

## 日本語基礎能力テストの特性 (4) — 項目反応理論に基づく漢字読み取りテストの測定性能の分析 —

倉元 直樹, \* 佐藤 洋之, \*\* 伊藤 博美 \*\*\*

\* 東北大学アドミッションセンター・東北大学大学院教育情報学教育部

\*\*\* 東北大学大学院教育情報学教育部

\*\* 秋田県立本荘高等学校

**要旨:** 本研究の目的は、倉元・伊藤・佐藤 (2004) で妥当性を確認した日本語基礎能力テスト (平他、1998) の漢字読み取り能力テストの項目について、項目反応理論に基づき、その測定性能を分析することである。倉元他 (2004) による分析の結果、項目作成から10年余りを経た現在でも、全体として十分機能することが示された。本研究では、項目プール全体974項目を分析の対象とする。

テスト情報曲線から、項目プールは全体としては測定性能のピークを2つ持つ、双峰形の特徴を持つことが分かった。元々作成された11の版のうち、小中学生用の各版では、小学1年生版を除き、それぞれ測定対象とする集団の平均よりも尺度値にして2.5~3.5程度下に最大の測定性能を有することが分かった。また、高校生用の各版は、それぞれ狙いとする範囲に高い識別性能を持つことが分かった。

以上、本研究の結果からは、統計的観点においても尺度が十分機能することが確認された。

**キーワード:** 項目反応理論、テスト情報量、項目プール、漢字、ロジスティック・モデル

### 1. 目的

日本語基礎能力テスト (平・前川・小野・林部・内田、1998) は、項目プールが完成してから6年、最後の項目が作成されてから12年という歳月が経過している。項目の実用性を内容的に再検討する必要性が迫られていた。日本語語彙理解力テスト (以後、「語彙テスト」と略記する) 部分については、伊藤・佐藤・倉元 (2003)、漢字読み取り能力テスト (以後、「漢字テスト」と略記する) 部分のうちの中学生用、高校生用項目については、倉元・伊藤・佐藤 (2004) によって国語教育的見地から再評価が行われ、ほとんどの項目は現在でも問題なく使用できることが確認されている。また、語彙テスト項目に関しては、佐藤・伊藤・倉元 (2003) によって、その測定性能に問題がないことが確認されている。

本研究では、それらの分析結果を元に、漢字テストの項目プール<sup>1</sup>、および、冊子形式となっている11の版の測定論的性質について分析を行い、その測定性能を再吟味することを目的とする。

日本語基礎能力テストはいずれも項目反応理論に

基づき、2パラメタ・ロジスティック・モデルによって尺度化されている。佐藤他 (2003) では、その特長を生かし、情報関数に基づいて語彙テスト項目の測定性能の分析を行っている。本研究でも同様の観点から分析を行う。

なお、佐藤他 (2003) では、元々日本語基礎能力テストの項目が作成された際の11種類の版について測定性能を分析している。項目反応理論の利点は項目が異なっても同一尺度上で特性値を表すことが可能な点にある。したがって、版毎の分析にこだわる必要はないが、日本語基礎能力と作成の経緯、専門知識を持たない一般のユーザーに対する利便性などを考え、佐藤他 (2003) と同様に版毎の測定性能の特徴についても分析を加えることとする。

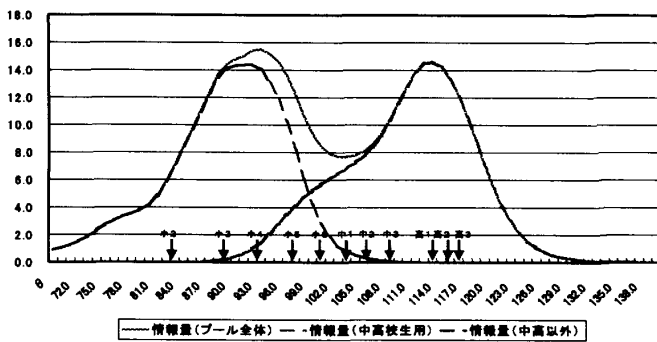
### 2. 項目プールの測定性能

平他 (1998) によると、漢字テストの項目プールは、語彙テストと同様に標準化を目的とした線形変換が施されており、被験者の能力値がほぼ平均100、標準偏差15となっている。

倉元他 (2004) では、平他 (1998) の日本語基礎能力の漢字テストの項目プールのうち、中学生・高校生用の517項目について国語学的見地からの再検討が行われた。その結果、全ての項目が「レベルA」、および、「レベルB」の評価を受けた。すなわち、利用に当たって特に注意を要する項目は存在しない。したがって、本研究では、小学生用も含む974項目の項目プール全体について、その測定性能を論じることとする。

漢字テストの項目プールについて情報関数を計算したところ、結果は図1に示す通りであった。

図1. 漢字テスト項目プールのテスト情報量



図中の実線は項目プール全体の情報関数である。また、点線は中学生用・高校生用項目とそれ以外<sup>2</sup>の項目とを分けて情報関数を描いたものである。さらに、矢印で示したのは項目プール作成時の被験者集団の学年別平均値<sup>3</sup> (平他、1998) である。項目プール全体として、 $\theta = 84$ 付近に情報量のピークがあった語彙テスト (佐藤他、2003) とは異なり、漢字テストの情報関数は双峰形の形状となっている。小学3年生から小学5年生程度、中学3年生から高校生程度の範囲は精度良く能力を識別できるが、小学6年生から中学2年生程度の範囲で精度がやや劣っている。なお、中学3年生の平均値と高校生の平均値が相対的に大きく開いているのは、項目作成時に調査対象となった高校生の被験者の中に、進学校の生徒が多かったことに由来するのではないと思われる。

点線で示した2つのグラフを見ると、前者にアンカー項目を含めたことにも由来するが、中・高校生用の項目が下の学年の方まで比較的高い精度を保っているのに対し、小学生程度の問題のうち、高学年用の問題の情報量がやや足りないことが分かる。いずれにせよ、今後、項目プールを拡充する際には、

小学校高学年から中学生程度の問題を重点的に補強する必要があると思われる。

また、小学1年生程度の能力に関しては、この項目プールでは十分に識別できていないことも分かった。小学校入学直後は多くの子どもたちにとって漢字を覚え始める時期でもあり、テストとして用いることができる文字は限られている。精度の良い測定を行うための項目作成には限界があると思われる。漢字テストの場合、基本的には小学2年生程度以上の能力を持つ被験者を測定対象と考えた方が無難であろう。

### 3. 版別分析結果

#### 3.1. 小学1年生版～中学生版

図2.1～2.7は各版ごとに調査対象として想定されている学年の被験者の漢字読み取り能力 $\theta$ の推定値

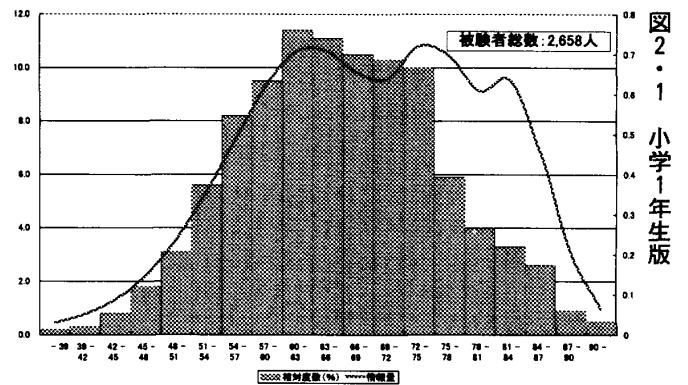


図2・1 小学1年生版

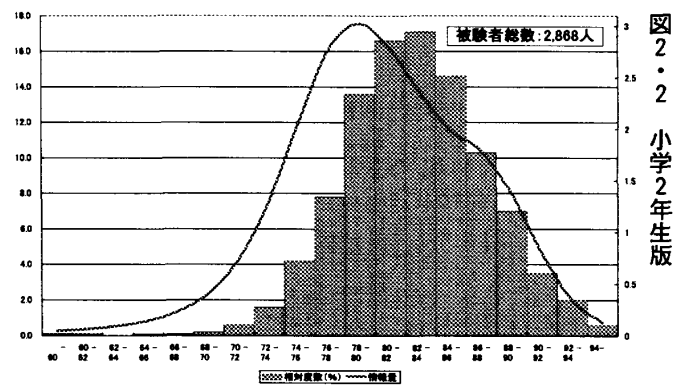


図2・2 小学2年生版

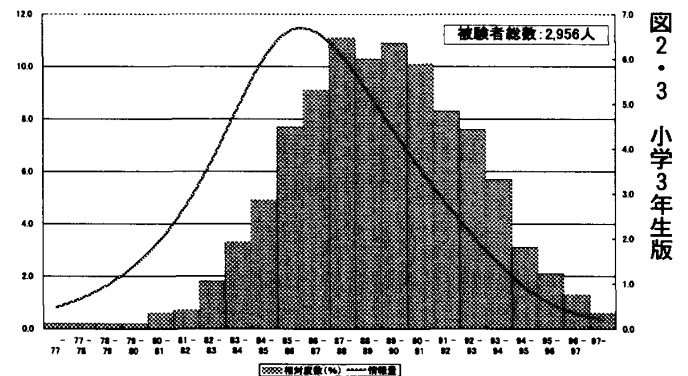


図2・3 小学3年生版

の分布とそれぞれの版の情報関数を重ね合わせて図示したものである。なお、先述の通り、それぞれの被験者は項目プール作成時に調査対象となった者である。

小学1年生の被験者の能力分布は、ピークの幅が広く、他の学年とは違った特異な形状をしている。能力推定値のモードは $\theta = 61-62$ 付近であるが、 $\theta =$

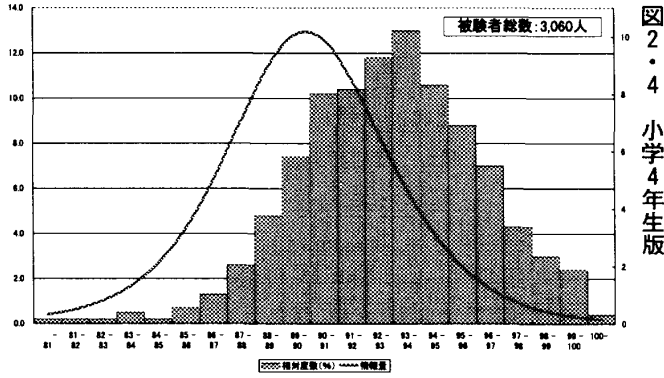


図2・4 小学4年生版

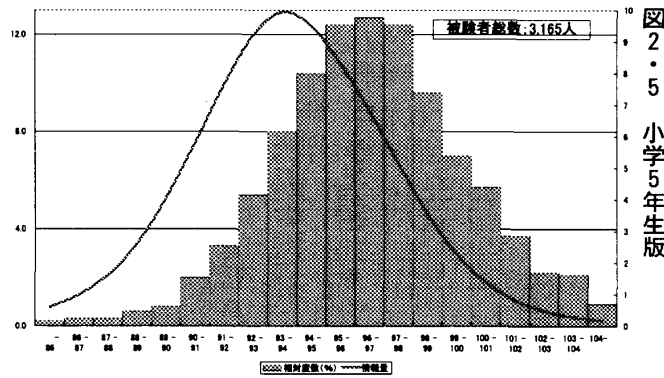


図2・5 小学5年生版

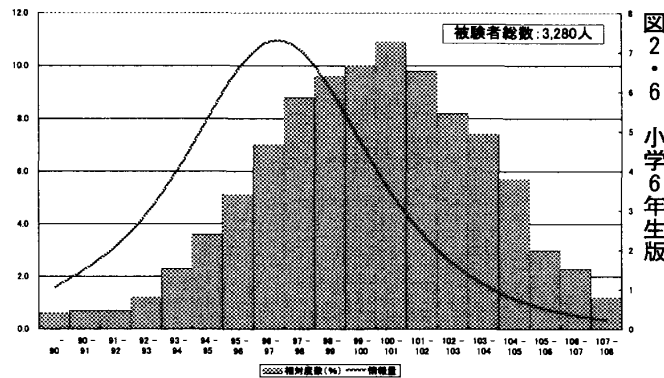


図2・6 小学6年生版

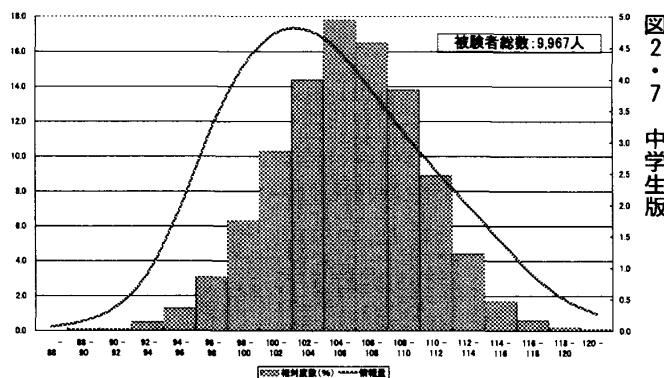


図2・7 中学生版

60~75付近に多くの被験者の能力推定値が分布しており、平均値は $\theta = 65.8$ となっている。一方、情報関数も多峰形を示しており、 $\theta = 60\sim 85$ 付近の広い範囲で比較的高い測定精度を示している。項目プール全体としての識別性能は十分とは言えないにしても、小学1年生版の項目においては被験者の能力分布に見合った範囲の測定精度が確保されている。小学1年生程度の能力は、項目プール全体としては十分な測定精度を保った測定は難しいが、小学1年生用の版を用いれば、ある程度の精度での測定は可能である。小学1年生版もそれ自体としては、被験者の能力に見合った項目から構成されていると言えるのである。

小学2年生以上の漢字読み取り能力の分布は、比較的美しい単峰形の分布となっている学年が多い。しかし、いずれも、項目の測定性能のピークは被験者の能力分布に対してやや下方に位置しており、全体として想定されていた学年集団にとっては多少易しめであったことがうかがえる。具体的には、小学2年生版では能力分布の平均値が $\theta = 82.6$ に対し、情報関数のピークが $\theta = 79$ 付近、小学3年生版では能力分布の平均値が $\theta = 89.0$ に対し、情報関数のピークが $\theta = 86$ 付近、小学4年生版では能力分布の平均値が $\theta = 92.9$ に対し、情報関数のピークが $\theta = 90$ 付近、小学5年生版では能力分布の平均値が $\theta = 96.5$ に対し、情報関数のピークが $\theta = 93.4$ 付近、小学6年生版では能力分布の平均値が $\theta = 100.0$ に対し、情報関数のピークが $\theta = 97$ 付近と、各学年尺度値にしておよそ3程度のズレがある。ただし、測定性能が高い範囲と被験者の能力分布が非常に大きく離れているわけではなく、実用性に問題はない。また、実際に項目プールを利用する場合には、オリジナルの版をそのまま利用する必要はないので、これらの結果を目安として、適切な項目を選んでテストを構成すればよい。

一方、中学生版に関して、能力分布の平均値が $\theta = 105.5$ に対して情報関数のピークが103付近となっており、やはり、最大の測定性能を持つ領域が被験者の能力分布のピークの下に位置している。この点では、中学生版も小学生用の各版と違いはさほど大きくはない。

語彙テストと比較すると、漢字テストの項目プールは大きく、個々の項目の識別力(項目パラメタ $a$

の値)も大きい傾向がある。したがって、全般に渡って測定性能は基本的に高いと言える。しかしながら、小学校から中学校程度の能力の範囲においては、元々ターゲットとしていた学年の能力に合わせた測定という意味ではやや易しい方向に偏る傾向があり、語彙テスト項目の各版ほどには被験者の能力分布と一致していないことが分かった。

### 3.2. 高校生91年度版、92年A～C版

図3は、前述の図2.1～2.7同様に、高校生版として作成された漢字テストの被験者の能力 $\theta$ の推定値の分布、および、版ごとの情報関数を示したものである。ヒストグラム(たて棒グラフ)は能力推定値の分布、実線は91年度版、破線は92年A版、点線は92年B版、一点鎖線は92年C版の情報関数を表している。

高校生用の4つの版は、全て高校生程度の漢字読み取り能力の測定を意図して作成されたものであるが、91年度版は全般的に幅広い範囲を網羅し、92年A版は特に能力の高い被験者、92年C版は高校生としては能力の低い被験者、92年B版はその中間の能力の被験者に精度の高い測定を目的として作成されたものである。

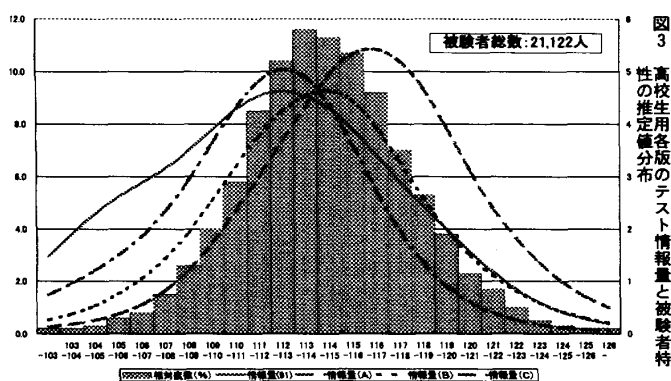


図3に示したように、2万人を超える被験者の漢字読み取り能力推定値のモードは $\theta = 114$ 付近、平均値は114.4である。

それぞれの版の情報関数のピークは91年版が $\theta = 112-113$ 付近、92年A版が $\theta = 116-117$ 付近、92年B版が $\theta = 114-115$ 付近、92年C版が $\theta = 112-113$ 付近であった。各版とも、中学生版よりかなり難しいことが共通であると同時に、それぞれ版の特徴がある。

まず、91年度版については、情報量のピークは92年C版とほぼ同じところに位置しているが、上下に

幅広い範囲の精度を確保している。項目数が139項目と、他の版(103～122項目)より多いので全体として情報量が大きいののは必然だが、それが一点に集中するのではなく、幅広い範囲の能力の精度の良い推定という目的につながっているという意味では、作成の趣旨に適っている。また、被験者の能力推定値の分布と重ね合わせて比較したときに、B版はそれとほぼ重なる形状の情報関数となっており、A版が高い能力範囲、C版が低い能力範囲をカバーしていることから、ほぼ狙い通りの推定精度が確保されていると言えよう。

### 4. まとめと課題

本研究の分析から、日本語基礎能力テストの低位尺度である漢字読み取り能力テストが、現時点で実用的にはほぼ問題なく利用できることが分かった。今後の課題としては項目プールの更なる拡充が挙げられる。特に、小学校高学年程度から中学生程度をカバーする項目、漢字を覚え始める小学1年生程度の項目、高校段階を超えるような項目の尺度化が成され、項目プールに加えられるならば、さらに利用価値が高まると考えられる。

日本語基礎能力テストのような基礎的な尺度は様々な目的で利用されてこそ、その持てる力を発揮することができる。各種、被験者の能力の個人差に関する評価が必要となる研究において、能力の1つの要因としての測定、漢字の読み取り能力そのものの研究、あるいは、項目反応理論の理論的研究の素材など、今後も様々な分野で広く活用されることが期待される。

#### 注釈

- 倉元他(2004)の分析においては、中学生用、高校生用の517項目のみが対象であったが、本研究の分析には小学生用項目も含む項目プール全体974項目である。
- 小学生用項目から中学生版、高校生用の各版とのアンカー項目を除いたもの。
- 大学入試センター研究紀要No.30に掲載された訂正表の数値を用いた。小学1年生の平均値( $\theta = 65.8$ )はグラフの範囲外である。

文献

伊藤博美・佐藤洋之・倉元直樹 (2003). 日本語基礎能力テストの特性 (1) -国語教育から見た語彙理解力測定項目の内容評価-, 教育情報学研究 (東北大学大学院教育情報学研究部・教育部紀要)、第1号、15-24.

倉元直樹・伊藤博美・佐藤洋之 (2004). 日本語基礎能力テストの特性 (3) -国語教育から見た漢字読み取り能力測定項目の内容評価-, 教育情報学研究 (東北大学大学院教育情報学研究部・

教育部紀要)、第2号 (7-15).

佐藤洋之・伊藤博美・倉元直樹 (2003). 日本語基礎能力テストの特性 (2) -項目反応理論に基づく測定性能の分析-, 教育情報学研究 (東北大学大学院教育情報学研究部・教育部紀要)、第1号、25-30.

平直樹・前川眞一・小野博・林部英雄・内田照久 (1998). 日本語基礎能力テストの項目プールの作成、大学入試センター研究紀要、No.28、1-12.

## Properties of the Broad-Range Japanese Fundamental Language Skills Test IV: Statistical Analysis of Chinese Character Pronunciation Subscale based on Item Response Theory

Naoki T. Kuramoto , \* Hiroyuki Sato , \*\* Hiromi Ito \*\*\*

\* Admission Research Center of Tohoku University; Education Division,  
Tohoku University Graduate School of Educational Informatics

\*\* Education Division, Tohoku University Graduate School of Educational Informatics

\*\*\* Honjo Akita Prefectural High School

The purpose of this study is to reexamine the statistical properties of Chinese Character Pronunciation items, which consists of the subscale for the Broad-Range Japanese Fundamental Language Skills Test (BRJFLST) developed by Taira et al. (1998). Kuramoto, Ito, and Sato (2004) showed the validity of the item pool, even after 10 years since item construction.

According to the test information curve, the item pool showed dual-modal shape that means it has two peaks of the most discriminating range on the theta scale. Each of the original 7 testlets, designed for elementary and junior high school level showed the peak of measurement efficiency at 2.5 to 3.5 under the average level of its target population, except for the form grade 1. Each of the 4 versions for senior high school student level showed highly discriminating range just as designed for.

The present study showed that the scale was also effective from statistical viewpoints.

**Key words:** Item Response Theory, test information, item pool, Chinese Character, logistic model