## 注

1. 本文中の「3 等化の方法」および英文アブストラクトの文責は前川に、それ以外の部分の文責は平にある。
2. 本研究の方法論的な制約により、項目プール全体の公開は残念ながら差し控えざるを得ない。本研究で作成したテストの利用を希望される方は著者宛に御連絡をいただきたい。

## 謝 辞

小中学生用項目の作成・分析には、小野・林部以外に繁樹算男、岡崎勉、市川雅教、木下ひさし、牧野泰美の各氏が貢献した。高校生用問題の作成には、平・小野・林部以外に米山千佳子氏が貢献した。これらの方々には心から御礼を申し上げる。

## 文 献

東 洋・梅本堯夫・芝 祐順・梶田叡一 (1988). 現代教育評価事典, 金子書房  
Hambleton, R.K., & Swaminathan, H. (1985).

*Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.

服部 環 (1990). 個人差に応じたテスト方式による語彙理解力の測定, 教育心理学研究, 38, pp. 445-454.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Welsey.

前川眞一 (1991). パラメタの推定 芝 祐順 (編), 項目反応理論 基礎と応用, 東京大学出版会 pp. 87-129.

前川眞一・平 直樹・小野 博・林部英雄・米山千佳子 (1993). 帰国子女の日本語能力測定のための高校生用漢字・語彙テストの開発 (5), 一般項目反応理論 (GIRT) を用いた測定の方法論一, 日本教育心理学会第 35 回総会発表論文集, p. 483.

Mislevy, R.J. & Bock, R.D. (1989). *PC-BILOG3: Item analysis and test scoring with binary logistic models*. Mooresville: Scientific Software.

野口裕之・芝 祐順・丹 直利 (1983). 語彙理解尺度の研究 II, 一項目固定版と適応形テストによる測定— 東京学芸大学紀要 第 1 部門, 34, pp. 101-114.

大友賢二 (1996). 項目応答理論入門 言語テスト・データの新しい分析法, 大修館書店

小野 博・繁樹算男・林部英雄・岡崎 勉・市川雅教・木下ひさし・牧野泰美 (1989). 日本語力検査の開発, 昭和 61-63 年度科学研究費報告書, 試験研究 (1) 課題番号 6181003 (未公刊).

小野 博・平 直樹・林部英雄・米山千佳子・前川眞一 (1993). 帰国子女の日本語能力測定のための高校生用漢字・語彙テストの開発 (3) 一項目プールの完成—, 日本教育心理学会第 35 回総会発表論文集, p. 481.

芝 祐順 (1978). 語彙理解尺度作成の試み, 東京大学教育学部紀要, 17, pp. 47-58.

孫 媛・芝 祐順 (1990). 特異な反応パターンを示す被験者の能力測定, 一般項目反応理論の適用一, 教育心理学研究, 38, pp. 360-368.

平 直樹 (1995). 物語作成課題に基づく作文能力評価の分析, 教育心理学研究, 43, pp. 134-144.

- 平 直樹・武井澄江・荻野美佐子 (1989). 適応形質問紙形式の乳幼児用言語理解尺度への応用, 教育心理学研究, 37, pp. 392-399.
- 平 直樹・小野 博・林部英雄 (1992). 高校生用日本語語彙理解力テストの開発 —(1) 試作問題の精選—, 大学入試センター研究紀要, 21, pp. 107-143.
- 平 直樹・小野 博・林部英雄 (1993a). 高校生用日本語能力テストの開発, —(2) 漢字読み取りテストの項目分析と誤答パターンについて—, 大学入試センター研究紀要, 22, pp. 53-85.
- 平 直樹・前川眞一・小野 博・林部英雄・米山千佳子 (1993b). 帰国子女の日本語能力測定のための高校生用漢字・語彙テストの開発 (4) —項目反応理論 (IRT) による尺度化と等化—, 日本教育心理学会第35回総会発表論文集, p. 482.
- 平 直樹・前川眞一・小野 博・林部英雄・米山千佳子 (1995). 高校生程度の日本語能力テストの開発, 一語彙理解テスト・漢字読み取りテストの尺度化—, 教育心理学研究, 43, pp. 68-73.
- Tatsuoka, K.K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach, In P.D. Nichols, S.F. Chipman, & R.L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 327-359). Hillsdale, NJ: Lawrence Erlbawn Associates.

# Development of the Item Pool for the Broad-Range Japanese Fundamental Language Skills Test

Naoki TAIRA\*  
Shin-ichi MAYEKAWA\*\*  
Hiroshi ONO\*\*\*  
Hideo HAYASHIBE\*\*\*\*  
Teruhisa UCHIDA\*\*\*

## Abstract

A Broad-range Japanese Fundamental Language Skills Test designed to evaluate student's comprehension of basic Japanese words was developed. The present test consists of two subtests: the Basic Vocabulary Test (Vocabulary Test) consisting of 471 basic Japanese word multiple-choice items, and the Basic Chinese Character Pronunciation Test (Kanji Test) consisting of 977 free-response items. Both subtests covers the age-group of 6 through 18.

For each subtest, 11 preliminary forms were constructed and administered to the students from the first grade of elementary schools to the third grade of senior high school. Item parameters were separately calibrated for each form using approximately 3,000 students and then equated to form a common scale. On the common scale of the item pool, for the Vocabulary Test, the mean ability score of each grade ranges from 80 to 120, and for the Kanji Test, 67 to 116. The standard deviation of each grade is about 4 in the lower and about 10 for the higher for the Vocabulary Test, and for the Kanji test, about 13 in the lower and about 4 in the higher grade.

**Key words:** item pool, test equating, Japanese language skills test, vocabulary test, Chinese character test

---

\* Admission Method Section, Research Division, The National Center for University Entrance Examinations.  
2-19-23 Komaba, Meguro-ku, Tokyo 153-8501 Japan. e-mail: taira@rd.dnc.ac.jp Telefax: 03-5478-1297  
\*\* Scholastic Aptitude Section, Research Division, The National Center for University Entrance Examinations.  
\*\*\* Special Examination Section, Research Division, The National Center for University Entrance Examinations.  
\*\*\*\* Faculty of Education and Human Sciences, Yokohama National University.

## 大学入試センター研究紀要 No. 28 の訂正について

倉元直樹（東北大学 アドミッションセンター）\*  
前川眞一（大学入試センター研究開発部 進学適性研究部門）  
小野博（大学入試センター研究開発部 特別試験研究部門）  
林部英雄（横浜国立大学 教育人間科学部）  
内田照久（大学入試センター研究開発部 特別試験研究部門）

大学入試センター研究紀要 No. 28 「日本語基礎能力テストの項目プールの完成」 8ページ「表5」の漢字テストの数値に誤りがありましたので、訂正いたします。また、8ページ「表6」の「同一学年内偏差値」についても、一部、誤った値に基づいて計算が行われていたため、訂正いたします。数値を訂正した箇所については、表中に下線を引いて表示いたしました。

さらに、1ページ「要約」の第2パラグラフ4～5行目、6行目を以下のように訂正いたします。

〔1ページ第2パラグラフ4～5行目〕

（誤）・・・漢字テストでは67から・・・

（正）・・・漢字テストでは66から・・・

〔1ページ第2パラグラフ6行目〕

（誤）・・・漢字テストの場合は、下の学年で約13、上の学年で約4であった。

（正）・・・漢字テストの場合は、小学校1年生の約10を例外として、3～5程度となった。

なお、「表5」の訂正により、7ページ左段下から3行目～右段下から2行目のパラグラフの記述に意味がなくなりましたので、全文削除いたします。

\*：大学入試センター紀要 No. 28 発表当時、「大学入試センター 試験方法研究部門」所属。旧姓「平」。

表5 各学年、版ごとの被験者の能力推定値の平均と標準偏差（訂正）

版	語彙 ( $\theta_v$ )	漢字 ( $\theta_k$ )
	平均 (s d)	平均 (s d)
小学校1年生	79.7 (4.2)	<u>65.8</u> (9.9)
小学校2年生	82.7 (4.1)	<u>82.6</u> (4.7)
小学校3年生	85.9 (3.7)	<u>89.0</u> (3.5)
小学校4年生	89.0 (3.7)	<u>92.9</u> (3.3)
小学校5年生	92.5 (4.8)	<u>96.5</u> (3.3)
小学校6年生	96.4 (5.9)	<u>100.0</u> (3.8)
中学1年生	98.4 (6.1)	102.9 (3.9)
中学2年生	102.2 (7.2)	105.6 (3.9)
中学3年生	105.9 (8.4)	<u>108.0</u> (4.2)
高校1年生	113.7 (9.2)	113.0 (3.5)
高校2年生	117.4 (10.1)	<u>114.3</u> (3.4)
高校3年生	120.4 (10.5)	115.6 (3.5)
中学生版	102.4 (8.0)	105.5 (4.5)
高校生91年版	117.1 (10.1)	<u>114.3</u> (3.6)
高校生92年A版	123.2 (10.3)	116.3 (3.1)
高校生92年B版	116.9 (10.0)	114.5 (3.3)
高校生92年C版	112.0 (8.9)	<u>112.6</u> (3.6)

表6 平 (1995) の再分析結果 (訂正)

	平均 (s d)	学年標準(階)	同一年内偏差値
中学1年生語彙	114.9 (5.6)	高校1~2年生程度	77.0
中学1年生漢字	113.0 (2.6)	高校1年生程度	75.9
高校3年生語彙	126.0 (9.1)	高校3年生レベル	55.3
高校3年生漢字	118.7 (5.1)	高校3年生レベル	58.9