

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—

倉元 直樹, * 伊藤 博美, ** 佐藤 洋之 ***

* 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

** 秋田県立本荘高等学校

*** 東北大学大学院教育情報学教育部

要旨: 本研究の目的は、日本語基礎能力テスト (平他、1998) の漢字読み取りテスト項目について国語教育の観点からの妥当性を吟味することである。本研究では、「漢字読み取り能力」下位尺度の項目プールの中から、中学生、高校生程度の測定を目的として作成された517項目について分析の対象とした。

まず、項目内容を分類する基準として、「レベルA~C」を設定した。「レベルA」は「通時的に意味や用法が安定しているもの」であり、「レベルB」は「レベルAに準じ、かつ、いくつかの観点で特徴が見られるもの」である。「レベルC」は「使用機会がない、ないしは、特殊な分野のみで使われるもの」である。さらに、各レベルに2~5の下位カテゴリーを設けた。

結果的に全体の約85%が「レベルA」に分類され、「レベルC」は皆無であった。現在でも、全体として、十分に漢字読み取り能力の測定に通用する尺度であることが確認された。

キーワード: 国語教育、高校生、項目プール、漢字、測定

1. 目的

本研究の目的は、平・前川・小野・林部・内田(1998)で作成された日本語基礎能力テスト項目プールのうち、漢字読み取り能力テスト (以後、「漢字テスト」と略記する) の各項目のうち、基本的に高校生レベルの能力測定を想定して作成された項目について、主に国語教育の観点から再吟味を行うものである。

日本語基礎能力テストの項目プールは、日本語語彙理解力テスト (以後、「語彙テスト」と略記する) 468項目、漢字テスト974項目から成る¹。下位尺度ごとに、項目反応理論²に基づき2パラメタ・ロジスティック・モデルで尺度化されている (佐藤・伊藤・倉元、2003a) のが項目プールの特徴である。測定対象は、語彙テストと同様に、小学校1年生程度から高校生程度までと幅広い。

漢字テストの項目は、幹に示された短文に埋め込まれた単語や熟語の読みを答える形式である。項目は、傍線で示される。なお、一部に1つの短文に複数の項目が問題として埋め込まれているケースがある。解答者は、漢字の読みの部分だけを指定された

解答欄にひらがなで書き込むことになる。「送りがな」が必要な漢字についても、その振り方は問題文の中で明示されているので、基本的には測定 (採点) の対象とはならない。元々、項目プールが作成された際には11種類の版 (小学校1年生版~6年生版、中学生版、高校生91年版、高校生92年A版~C版) に分けて実施されていた。アンカー項目として複数の版に採録されている項目もある。各版には、その後の項目分析の結果項目プールから削除された項目も含まれている (以上、伊藤・佐藤・倉元、2003)。

日本語基礎能力テストは、まず、小・中学生用の項目が小野・繁樹・林部・岡崎・市川・木下・牧野 (1989) によって開発され、その後、語彙テストと漢字テストの部分に高校生用の項目が付け加えられた (平・小野・林部、1992; 平・小野・林部、1993)。さらに、高校生用項目の拡充が行われた上で高校生用に関する項目プールが作成された (平・小野・前川・林部・米山、1995)。最終的に、小・中学生用項目との等化が行われ (平他、1998)、現在に至っている。

項目反応理論によって尺度化されているため、日本語基礎能力テストには、項目プールから被験者の能力に合わせて自由に冊子 (form) を編纂することができるという利点がある。したがって、応用範囲も広い。現在までに、作文能力を純粹に抽出して評価する目的で、基礎的日本語能力を統制するための測定 (平、1995)、新しく開発された総合試験問題で測定されている能力の要因分析の一部 (倉元、2002) といった形で日本語基礎能力以外の能力に寄与する統制要因としての日本語基礎能力の測定や、高校生段階の日本語基礎能力そのものの経年変化の測定 (佐藤・伊藤・倉元、2003b)、さらに、語彙テストの一部の項目については、項目反応理論をベースにした学習診断システム (Rule Space Methodology: ルールスペース法) への応用 (倉元・龍岡、2001; 倉元・スコット・笠居、2003) といった形で活用されてきた。

しかしながら、漢字テストにおいては、特に中・高校生用として作成された517項目について、国語教育的に項目内容を再吟味し、項目プールの内容的妥当性に裏づけを与えることが必要である。項目が作成されてから、現時点 (2004年) までに少なくとも10年以上の歳月が経過していること、項目作成の過程において、専門家 (高校の国語教師等) の関与があったものの、必ずしも国語教育の観点からの吟味が完全ではなかったことがその理由である。なお、語彙テストについては小学生用から高校生用まで、既に国語教育的観点からの項目内容の吟味を終えており、実用性の面では十分満足できる結果が得られている (伊藤他、2003)。

本稿では、最初に中等教育、特に高校の国語教育のうち、漢字の指導の現状について簡単にまとめる。その後、それぞれの項目の評価を行うこととする。

2. 国語教育の現状と漢字指導の位置づけ

学校教育において基礎的な言語能力の育成を多く負っている教科は国語であるが、現在移行期にある新学習指導要領において小学校国語、中学校国語、高等学校国語の教科目標はそれぞれ以下のようになっている。

(1) 小学校国語

国語を適切に表現し正確に理解する能力を育成

し、伝え合う力を高めるとともに、思考力や想像力及び言語感覚を養い、国語に対する関心を深め国語を尊重する態度を育てる。

(2) 中学校国語

国語を適切に表現し正確に理解する能力を育成し、伝え合う力を高めるとともに、思考力や想像力を養い言語感覚を豊かにし、国語に対する認識を深め国語を尊重する態度を育てる。

(3) 高等学校国語

国語を適切に表現し的確に理解する能力を育成し、伝え合う力を高めるとともに、思考力を伸ばし心情を豊かにし、言語感覚を磨き、言語文化に対する関心を深め、国語を尊重してその向上を図る態度を育てる。

教科目標は、中学校は小学校、高等学校は小・中学校のものをそれぞれ受けたものである。また、新指導要領では旧要領と比較して多くの点の変更が行われている。

例えば高等学校においては、「国語を適切に表現し的確に理解する能力を育成」という従前のものに加え、「伝え合う力」を高めることが位置づけられている。「伝え合う力」とは端的に言えば、「円滑に相互伝達、相互理解を進めていく能力」のことであるとされるが、これは新指導要領が社会人として必要とされる言語能力の育成を重視していることと密接に関連している (伊藤他、2003)。

また、我が国の言語文化はその初期において漢字に負っている部分が大きいという国語史的事実をふまえると「言語文化に対する関心を深め」という表現からは、漢字文化の重要性を主張していることも少なからずうかがえる。「高等学校学習指導要領解説」では、話し言葉、書き言葉それぞれの歴史的事実に基づいた様々な言語形態に関する広くかつ深く関心を求めているからである (文部省、1999)。

必履修科目としての「国語表現Ⅰ」では、内容の取り扱い上の配慮事項として、「常用漢字の読みに慣れ、主な常用漢字が書けるようにする」ことを求めている。また、「国語総合」では、「言語事項」の一項目に「常用漢字の読みに慣れ、主な常用漢字が書けるようになること。」と設定されている。そこでは中学校で学習済みの漢字に関する音訓の習熟も意識しており、漢字が担う語彙的側面の大きさも意識しな

がら、漢字全体についての深い理解を求めているのである。「話すこと・聞くことを主とする指導」に15単位時間程度、「書くことを主とする指導」に30単位時間程度の配当時間が明示されていることとあわせると、今後漢字教育に要する時間は多くなると言えるはずである。ちなみに、日本漢字能力検定協会主催によると、同協会主催の漢字検定は年間200万名程であり、毎年増加しているとのことである。漢字の習得とその適切な運用は、学校教育はもとより一つの社会的要請になっているのである。

以上の点から見た場合、今後の国語教育において、豊かな表現力の基礎となる漢字の習得と適切な運用能力はその目標の根幹と深く関わってくると言えるだろう。学校現場では漢字に関する習得の程度を測定するための有効なテストの存在は欠かせないものとなり、ますます良質なものが求められる傾向が強くなることは間違いない。

日本語基礎能力テストは、今後想定される国語教育からの要請にも極めて有効な存在となるはずである。それゆえテストの項目に関しては、教育現場での運用の実用性という観点から、適切性および妥当性に関して十分に吟味される必要がある。以下、主として国語教育の観点から、日本語基礎能力テストにおける漢字テストの各項目特性を中心に検討した結果を報告する。

3. 漢字テスト項目評価基準

漢字テストの項目は、「読み」に特化したものであり、漢字の運用能力全体の測定をそのまま可能にするものではない。したがって前項の議論を受けるととともに、読み取りテスト特有の性質も加味しつつ、内容的な吟味を行う必要がある。

漢字テスト全体の中で、小学校段階の項目は、基本的に学習用の漢字の範囲で作成されたものである。小学校での漢字学習の主基準となる学年別漢字配当表に大幅な変更はみられず、また日常生活での使用に関しても10年余りの短い間で項目としての使用に問題となるような社会環境の変化等は起こっていないと考えられる。したがって、ここでは、中学生用・高校生用として作成された517項目のみに関して内容的な検討を行うこととする。

以下に評価の観点を示す。

(1) レベルA

レベルAは、以下の観点と基準にしたがった³。

基準1：意味に加え、使用機会や場面も通時的に安定しており、今後においても安定した継続使用がなされていくと思われるもの。

基準2：上記「基準1」により、各種のテストその他でも安定して取り上げられる機会が多いと思われるもの。

「レベルA」は、使用に関する通時的安定性を主基準とし、日常的な使用頻度の高さに加え、類義語との弁別の必要性等から各種テストにおいても様々な形式で出題される機会が多いと思われるもの、という観点から設定したものである。また、テストの出題形式（文形式）に関連し、文意からの類推可能性が極端に高いと思われる項目は含まれていない。実生活での使用頻度の高さ、理解に関する社会的な要求性の高さという点において安定性の高い項目である。

(2) レベルB

レベルBは、項目の安定性においてレベルAに準ずるものである。以下にその基準を示す。

基準1：いわば試験用の漢字といった傾向の強いものであり、就職・公務員等の試験では頻出であるが、日常的使用はひらがな表記が普通であり、漢字としての使用頻度は比較的少ないと思われるもの（非日常的・試験用）。

基準2：使用頻度において減少傾向にあると思われるもの、あるいは、意味・使用機会において通時的变化がみられるもの、関連して使用に世代差がある（若年層は用いない、等）と思われるもの（通時的变化）⁴。

基準3：設問の表現から類推によって正答できる可能性があり、漢字の読みの理解度との直接的な関わりがやや希薄と思われるもの（類推可能性）。

基準4：漢字の読みと語彙理解の問題の複合形になっており、正答にいたるには語彙に関わる能力がある程度必要と思われるもの（語彙との複合型）。

基準5：教科・科目に関わる内容で頻出の語であり、領域固有の知識や特定の分野で学習

する可能性があるもの（教科内容との複合型）。

「基準1」は、各種の試験問題として頻出のものであるが、日常生活においては比較的使用頻度が低いと思われるものである。学校教育現場での試験問題作成は、過去に実施したテストデータや、それと関連した参考書等に拠るところが大きい、という現実をふまえると、漢字テストの問題自体が大きく変化するとは考えられにくい。したがって学校教育で「頻出漢字」とみなされるものであっても、日常的に使用される頻度は決して高くない、と思われる場合もある。「基準1」に含まれるものは、日常的なものでありながら、日常性を保証するものに実使用以外の面も含む可能性があると思われる項目である。

「基準2」は、使用頻度において通時的変化や属性差が見られるものである。そうした項目では、漢字の使用はもとより、語彙としての理解や使用にも世代差が生じている可能性もある。被験者の言語形成期・言語生活において日常的なものか否か、あるいはまた、世代間で影響力の大きさが異なる各種メディアでの使用状況の違い等による影響が挙げられる。このように「基準2」は、通時的変化やそれに関連したコーホート効果が含まれる可能性もあるという点に基づく判断基準である。

「基準3」は、漢字テストの形式による類推可能性の高さに基づくものである。本テストが「文形式」をとる以上、文全体の意味から傍線部分の「読み」が類推できる場合もあると言えるだろう。本テストの項目からは慣用句的要素の強いものは除外しているが、文全体において使用上の安定性が高く、なおかつ傍線部分の漢字の「読み」に関して正解以外の「読み」の代替可能性が低いものである場合、正しい読みを理解していなくとも正解にいたる可能性があることは否定できない。ただし、こうした点は、言葉に関するテストに本来的に付随するものであり、テスト自体の安定性に大きく関わることはない。「基準3」は「レベルA」の項目と比較した際、類推可能性の比較的高いもの、という基準として設定したものである。

「基準4」は、語彙としての理解が漢字の「読み」を補助する面が少なくない、という観点によるものである。漢字は一般に「書き言葉」として表記の面から扱われることが多いが、そこには、表記と語彙

理解とが分離可能であるという前提が存在している。そうした扱いは、日常生活で自然に習得される基礎的な語彙の場合では有効な見方であるが、学習によって意味内容とあわせた形で新たに習得される漢字の場合、その理解は語彙としての理解と同時に終わる。このように「基準4」は、読みを答える際に語彙に関わる理解がある程度必要になるとと思われるものといった判断による基準である。

「基準5」は、主として学校教育において国語以外の教科・科目に関わる内容で扱う場合が多い、という基準に基づくものである。領域固有の知識や術語等の理解が読みの理解を補助しているものであり、日常的使用とは別の面で使用の安定性が保たれているという可能性がある、という判断基準によるものである。

以上がレベルBに関する評価基準である。

(3) レベルC

レベルCの項目は、以下の基準に基づいて判定した。

基準1：日常的な使用はほとんどなく、就職・公務員試験等での出題や学校で扱う教科書作品に出現するケースのみに留まっていると思われるもの。

基準2：専門用語など、特殊な分野で使用される傾向の強いものであり、設問として一般性に欠けるとと思われるもの。

「基準1」は、「レベルB：基準1」とも関連するが、日常生活を中心とした実使用面においては極めて限定されていると思われるものである。学校の国語の教材等における使用に限られたり、試験のために学習する漢字、いった要素が色濃く感じられるものである。

「基準2」は、教科・科目の枠外の専門用語など、特殊な分野に限って使用される傾向が強いものであり、そのまま読みの能力の測定に用いるのには不適切な面もあると思われるものである。

以下、各項目について、上記のレベル、基準に基づく評価を行う。

4. 項目評価結果

4.1. 項目プール全体の評価結果

はじめに項目プール全体の評価結果を提示する。表に示すように、項目プールの中の中学生用、高校

生用項目全体の比率としては、「レベルA」に該当する項目が84.9%（基準1が7.2%、基準2が77.8%）、「レベルB」が15.1%（基準1が4.1%、基準2が1.2%、基準3が1.0%、基準4が8.1%、基準5が0.8%）、「レベルC」に該当する項目は皆無であった。日常的使用を想定した「レベルA」の比率が圧倒的に高く、「レベルA」と比較すると日常的な使用がやや乏しい「レベルB」に該当する項目は少なかった。「レベルB」に該当する基準の中身も、大半は「基準4（基盤に語彙能力を必要とするもの）」と「基準1（試験で頻繁に取り上げられている漢字）」であり、漢字の読み取り能力の測定道具としては基本的に問題のないものである。

以上のことから、項目プール全体の妥当性とバランスの良さを確認することができた。

なお、次節で詳述するが、版の違いにより項目比率が若干異なっている。この比率の違いは基本的には難易度の問題に起因すると思われる。各版とも目的とする被験者の能力分布にうまく合致したものとなっていることが確認できる（詳細は、倉元・佐藤・伊藤 [2004] を参照のこと）。

4.2. 各版の評価結果

次に、版ごとに評価した結果について述べる。各評価レベル・基準による項目比率の版比較から以下の事が確認できる。

中学生版では、「レベルA：基準1」に該当する項目が他の版と比較すると際立って多かった。日常的にも使用頻度が高く、意味や使用機会も通時的に安定した項目であることが、中学生程度の能力を測るのに適しているのであろう。

表. 各レベル、各基準に分類された項目の数

		レベルA		レベルB					計
		基準1	基準2	基準1	基準2	基準3	基準4	基準5	
プール全体	項目数	37	402	21	6	5	42	4	517
	比率	7.2%	77.8%	4.1%	1.2%	1.0%	8.1%	0.8%	100.0%
中学生版	項目数	22	123	0	4	4	13	3	169
	比率	13.0%	72.8%	0.0%	2.4%	2.4%	7.7%	1.8%	100.0%
高校生 91 年度版	項目数	8	112	6	0	0	13	0	139
	比率	5.8%	80.6%	4.3%	0.0%	0.0%	9.4%	0.0%	100.0%
高校生 92 年 A 版	項目数	4	88	11	1	0	18	0	122
	比率	3.3%	72.1%	9.0%	0.8%	0.0%	14.8%	0.0%	100.0%
高校生 92 年 B 版	項目数	5	85	6	1	1	6	0	104
	比率	4.8%	81.7%	5.8%	1.0%	1.0%	5.8%	0.0%	100.0%
高校生 92 年 C 版	項目数	5	95	1	0	0	1	1	103
	比率	4.9%	92.2%	1.0%	0.0%	0.0%	1.0%	1.0%	100.0%

高校生用92年A版は全体として最も能力の高い被験者層を想定して作成された難易度の高い版である。そのために、「レベルB：基準4」に該当する項目が多くなったと思われる。難易度の高い漢字に関して、語彙理解力が基礎として必要となるのは自然である。また、「レベルB：基準1」に該当する項目も散見される。日常的使用はやや少ないものの、各種試験には比較的取り上げられることが多い項目である。この点もこの版の目的とする被験者の能力分布の高さに起因すると思われるので、版としての基本的な特徴が現れた結果となっている。

高校生用92年C版は高校生用の中では最も能力の低い被験者層を想定して作成された版である。90%以上の項目が「レベルA：基準2」に該当し、テスト項目としての質の高さが際立っている。

高校生用91年度版と高校生用92年B版は難易度的に両者の中間を目指して作成された版である。そのためか、項目の内容評価も概ね高校生用92年A版とC版の中間的な結果となっている。

4.3. 項目パラメタ値による評価結果

ここでは、漢字テスト項目の識別力パラメタ a 、困難度パラメタ b に関して、各評価基準との関連性を探り、各版の比較を行う。

項目パラメタ値の分布に版による違いが見られるので、語彙テスト項目について分析した伊藤他 (2003) と同様に、項目プール作成過程においてそれぞれの版を受験した解答者の特性値の分布を用いて標準化した値を用いる⁵。

なお、図1は、 $Za_{(i)}$ の自然対数を取ってその平均値を表示したものである。

まず顕著な特徴として、「レベルA」と「レベルB」を比較したときに、「レベルA」の識別力が高い、という傾向が見られないことである。むしろ、中学生版を除き、高校生用各版においては「レベルA：基準1」の値が低いことが確認できる。特に、全体として最も難易度が高い高校生用92年度A版においてはその傾向が顕著であり、「レベルA」の判定基準は、難易度的に中学生程度の項目に特に合致した基準であるということが推測されるものである。

「レベルB」の各基準に分類された項目は、数が少ないうえに版による偏りがあるために確たることは言えない。しかしながら、その中でも比較的項目数が多かった「レベルB：基準1」、「レベルB：基準4」を中心に解釈を加えることとする。「レベルB：基準1」の項目の識別力は、各版ともに平均値を上回っている。試験で頻繁に取り上げられていることから、漢字に対する意識的なトレーニングの成果として生まれる漢字読み取り能力の違いを識別できているのかもしれない。また、「レベルB：基準4」は、高校生用92年度B版、高校生用92年度C版を除き、平均的な識別力を下回っている。平均値を上回った2版においては該当する項目数が少ないので除いて考えるとすれば、能力分布の上下に関わらず、やや識別性能に劣ると言える。先述の通り、解

答の基盤となる語彙理解力が一種の誤差的な要因として影響し、識別力を相対的に押し下げていると考えるべきなのかもしれない。

次に項目パラメタb（困難度）に関し、標準化した値の平均値のグラフを提示する。

まず、「レベルA：基準2」の値はいずれの版でも平均値付近であるが、項目数が圧倒的に多いことが影響していると考えられる。「レベルA：基準1」の項目はいずれの版でも平均的な困難度を下回っており、識別力の様相から「レベルA」の判定基準が中学生程度の項目と合致しているのではないかという推測の傍証となっている。

次に、「レベルB：基準1」、「レベルB：基準4」についてであるが、いずれも全て平均的な困難度を上回っている。「レベルB：基準1」の各項目が意識的な漢字のトレーニングによって養成される漢字読み取り能力を反映しており、「レベルB：基準4」は基盤となる語彙能力を反映しているのではないか、という推測にも合致した結果である。ただし、以上の推測が妥当かどうかという点については、今後、さらに慎重な検討が必要なのは言うまでもない。

以上、識別力、困難度について、版、レベルごとに確認した。

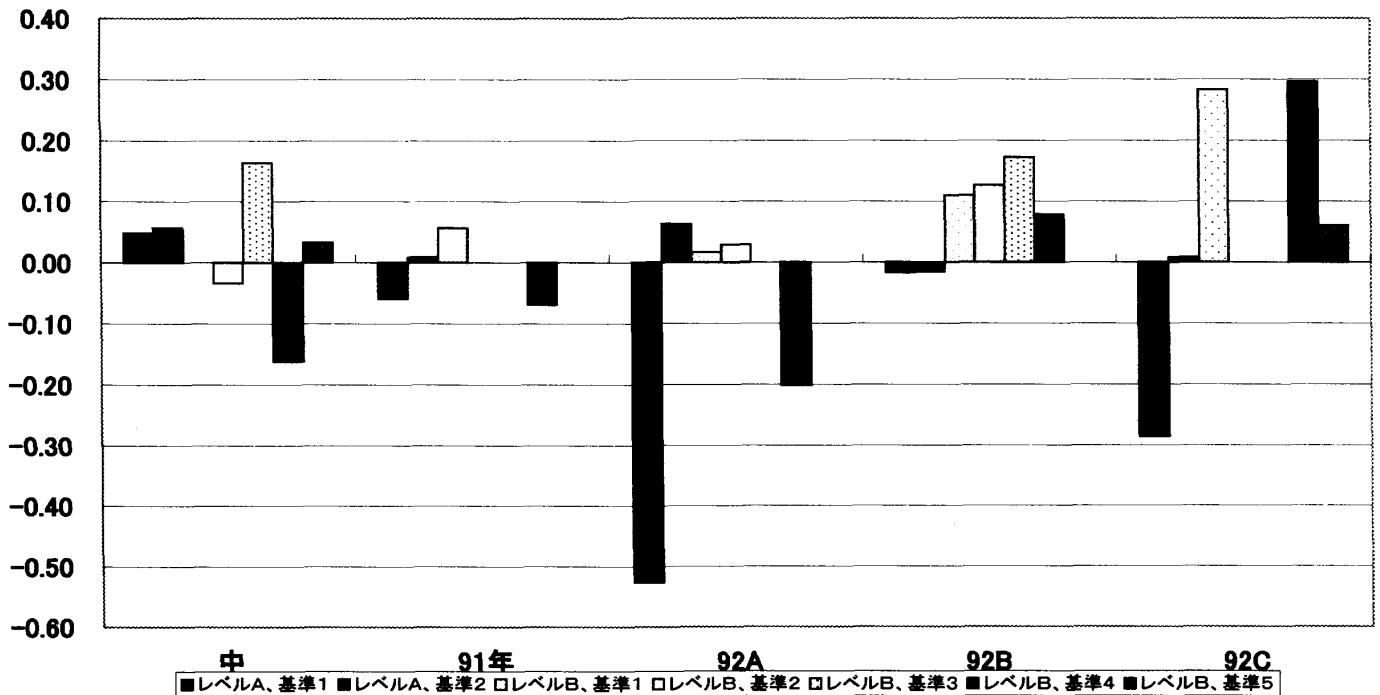


図1. 漢字テスト項目パラメタaの幾何平均値に対する比の値（自然対数の平均値）

5. まとめ

以上、日本語基礎能力テストについて、主として国語教育の観点から漢字テスト項目に関する評価を中心に論じてきた。あらかじめ項目評価基準を設定し、それに基づき各項目の評価を行ったが、中学生、高校生段階の全ての項目が「レベルA」および「レベルB」に含まれることが確認された。実使用面を考慮した場合のバランスは十分に高いと言える。

日本語基礎能力テストは、語彙理解力と漢字読み取り能力という2つの下位尺度を持つことが大きな利点である。語彙と漢字の習得過程は互いに密接に関連しているものの、必ずしも同一とみなすことはできない。特に、日常生活において言葉を取り巻く急激な環境の変化はこれらの関連性に大きな影響を及ぼしているかもしれない。高校生段階までの子どもたちの生活の中で、読書、直筆で文字を書く機会が減少し、変わってメールやインターネットの普及によって様々な書き手の文章を目にする機会、キーボードから文字を打ち込む機会が増えている。学校教育における教育課程の変化やこれらの日常生活の変化がことばや知性の発達にどのように影響しているのか、ことばの認知やことばの使用における世代間の差はどのように推移しつつあるのか、ことばとコミュニケーション能力の関係はどうなっているの

か、例えば、こういった課題に対してアプローチを行うための基礎的な資料を提供する道具として、日本語基礎能力テストは大いに利用価値があると思われる。

一方、漢字テストに特有の問題設定に対する応用可能性も残されている。最も大きな点は、「読み取りテスト」というテスト形式に関するものである。語彙テストのように誤答選択肢を用意し、多肢選択形式にした場合、あるいは、「書き取り」形式のテストとした場合、被験者の能力はどのように変わってくるのだろうか。このような課題に取り組むには、同一の被験者に多様な解答形式によるテストを実施する必要性がある。厳密な比較をするためには、通常、同一の問題を課さなければならないが、1つの形式のテストが残りの形式のテストのヒントとなってしまうという欠点がある。その点、日本語基礎能力テストは項目反応理論を用いて尺度化されているため、異なる項目を用いても結果は比較可能であり、このような構造的な技術的問題は解決できる。方法的に細部を詰める必要はあるが、「読み取り」の能力を機軸として、他の解答形式で測られる能力との関連を調べることも、有望な応用可能性の1つと言えよう。

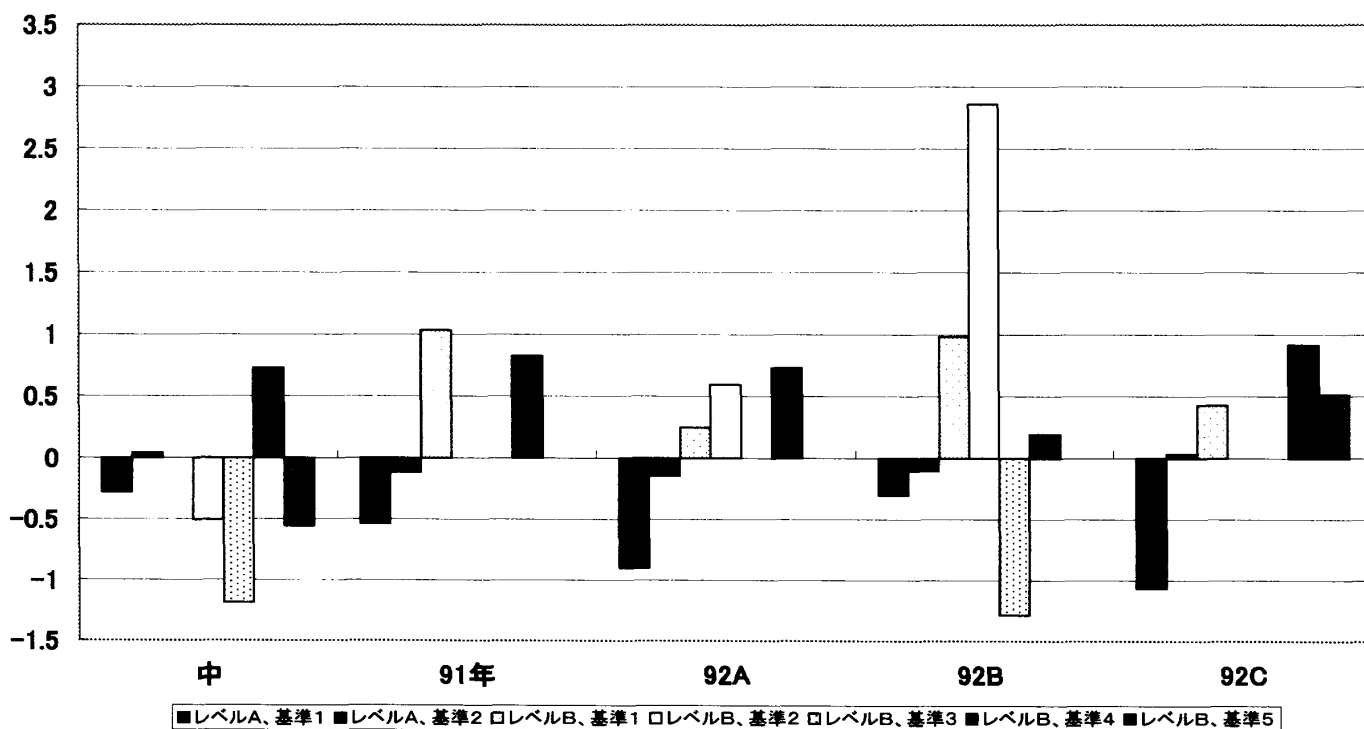


図2. 漢字テスト項目パラメタbの標準化した平均値

注釈

- 1 伊藤他 (2003) が執筆された時点での項目数は、語彙テスト470項目、漢字テスト977項目となっていた。その後、項目プール全体についてさらに慎重な吟味を行い、現時点で使用すべきではないと判断したものについて、語彙テストでは2項目、漢字テストでは3項目削除した結果、現在数に至っている。
- 2 Item Response Theory の訳である。項目応答理論とも呼ばれている。
- 3 伊藤・佐藤・倉元 (2003) における語彙テスト項目「レベルA」と同一である。
- 4 伊藤・佐藤・倉元 (2003) における語彙テスト項目「レベルA：基準2」と同一である。
- 5 すなわち、 x 版において標準化された項目パラメタの値は、

$$Z_{a(x)} = a/a_g \quad (1)$$

$$Z_{b(x)} = (b - \bar{\theta}_{(x)})/s_{\hat{\theta}_{(x)}} \quad (2)$$

ただし、 $Z_{a(x)}$ は x 版で標準化された項目パラメタ a の値、 a_g は x 版に含まれる項目パラメタ a の幾何平均、 $Z_{b(x)}$ は x 版の被験者パラメタの推定値 $\hat{\theta}$ の平均、分散によって標準化された項目パラメタ b の値、 $\bar{\theta}_{(x)}$ は、 x 版を受験した被験者の θ 推定値の平均、 $s_{\hat{\theta}_{(x)}}$ は、その標準偏差を表す (伊藤他、2003)。

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 (2002). 日本語語彙理解力テストの実施結果について、大学入学者選抜資料としての総合試験の開発的研究、平成11~13年度日本学術振興会科学研究費補助金 (基盤研究(B))、研究課題番号11551004、研究代表者 柳井晴夫、研究成果報告書、69-75.
- 倉元直樹・佐藤洋之・伊藤博美 (2004). 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 -、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 17-21.
- 倉元直樹・スコット寿美・笠居昌弘 (2003). 日本語語彙理解力テストの妥当性についての検討 - ルールスペース法を用いた認知論的分析 -、教育心理学研究、第51巻、413-424.
- 倉元直樹・龍岡菊美 (2001). ルールスペース法 (RSM) - テストの診断的利用に対する測定論的アプローチ -、平成11~12年度日本学術振興会科学研究費補助金 (奨励研究(A))、研究課題番号 11710087、「大学入試データによる学力の認知構造分析 - 総合試験設計のための基礎研究 -」、研究成果報告書.
- 文部省 (1999). 高等学校学習指導要領解説 国語編、東洋館出版社
- 小野博・繁榊算男・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.
- 佐藤洋之・伊藤博美・倉元直樹 (2003a). 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.
- 佐藤洋之・伊藤博美・倉元直樹 (2003b). 高校生における日本語基礎能力の測定 - ここ10年余りの変化を中心に -、日本行動計量学会第31回大会発表論文集、224-227.
- 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、第43巻、134-144.
- 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.
- 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、No.21、107-135.
- 平直樹・小野博・林部英雄 (1993). 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、No.22、53-82.
- 平直樹・小野博・前川眞一・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 - 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、第43巻、68-73.

Properties of the Broad-Range Japanese Fundamental Language Skills Test III: Item Contents Review of Chinese Character Pronunciation Subscale from the Viewpoint of First Language Education

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

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

*** Honjo Akita Prefectural High School*

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

The purpose of this study is to reinvestigate the Chinese Character Pronunciation items of the Broad-Range Japanese Fundamental Language Skills Test (BRJFLST) developed by Taira et al. (1998). We examined 517 items designed for measuring junior and senior high school students' ability level.

Items were reviewed from the viewpoints of Japanese language education as the first language. Three levels of category were set for the purpose of evaluation: Level A as the 'considerably stable in the sense of meaning and social significance over time', Level B as 'fairly stable in the sense of meaning and social significance over time, having some features from a particular viewpoint' and Level C as 'limited use in daily life or technical terms of specific fields'. Each level was divided into two to five categories according to other criteria mainly set from the viewpoints of usage and exposure in the testing situations.

The result showed that 85% of items were classified into level A, while not a single item was evaluated as level C. It verifies that the item pool is still functioning well as a measurement tool for basic skills of the Japanese language at target as a whole, even a decade passed since the items were created.

Key words: first language education, high school students, item pool, Chinese character, measurement