



Title	文法・漢字プレースメントテストの信頼性と項目識別力の検証
Author(s)	大和, 祐子
Citation	日本語・日本文化. 2016, 43, p. 27-52
Version Type	VoR
URL	https://doi.org/10.18910/56968
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

〈研究論文〉

文法・漢字プレースメントテストの 信頼性と項目識別力の検証

大和 祐子

1. はじめに

本研究は、留学生向けの文法・漢字プレースメントテストの信頼性とその利用可能性について検証するものである。日本語教育現場では、コースの開始時、コースの途中、コースの終了時など様々な場でテストを行い、学習者の日本語能力を測定する。このようなテストは、(1) 妥当性、(2) 信頼性、(3) 問題の包括性、(4) 実施容易性、(5) 採点の客観性、(6) 採点の容易性、(7) 解釈の容易性、(8) 利用の容易性、(9) 経済性を備えている必要がある(石田、1992)。その中でも、特にテストの(1) 妥当性と(2) 信頼性は欠かせないものである。

教育機関でのプレースメントテストは、診断的評価の代表的な例としてしばしば挙げられる。プレースメントテストの最大の目的は、適切なクラス配置を行うことである。そのために、プレースメントテストで留意すべき点として、受験者間の日本語能力の違いが明確になるだけの多種多様な項目(問題)を準備すること、それらの項目は受験者の能力を識別するものであること、そしてその項目が易しいものから難しいものへと順に配置されていること、などが挙げられる(伊東、2008)。また、主たる受験者の想定される日本語能力によっても、どのようなプレースメントテストがふさわしいのかは異なる。

プレースメントテストの適切性を統計的に検討した研究としては、三枝(1986、1987)、酒井(1990)、今村(2001)、小森(2011)、渡部(2012)などがある。これらは各教育機関で実施されているプレースメントテストを基に分析したものである。主に、多肢選択式の文法・語彙・漢字テストの項目分析を行った研究(三枝、1986、1987; 酒井、1990)、聴解テストまたは読解テストの項目分析とその結

果がクラス分けの結果を統計的に説明することができるかを調べた研究（小森、2011；渡部、2012）があり、それぞれの研究からブレースメントテスト項目作成において有益な知見を得ることができる。しかし、今村（2001）でも指摘されているように、各教育機関の受験者層、クラス編成、教育内容などを考慮したテストの分析が欠かせない。

大阪大学日本語日本文化教育センターにおいても、ブレースメントテストの受験者層は近年変化している。本稿で取り上げるブレースメントテストが最初に作成された段階では、本センターへ留学してくる学生のうち、一定数の日本語未習者が含まれていた。そのため、ブレースメントテスト作成にあたっては、日本語の言語知識と呼ばれる文法知識・漢字知識を測定する項目を数多く準備することが重要であった。しかし、現在は日本語日本文化研修留学生をはじめとする日本語既習者を受け入れることがほとんどである。また、学習者の母国での日本語学習環境も多様になっており、そのような多様なバックグラウンドを持つ留学生の日本語能力をどう評価すべきかは、現在の大きな課題となっている。さらに、近年、日本語を用いて課題達成するために、その根底にあるものとして、日本語の言語知識をとらえるのが一般的になっている。このようなブレースメントテスト受験者層の多様化、日本語能力観の変容とそれに伴う教育内容の変化を考慮したテストの再分析が必要である。

そこで、本稿では大阪大学日本語日本文化教育センターで使用されているブレースメントテストを分析し、現在実施している文法・漢字知識を問うブレースメントテストの有用性を検討する。

2. 研究目的

本研究では、古典的テスト理論（classical test theory：CTT）を用いて、現行のブレースメントテストを再分析する。留学生向けの文法・漢字ブレースメントテスト（以下、文法・漢字テスト）の信頼性を検討するために、文法・漢字テスト全体の記述統計量を示すとともに、信頼性係数（クロンバックの α 係数）、各項目の項目困難度（通過率）、識別力（点双列相関係数）から文法・漢字テストが有用なものであるか、検討する。次に、G-P 分析（Good-poor analysis）を通して、

項目困難度や識別力が著しく高かった項目、低かった項目、その他特徴的であった項目を中心に、項目を質的に分析する。最後に、文法・漢字テストの有用性を高めるために、どのような改訂が望ましいか提案する。

3. 文法・漢字テストの概要

本研究の分析対象となる漢字・文法テストは、大阪大学日本語日本文化教育センターでクラス分けに使用されているテストである。日本語科目を履修する留学生は、正規生か否かにかかわらず、科目選択までに受験するよう指示されている。なお、このテストは、オンラインで受験可能なものである。

このプレースメントテストは、文法問題 164 問、漢字問題 220 問からなる。文法問題はさらに A 問題 70 項目、B 問題 94 項目に分類することができ、項目は 4 者択一もしくは 3 者択一の問題である。表 1 は、各セクションの問題構成と出題された問題例である。表 1 の通り、文法テストとはいえ、A5 セクションの問題のように語彙知識を問う項目、B4 セクションの問題のように（読解問題とまでは言えないものの）短いテキストを読み内容の理解を問う項目があることが分かる。このような構成でテスト問題を作成した背景には、試験開発当時の受験者の日本語能力が影響していると考えられる。試験開発当時、主たるプレースメントテスト受験者の日本語レベルが初級であったこと、コースでは東京外国語大学留学生日本語教育センターの『初級日本語』が使用されていたことを勘案し、A 問題は初級前半レベル、B 問題は初級後半レベルの文法項目をターゲットとして作題された。なお、漢字能力の影響を排除するために問題文に含まれる漢字にはふりがながふられ、問題文のその他の語に関しても難易度が調整された。

一方、漢字問題はそのターゲットのレベルにより、A 問題、B 問題、C 問題に分けられる。表 2 に示した通り、A 問題では漢字の「書き」を問う問題が計 50 項目（A1 セクション、A3 セクション）、漢字の「読み」を問う問題が 30 項目あり、B 問題と C 問題ではそれぞれ漢字の「読み」を問う問題が 30 項目、漢字の「書き」を問う問題が 40 項目出題された。以上の漢字テストの全ての項目は、5 者択一の問題である。試験開発当時、主たるプレースメントテスト受験者の日本語レベルが初級であったこと、コースでは東京外国語大学留学生日本語教育セン

表 1 文法ブレースメントテストの構成と問題例

レベル	セクション	項目数	問題例				錯乱肢①	錯乱肢②	錯乱肢③
			設問文	正答	に	が			
A	1 適切な助詞を問う問題	15	【A1-II】父はまいあさ ぎんこうのまえ（ ）とおつて会社へ行きます。	を				で	
	2 質問に対して文法的に正しい答えを選ぶ問題	20	【A2-20】あなたのともだちは、どこかへよりこうしたいといっていますか。	はい、ともだちはおきなわへ行きたいです。	はい、ともだちはおきなわへ行きたいです。	はい、ともだちはおきなわへほしがつています。		—	
	3 語を適切な活用を問う問題	15	【A3-14】りーさんはあしたのパーティーに_____にちがひありません。	くる	きて	こよう		きます	
	4 文意に合う語表現を選ぶ問題	10	【A4-5】くつをはいた_____へやにはいってはいけません。	まま	ながら	ないで		まで	
	5 文脈に合う語を選ぶ問題	10	【A5-10】あしたはわたしのたんじょうびパーティーですから、_____きてください。	かならず	たぶん	めったに		おそらく	
B	1 文意に合う語表現を選ぶ問題	50	【B1-34】今夜から明日の朝（ ）大雨になるそうだ。	に対して	に関して	について		にかけて	
	2 文意に合う後件を選ぶ問題	15	【B2-5】先生の言うことが、かならずしも_____。	正しいとはかぎらない	きくべきだ	間違っていることもある		認められるだろう	
	3 下線部と類似した意味の表現を選ぶ問題	25	【B3-19】田中さんが授業を休むとは <u>あえがたい</u> 。	考えにくい	考えやすい	考えられるかもしれない		考えたい	
	4 短いテキストを読み、内容理解を問う質問に答える問題	4	【B4-1】日本人は、道で人と会ったときに、よく「どちらへいらっしゃいますか。」とたずねます。外国人からすると、「どこへ行こうと、あなたには関係ないですよ。わたしの自由です。」と言いたくなるかもしれません。しかし、これはただのあいさつなのです。だから、聞かれた人は「ちやうとそこまで。」とこたえればいいのです。「今からご飯を食べ、映画を見た後で本を買って帰るつもりです。」などと正直にこたえろと、聞いた人も困ってしまいます。 上の文章の内容に合っているものを一つえらびなさい。	「どちらへいらっしゃいますか。」とたずねる人は、本当にそれを知りたい。 「どちらへいらっしゃいますか。」という質問には、正直にこたえたい方がいい。	外国人のあいさつは、日本人のあいさつと関係がない。	日本人では、道で会ったときに「飯や映画」の話をするのはよくないと思われる。			

ターの『初級日本語』が使用されていたことを勘案し、A問題は入門レベル、B問題は初級前半レベル、C問題は初級後半レベルの漢字をターゲットとして作題された。各セクションで出題された問題例は、表2に示す通りである。

表2 漢字プレースメントテストの構成と問題例

レベル	セクション	項目数	問題例					
			設問文	正答	錯乱肢①	錯乱肢②	錯乱肢③	錯乱肢④
A	1 漢字書き問題 (字形の類似を選択肢構成としたもの)	20	【A1-13】 私は しょうらい <u>い</u> 者に な りたいです。	医	匡	区	匠	巨
	2 漢字読み問題	30	【A2-8】 <u>切手</u> を ください。	きって	きうて	きて	きつて	きいて
	3 漢字書き問題 (同音異字または意味を選択肢構成としたもの)	30	【A3-28】 ながい <u>き</u> 間、国へ かえってい ました。	期	気	記	機	季
B	1 漢字読み問題	30	【B1-6】 めずらしい <u>植物</u> だ。	しょくぶ つ	どうぶつ	おきもの	いきもの	わかもの
	2 漢字書き問題	40	【B2-4】 その川はとても <u>ふかい</u> 。	深い	泣い	泳い	浅い	洗い
C	1 漢字読み問題	30	【C1-26】 <u>謝罪</u> してもらった。	しゃざい	しょくざ い	むざい	かんしゃ	ちんしゃ
	2 漢字書き問題	40	【C2-26】 車の会社に <u>しゅう</u> 職 しました。	就	集	着	修	祝

4. 分析対象

本研究では、プレースメントテストを受験した留学生 271 名（2014 年 9 月来日留学生 131 名、2015 年 9 月来日留学生 140 名）のデータを用いた。本研究では受験者 1 人 1 人に着目した被験者分析ではなく、プレースメントテストで出題された項目（問題）に着目した項目分析を行った。項目分析は、各項目の困難度 (difficulty)、識別力 (discrimination power)、多肢選択式の項目の場合は各選択肢の選択状況などについて検討される（野口・大隅、2012）。

5. 結果

5.1. 文法テスト

5.1.1. 記述統計量

本稿で対象とした留学生 271 名の文法テストの得点平均と標準偏差は、表 3 の通りである。テストは 1 問 1 点として 164 点満点で算出された。また、文法テストの点数で受験者は上位群（149 点以上）・中位群（122 点以上 148 点以下）・下位群（121 点以下）の 3 グループに分類された。その結果、上位群が 89 名、中位群が 92 名、下位群が 90 名となった。

表 3 文法ブレースメントテストの平均と標準偏差

	満点	全体 (N=271)		上位群 (N=89)		中位群 (N=92)		下位群 (N=90)	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
文法テスト全体	164	131.23	24.35	155.39	3.61	136.83	8.20	101.61	14.61
A1セクション	15	12.98	2.19	14.58	0.60	13.38	1.46	10.99	2.29
A2セクション	20	17.91	2.07	19.39	0.70	18.30	1.36	16.04	2.18
A3セクション	15	13.52	1.76	14.51	0.68	13.92	1.01	12.12	2.18
A4セクション	10	7.82	2.02	9.31	0.76	8.30	1.21	5.84	1.96
A5セクション	10	9.87	0.42	9.94	0.23	9.98	0.15	9.68	0.63
B1セクション	50	38.58	9.65	47.72	1.48	40.93	3.77	27.12	6.60
B2セクション	15	10.42	3.33	13.63	1.14	10.61	1.97	7.06	2.56
B3セクション	25	17.32	5.69	22.70	1.44	18.39	2.80	10.90	4.06
B4セクション	4	2.82	1.11	3.61	0.54	3.00	0.84	1.86	1.06

5.1.2. 信頼性

文法テストの信頼性を調べるために、クロンバックの α 係数を求めた。クロンバックの α 係数は、テストの信頼性を調べる際に一般的に使用されている推定法である。 α 係数は 0 から 1 の間の数値で出されるが、能力テストの場合、 α 係数が 0.8 以上であることが望ましいとされている（中村、2002）。

本調査で用いた文法テスト（164 問）の場合、 $\alpha = 0.967$ と極めて高い値が出たことから、この文法テストの信頼性は高いと判断できる。

5.1.3. 各項目の困難度と識別力

まず、各項目の困難度を算出した。通常、困難度の指標として通過率（正答

率) が用いられる。これは、正答者数の全体数に対する比率のことである。この通過率は全員が不正解であった場合は 0、全員が正解であった場合は 1 となる。したがって、通過率は 0 から 1 の間で示され、この項目が 0 に近い項目ほど易しい項目であり、1 に近い項目ほど難しい項目である。表 4 に示すのは、文法テストの困難度の平均をセクションごとにまとめたものである。

表 4 文法プレースメントテストの項目困難度平均

	全体 (N=271)	上位群 (N=89)	中位群 (N=92)	下位群 (N=90)
文法テスト全体	0.811	0.945	0.842	0.645
A1セクション	0.866	0.972	0.892	0.733
A2セクション	0.896	0.970	0.915	0.802
A3セクション	0.901	0.967	0.928	0.808
A4セクション	0.782	0.931	0.830	0.584
A5セクション	0.987	0.994	0.998	0.968
B1セクション	0.772	0.954	0.819	0.542
B2セクション	0.695	0.909	0.707	0.470
B3セクション	0.693	0.908	0.736	0.436
B4セクション	0.705	0.902	0.750	0.464

プレースメントテストは、期末試験などのような到達度を測るテストとは異なるため、受験者の正答率が高いことが重要ではなく、受験者の能力を識別できることが重要である。そう考えると、A5 セクションの語彙を問う項目(全 10 項目)は、全体の困難度平均が 0.987 と受験者層全体の日本語能力から考えると易しすぎる項目であったことが分かる。特に、グループ間の困難度の差も見られないことから、プレースメントテストには不適当な困難度の項目だったと言えよう。

次に、各項目の識別力を算出した。識別力は、各項目が測定したい能力に対して受験者間の水準の差を正確に反映できるかを表し、点双列相関係数(受験者の当該項目の正誤とテスト全体の正答数との相関)で表すものである(野口・大隅、2012)。識別力は、-1.000 から +1.000 の範囲内で示される。1 に近いほど受験者間の水準の差を正確に反映できると言われている。逆に、この数値が 0 に極

めて近い、もしくはマイナスになる場合には、項目として受験者の能力を識別できないことから、項目を改善する必要があると考えられている。例えば、項目の困難度（通過率）が著しく高い場合、著しく低い場合には、識別力は低くなる。一般的には、識別力を表す点双列相関係数が 0.3 以上であれば、受験者を識別できる良問であるとされている（中村、2002）。

本調査で用いた文法テスト（164 問）の場合、各項目の点双列相関係数の平均は 0.386 で、一般的に良問とされる 0.3 を超えている項目がある程度あることから、概ね良問がそろっていることが分かった。ただし、識別力の高い項目とそう

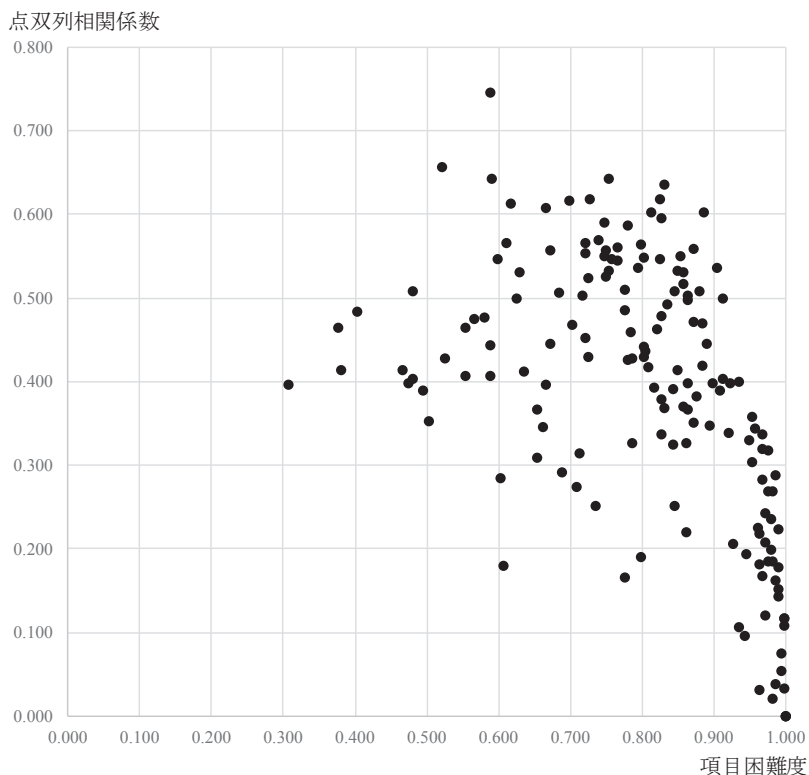


図 1 文法プレースメントテストの項目困難度と識別力

でない項目のばらつきは大きかった。図1は、文法テストの各項目の項目困難度(通過率)と識別力(点双列相関係数)の散布図である。

これを見ると、広い範囲から様々な困難度の項目を出題することが望ましいとされるプレースメントテストとしては、やや困難度の低い項目が多く、それらの中には点双列相関係数が高いとは言えない項目があることが分かる。図1からも分かるように項目困難度が低い項目が必ずしも識別力の低い項目であるとは言えないが、識別力が低い項目がどのようなものであるか、またこのテストの総合得点の上位群・中位群・下位群の各群の解答行動から項目としての問題点は何か検討する必要がある。

5.1.4. G-P 分析

G-P 分析 (Good-poor analysis) とは、本稿で対象としているテストのような多肢選択式の問題の各選択肢の選択状況から、項目の良し悪しを検討するものである(野口・大隅、2012)。ここでは、識別力が高かった項目と低かった項目、及びその他特徴的な項目を取り上げ、各選択肢の選択率などから項目の良し悪しを検討する。

164 問のうち、識別力が最も高かったのは【A1-I1】で、この項目の識別力を表す点双列相関係数は、0.747 であった。選択肢の * は正答を表し、各選択肢の() 内は選択率を表している(以下、同じ)。なお、実際の問題文及び選択肢の漢字には、全てふりがながつけられている。

【A1-I1】

父は まいあさ ぎんこうのまえ () とおって会社へ行きます。

①に (32.47%) ②*を (58.67%) ③が (0.37%) ④で (7.01%)
(無答 : 1.48%)

この項目は、助詞を問う問題である。この項目について、対象とした 271 名を文法の総合得点で上位群 (149 点以上 ; N=89)、中位群 (122 点以上 148 点以下 ; N=92)、下位群 (121 点以下 ; N=90) に分けた 3 グループのそれぞれの選択肢の

選択率は、表 5 の通りである。

A レベルの項目として出題された【A1-11】は、受験者全体の項目困難度が 0.586 と、受験者にとって易しい項目ではなかったことが分かる。このような結果になったのは、場所を表す「ぎんこうのまえ」と共に使われることの多い助詞「①に」(32.47%)や「④で」(7.01%)が錯乱肢として有効に機能したためだと考えられる。特に「①に」は設問文にある「行きます」にのみ着目した受験者には魅力的な錯乱肢であったと思われる。表 5 を見る限り、「①に」の選択率は、上位群 (29.21%) より中位群 (32.61%) で、中位群より下位群 (35.56%) で高かったことが分かる。以上のように【A1-11】は、設問文や選択肢構成により、項目困難度が上がり、結果として受験者の能力の識別に成功したと言える。

表 5 【A1-11】各選択肢の選択率

選択肢	①	②*	③	④	無答
全体 (N=271)	88 32.47%	159 58.67%	1 0.37%	19 7.01%	4 1.48%
上位群 (N=89)	26 29.21%	55 61.80%	0 0.00%	6 6.74%	2 2.25%
中位群 (N=92)	30 32.61%	55 59.78%	0 0.00%	6 6.52%	1 1.09%
下位群 (N=90)	32 35.56%	49 54.44%	1 1.11%	7 7.78%	1 1.11%

次に、他のセクションの項目で識別力が高かった項目として【B1-49】を挙げる。この項目の識別力を表す点双列相関係数は 0.658 であった。

【B1-49】

試験に合格したという知らせを聞いて嬉しさの（ ）、涙が出てきた。

① *あまり (52.03%) ②とても (1.11%) ③以上 (31.0%)

④さらに (14.02%)

(無答：1.85%)

この項目について、対象とした 271 名を文法の総合得点で上位群 (149 点以上; N=89)、中位群 (122 点以上 148 点以下; N=92)、下位群 (121 点以下; N=90) に分けた 3 グループのそれぞれの選択肢の選択率は、表 6 の通りである。

表 6 【B1-49】各選択肢の選択率

選択肢	①*	②	③	④	無答
全体 (N=271)	141 52.03%	3 1.11%	84 31.00%	38 14.02%	5 1.85%
上位群 (N=89)	46 51.69%	2 2.25%	27 30.34%	11 12.36%	3 3.37%
中位群 (N=92)	53 57.61%	0 0.00%	31 33.70%	8 8.70%	0 0.00%
下位群 (N=90)	42 46.67%	1 1.11%	26 28.89%	19 21.11%	2 2.22%

この項目の正答の選択肢は「①あまり」で 4 つの選択肢の中では最も受験者に選ばれているが、「③以上」も各群の受験者にも 30% 前後選択されており、どのグループの受験者にとっても魅力的な選択肢であったと考えられる。さらに、下位群については「④さらに」も 21.11% 選ばれており、上位群 (12.36%)・中位群 (8.70%) とは異なる解答行動を見せている。このことは、下位群においては、「④さらに」が有効な選択肢として機能していたことを意味している。

一方、改善する必要があるとされる項目も散見された。本調査で対象とした 164 問のうち識別力を表す点双列相関係数がマイナスとなる項目はなかったものの、0 となった項目は 3 項目あった。これらはいずれも、項目の困難度 (通過率) が 100% の項目、すなわち 271 名の受験者全員が正解した項目であった。次に示す【A5-2】は、点双列相関係数が 0 となった項目の 1 つである。

【A5-2】

あさおきたとき、_____ といいます。

- ① * おはよう (100%) ② こんばんは (0%)
 ③ いただきます (0%) ④ どういたしまして (0%)
 (無答: 0%)

この項目は日本語の定型表現の知識を問う問題である。その中でも極めて基本的な挨拶表現をターゲットとしており、他の錯乱肢も難易度の低いものである。この問題を見る限り、この項目に全員が正解できたのは、問題文や選択肢などに不備があるというより、むしろ受験者には易しすぎ、受験者の水準の差を反映するには適していなかった項目だと言える。

しかし、難しい項目であっても受験者の水準の差を反映しているとは言えない項目もあった。次に示す【B3-14】は、項目困難度 0.605 と決して易しい項目とは言えない項目であったが、点双列相関係数は 0.180 と識別力が高くない項目であった。

【B3-14】

国の基礎を作るのは教育にほかならない。

- ① * 教育以外にはない (60.52%) ②他の教育になる (1.11%)
 ③教育だけではない (7.75%) ④教育にちがいない (25.09%)
 (無答 : 5.54%)

この項目は、下線部の「教育にほかならない」と類似した表現で言い換える場合、適切なものを選ぶという問題である。この項目について、対象とした 271 名を文法の総合得点で上位群 (149 点以上 ; N=89)、中位群 (122 点以上 148 点以下 ; N=92)、下位群 (121 点以下 ; N=90) に分けた 3 グループのそれぞれの選択肢の選択率は、表 7 の通りである。

表 7 【B3-14】各選択肢の選択率

選択肢	①*	②	③	④	無答
全体 (N=271)	164 60.52%	3 1.11%	21 7.75%	68 25.09%	15 5.54%
上位群 (N=89)	52 58.43%	1 1.12%	6 6.74%	23 25.84%	7 7.87%
中位群 (N=92)	59 64.13%	2 2.17%	6 6.52%	22 23.91%	3 3.26%
下位群 (N=90)	53 58.89%	0 0.00%	9 10.00%	23 25.56%	5 5.56%

表7を見ると、上位群の正答「①教育以外にはない」の選択率（58.43%）が全体の選択率（60.52%）とほぼ同じことから、上位群の受験者にとっても正答を選ぶことは容易ではなかったということが分かる。また、錯乱肢の中では、「②他の教育になる」の選択率はどのグループでも低いが、「④教育にちがいない」の選択率はどのグループでも高い。さらに、上位群の「④教育にちがいない」の選択率（25.84%）は中位群（23.91%）、下位群（25.56%）と同程度であり、3グループのこの項目の解答行動は、ほぼ同じである。このことから、受験者の水準の差を反映することができなかったと推測される。したがって、【B3-14】をプレースメントテストで出題するには、中位群及び下位群の識別に有効な錯乱肢を再構成する必要がある。

5.2. 漢字テスト

5.2.1. 記述統計量

本稿で対象とした留学生 271 名の漢字テストの得点平均と標準偏差は、表8の通りである。テストは1問1点として220点満点で算出された。また、漢字テストの合計点で受験者は上位群（216点以上）・中位群（197点以上215点以下）・下位群（196点以下）の3グループに分類された。その結果、上位群が99名、中位群が90名、下位群が82名となった。

このグループ分けの結果からも分かるように、この漢字テストの受験者の上位

表8 漢字プレースメントテストの平均と標準偏差

	満点	全体 (N=271)		上位群 (N=99)		中位群 (N=90)		下位群 (N=82)	
		平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
漢字テスト全体	220	198.59	28.06	217.96	1.38	207.82	5.29	165.09	30.02
A1セクション	20	19.52	1.17	19.99	0.10	19.76	0.50	18.71	1.82
A2セクション	30	29.05	1.88	29.75	0.48	29.34	0.91	27.89	2.93
A3セクション	30	28.86	2.69	29.91	0.32	29.58	0.60	26.79	4.17
B1セクション	30	28.80	3.26	29.92	0.27	29.81	0.49	26.35	5.14
B2セクション	40	35.76	6.57	39.78	0.46	38.14	1.51	28.29	7.66
C1セクション	30	25.00	6.31	29.59	0.73	27.07	2.01	17.21	5.98
C2セクション	40	31.59	9.25	39.03	1.01	34.12	3.50	19.84	7.48

群・中位群にはほぼ満点に近い得点を取っている者もいた。ただし、このテストの最低点は 220 点満点で 30 点と受験者間の能力の差が大きかった。

5.2.2. 信頼性

本調査で用いた漢字テスト 220 問の信頼性を知るために、クロンバックの α 係数を求めた。その結果、漢字テスト（220 問）の信頼性係数は、 $\alpha = 0.983$ と極めて高く、基準となる 0.8 を大きく上回っていることから、テストに十分な信頼性があることが確認された。

5.2.3. 各項目の困難度と識別力

まず、各項目の困難度を算出した。表 9 に示すのは、文法テストの困難度の平均をセクションごとにまとめたものである。

表 9 漢字プレースメントテストの項目困難度平均

	全体 (N=271)	上位群 (N=99)	中位群 (N=90)	下位群 (N=82)
漢字テスト全体	0.905	0.992	0.951	0.773
A1セクション	0.974	0.999	0.988	0.935
A2セクション	0.966	0.992	0.978	0.930
A3セクション	0.959	0.997	0.986	0.893
B1セクション	0.956	0.997	0.994	0.878
B2セクション	0.885	0.994	0.954	0.707
C1セクション	0.821	0.986	0.902	0.574
C2セクション	0.775	0.976	0.853	0.496

表 9 の A レベルの各グループの困難度はほぼ全てのセクションで 0.900 を超えており、A レベルの項目がプレースメント受験者にとっていかに易しい項目であったかが分かる。このことから、少なくとも A レベルの項目は受験者の識別には有効ではないと考えられる。また、このプレースメントテストは、漢字能力上位群の受験者にとって、C レベルの項目であっても困難度が 0.986（C1 セクション）、0.976（C2 セクション）と易しすぎる項目であったことが分かる。上位群に限らず、中位群においても天井効果が出ていることは明らかである。

次に、各項目の識別力を算出した。その結果、本調査で用いた漢字テスト（220問）の場合、各項目の識別力の平均は0.464で、一般的に良問とされる0.3を超えていることから、概ね良問がそろっていることが分かった。ただし、識別力の高い項目とそうでない項目のばらつきは大きかった。次に示す図2は、文法テストの各項目の項目困難度（通過率）と識別力（点双列相関係数）の散布図である。

項目困難度の平均を見ても分かるように、本稿の受験者層には適応していると言えない項目もあったが、図2を見ると、項目困難度が1.0に近い項目であっても、識別力を表す点双列相関係数が高いものも低いものもあることが分かる。分析の対象となっている漢字プレースメントテストの中で、最も識別するとされた項目の点双列相関係数は0.717であるなど、0.6を超える項目が220問中59項目ある

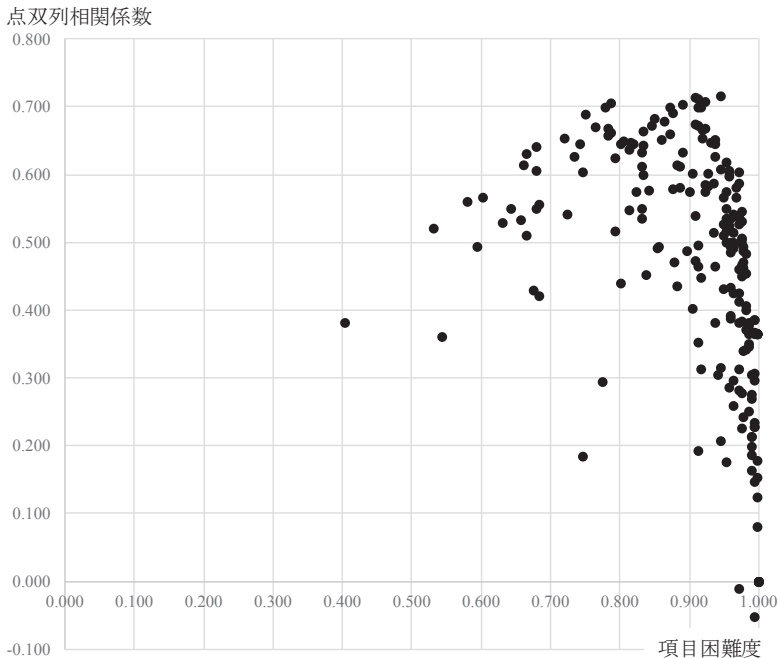


図2 漢字プレースメントテストの項目困難度と識別力

一方で、0.1 を下回る項目が 220 項目中 10 項目あった。そこで、識別力が低い項目がどのようなものであるか、またこのテストの総合得点の上位群・中位群・下位群の各群の解答行動から項目としての問題点は何か検討する。

5.2.4. G-P 分析

ここでは、識別力が高かった項目と低かった項目、及びその他特徴的な項目を取り上げ、各選択肢の選択率などから項目の良し悪しを検討する。220 問の漢字テストのうち、識別力が最も高かったのは【A3-14】の項目で、この項目の識別力を表す点双列相関係数は 0.717 であった。【A3-14】の問題文、選択肢と各選択肢の選択率を示す。

【A3-14】

かれが それを きいたら、とてもよろこぶでしょう。

①楽 (0.74%) ②*喜 (94.46%) ③悲 (1.85%) ④飛 (1.11%) ⑤遊 (0%)
(無答 : 1.85%)

一見、正答である「②喜」を選択した受験者が 94.46% と圧倒的に多く、受験者の水準の差を反映するには適していないかのように思われる。次に示す表 10 は、【A3-14】について、受験者 271 名を漢字の総合得点で上位群 (216 点以上 ; N=99)、中位群 (197 点以上 215 点以下 ; N=90)、下位群 (196 点以下 ; N=82) に分けた 3 グループの、それぞれの選択肢の選択率である。

表 10 【A3-14】各選択肢の選択率

選択肢	①	②*	③	④	⑤	無答
全体 (N=271)	2 0.74%	256 94.46%	5 1.85%	3 1.11%	0 0.00%	5 1.85%
上位群 (N=99)	0 0.00%	99 100.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
中位群 (N=90)	0 0.00%	90 100.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%
下位群 (N=82)	2 2.44%	67 81.71%	5 6.10%	3 3.66%	0 0.00%	5 6.10%

これを見ると、全体の正答率が 94.46% と易しい項目であったことは言うまでもないが、上位群と中位群の受験者は全員正答を選んでおり、誤答となる選択肢を選んだ受験者は、全て下位群の受験者であったことが分かる。つまり、このことから、中位群以上と下位群の漢字能力の差を反映している項目であるため、この項目は識別力のある項目であると判定されたと考えられる。ただし、各選択肢の選択率を見てみると、「⑤遊」は漢字能力の高低に関係なく、全く選択されていないことが分かる。つまり、この項目における「⑤遊」は選択肢として機能しておらず、受験者にとっての実質的な選択肢にはなっていなかったと推測される。今後、この項目を改善する余地があるとすれば、それは錯乱肢の工夫であると考えられる。

一方、受験者の識別に機能せず、項目の改善が必要とされる、点双列相関係数が 0.1 を下回る項目は 220 問中 10 項目あった。そのうち、7 項目は点双列相関係数が 0 の項目で、2 項目は点双列相関係数がマイナスとなった項目であった。これら 10 項目は全て A レベルの問題として出題されたものである。また、点双列相関係数が 0 である項目も、271 名の受験者全員が正答を選んだ、つまり正答率 100% の項目であったことから、受験者にとって難易度の低い問題で、受験者の識別には適していない問題であったことが分かった。次に示すのは、点双列相関係数が -0.011 であった【A1-3】と -0.052 であった【A2-10】の問題文、各選択肢とその選択率である。

【A1-3】

きょうは きん ようびです。

- ①全 (2.96%) ②全 (0%) ③*金 (97.05%) ④今 (0%) ⑤会 (0%)
(無回答 : 0%)

【A2-10】

いつも 3 じかんくらい 勉強 します。

- ①*べんきょう (99.26%) ②べんきょ (0.37%) ③べんきゅう (0%)
④へんきょう (0.37%) ⑤べきょう (0%)

（無回答：0%）

以上の2項目は、点双列相関係数がマイナスになったということで、正答を選べると想定される漢字能力が高い受験者が誤答を選び、そうでない受験者が正答を選びやすいということが起こっていることを意味している。しかし、問題文を見る限り、選択肢の構成等に問題があったとは考えにくい。【A1-3】は、正答である「③金」に字形が類似している文字を錯乱肢としている。錯乱肢には、漢字知識が豊富であると思われる受験者であってもなじみがないと思われる（常用漢字ではない）「②全」も含まれているが、この選択肢の選択率は0%であることから、未知の漢字の錯乱肢が異常に機能していたという可能性もない。一方、【A2-10】は「勉強」の読みを問うているが、長音・撥音・濁音の有無で錯乱肢が構成されており、構成自体に問題があるとは考えにくい。実際、この項目のように、読みの長短、清濁などから選択肢を構成することは、日本語能力試験の文字・語彙問題など漢字を多肢選択式の問題で問う場合でも一般的である。このように、点双列相関係数がマイナスになった2項目に問題そのものの不備は見当たらない。それでもこのような結果になった理由は、項目困難度が著しく高い、つまり極めて易しい項目であったためであると思われる。そのため、上位群あるいは中位群の受験者の中で数名、ケアレスミスなどで誤答を選択した場合などに大きく影響が出てしまうのだと推測される。

最後に、図2の散布図で多くの項目が困難度1.0に極めて近い値を示していた中で、特徴的であった項目を挙げる。【C2-38】は項目困難度が0.402と低く、識別力を表す点双列相関係数が0.382であった。この項目の設問文、選択肢と各選択肢の選択率を示す。

【C2-38】

ちかくの銀行で税金を おさめる。

- | | | |
|----------------|---------------|--------------|
| ①修める (4.43%) | ②収める (33.95%) | ③治める (2.58%) |
| ④*納める (40.22%) | ⑤押める (10.70%) | |

（無回答：8.12%）

次に示す表 11 は、【C2-38】について、受験者 271 名を漢字の総合得点で上位群（216 点以上；N=99）、中位群（197 点以上 215 点以下；N=90）、下位群（196 点以下；N=82）に分けた 3 グループの、それぞれの選択肢の選択率である。

表 11 【C2-38】各選択肢の選択率

選択肢	①	②	③	④*	⑤	無答
全体 (N=271)	12 4.43%	92 33.95%	7 2.58%	109 40.22%	29 10.70%	22 8.12%
上位群 (N=99)	1 1.01%	25 25.25%	0 0.00%	70 70.71%	3 3.03%	0 0.00%
中位群 (N=90)	3 3.33%	46 51.11%	3 3.33%	31 34.44%	6 6.67%	1 1.11%
下位群 (N=82)	8 9.76%	21 25.61%	4 4.88%	8 9.76%	20 24.39%	21 25.61%

この項目は、同訓異字を錯乱肢に 3 つ挙げている点が、他の項目と最も異なる点である。錯乱肢「①修める」「②収める」「③治める」は全て「おさめる」と読むことができるが、文脈に合う漢字ではないため誤答である。受験者 271 名の各選択肢の選択率を見ると、4 つの錯乱肢がどれも効果的に機能していることが分かる。この項目については、上位群であっても正答「④納める」を選択したのは 70.71% で、他の項目と比べて正答を選びにくかったことが分かる。表 11 を見ると、上位群にとって有効な錯乱肢となったのは「②収める」(25.25%) であったことが分かる。「おさめる」と読む漢字の中でも「納める」と比較的字義が似ている「②収める」が魅力的な選択肢であったのだと考えられる。中位群にとっても「②収める」(51.11%) が有効な選択肢として機能していたことは同様であるが、正答である「④納める」(34.44%) より「②収める」が選択されていた点が大きく異なる。一方、下位群は、上位群および中位群とは異なる解答行動をしていることが分かる。下位群は、正答「④納める」(9.76%) よりむしろ「②収める」(25.61%) と「⑤押める」(24.39%) を積極的に選択している。下位群に関しては無答の受験者が 25.61% であったことから分かるように、このグループの受験者には、この項目は難しく、正答であった「④納める」自体が未知であった可能性が高い。そこで、下位群の受験者は語頭を「お」と読むであろう「押

す」「押さえる」などで用いられる「⑤押める」を選択したと考えられる。

つまり、上位群と中位群の受験者は字義の情報により解答しようとしたのに対し、下位群の受験者は既知の漢字の読みを手掛かりに解答しようとしたと考えられる。いずれにせよ、異なる解答行動をする漢字能力の異なる受験者層が受験するブレースメントテストのような診断的評価の場では、このような選択肢構成は非常に有効であることが明らかになった。

6. 考察：テスト項目絞り込みの可能性

文法・漢字テストの信頼性は、双方とも高く、主たる受験者層の日本語能力がテスト作成当時と変化しているにもかかわらず、高い信頼性を維持していることが確認された。しかし項目数が多いほど信頼性係数が高く出る傾向がある上、テストが備えるべき条件の1つである実施容易性の観点から考えても、現行のブレースメントテストの項目数は、文法テスト（164項目）、漢字テスト（220項目）共に多すぎる。そのため、ブレースメント改訂の可能性の1つとして項目数を絞り込んでいくことも検討すべきであると考ええる。また、項目困難度が著しく高い、つまり易しい項目が多いこともブレースメントテストとしては望ましいとは言えない。そこで、現行のブレースメントテストの信頼性を維持した状態で項目を絞り込むことが可能か、という観点から考察したい。

文法・漢字ブレースメントテストの項目を絞り込んでいく上で参考になる情報は、各項目の項目困難度と識別力である。ブレースメントテストというテストの性質上、受験者にとって多様な困難度の項目を準備する必要があるため、項目困難度の情報は重要である。しかし、項目の良し悪しを項目困難度のみでは判断できず、識別力の情報も必要となる。そこで、項目が受験者の水準の違いを反映するのはどのような項目か、その傾向を知るために文法テストと漢字テストをそれぞれ識別力の高い順に並べ、順位をつけた（表12）。ここでは紙幅の都合により、文法テストと漢字テストの識別力を表す点双列相関係数上位20位と下位20位（文法テストは145－164位、漢字テストは201－220位）のみを挙げる。

表12を見ると、まず文法テストは、Aレベルの項目とBレベルの項目から構成されているが、点双列相関係数順の上位にはBレベルの項目が多く、下位に

表 12 文法・漢字プレースメントテストの点双列相関係数順位上位 20 位・下位 20 位

文法テスト(164項目)						漢字テスト(220項目)					
順位	項目番号			項目 困難度	点双列 相関係数	順位	項目番号			項目 困難度	点双列 相関係数
1	A	1	11	0.587	0.747	1	A	3	14	0.945	0.717
2	B	1	49	0.520	0.658	2	A	3	28	0.908	0.714
3	B	3	19	0.753	0.643	3	B	1	6	0.911	0.712
4	B	2	5	0.590	0.643	4	B	2	4	0.923	0.709
5	B	1	34	0.830	0.636	5	C	2	26	0.786	0.705
6	B	1	31	0.727	0.618	6	C	2	8	0.889	0.704
7	B	1	18	0.823	0.618	7	C	1	26	0.779	0.700
8	B	1	50	0.697	0.617	8	B	1	14	0.871	0.700
9	B	1	44	0.616	0.614	9	C	1	13	0.911	0.699
10	B	3	23	0.664	0.609	10	C	2	3	0.915	0.699
11	B	1	21	0.812	0.604	11	C	2	2	0.875	0.691
12	A	4	5	0.886	0.603	12	C	2	39	0.749	0.690
13	B	1	4	0.827	0.596	13	C	2	11	0.849	0.683
14	B	2	6	0.745	0.591	14	C	2	34	0.863	0.679
15	B	2	2	0.779	0.586	15	B	2	37	0.908	0.674
16	B	1	25	0.738	0.569	16	A	3	30	0.911	0.673
17	A	1	10	0.720	0.566	17	C	2	20	0.845	0.672
18	B	3	24	0.609	0.566	18	C	1	6	0.764	0.670
19	A	4	2	0.797	0.564	19	B	2	30	0.923	0.669
20	B	3	21	0.764	0.561	20	B	2	8	0.782	0.668
⋮						⋮					
145	A	2	15	0.967	0.168	201	A	2	3	0.989	0.200
146	B	2	14	0.775	0.167	202	A	2	13	0.911	0.194
147	A	1	3	0.985	0.164	203	B	2	3	0.989	0.187
148	A	1	9	0.989	0.152	204	A	2	30	0.745	0.184
149	A	2	3	0.989	0.145	205	A	1	20	0.996	0.179
150	A	1	1	0.970	0.121	206	A	3	8	0.952	0.177
151	A	5	9	0.996	0.118	207	A	3	6	0.989	0.165
152	A	5	3	0.996	0.118	208	A	2	22	0.996	0.153
153	A	5	6	0.996	0.108	209	A	1	4	0.993	0.148
154	A	2	13	0.934	0.108	210	A	1	8	0.996	0.125
155	A	3	1	0.941	0.098	211	A	1	1	0.996	0.082
156	A	2	4	0.993	0.075	212	A	1	7	1.000	0.000
157	A	2	9	0.993	0.056	213	A	1	9	1.000	0.000
158	A	5	4	0.985	0.039	214	A	1	10	1.000	0.000
159	A	1	6	0.996	0.033	215	A	2	20	1.000	0.000
160	A	5	5	0.963	0.032	216	A	3	1	1.000	0.000
161	A	2	8	0.982	0.022	217	A	3	2	1.000	0.000
162	A	2	1	1.000	0.000	218	A	3	17	1.000	0.000
163	A	5	1	1.000	0.000	219	A	1	3	0.970	-0.011
164	A	5	2	1.000	0.000	220	A	2	10	0.993	-0.052

は A レベルの項目が集中していることが分かる。つまり、難しい項目の方が易しい項目より識別力が高い項目が多いということになる。漢字テストも同様の傾向を示している。漢字テストの場合は、A レベル、B レベル、C レベルからなっているが、この中で最もレベルが高い C レベルの項目が、点双列相関係数順の上位に多く入っており、下位には A レベルの項目が多いことが分かる。もちろん、既に G-P 分析で見た通り、ターゲットとした文法項目や漢字のレベルが易しいものであったとしても、その問われ方や選択肢構成によって項目困難度は変化するが、概ね易しい項目は識別力が低くなりがちであると言える。そこで、仮に文法テストと漢字テストの易しい項目を削除し、項目数を大幅に減らした場合、今回実施したブレースメントテストと同様の信頼性を維持することができ、かつ受験者の能力を識別する項目がそろっているかを検討する。

まず、文法テストは、このテストの中で「難しい項目」とした B レベルの項目（94 項目）のみを受験者が受験したと想定した場合の、信頼性を表すクロンバックの α 係数と識別力を表す点双列相関係数を求めた。その結果、B レベルの項目 94 項目のクロンバックの α 係数は、 $\alpha = 0.962$ であった。A レベルと B レベルの計 164 項目を対象とした場合の α 係数は $\alpha = 0.967$ で、ほぼ同様の信頼性が確保されることが分かった。また B レベルの項目のみの点双列相関係数の平均も 0.462 であり、基準とされる 0.3 を超えることが分かった。つまり、文法テストは現行のブレースメントテストの B 項目のみを使用した場合でも、十分な信頼性と識別力があると言える。

次に、漢字テストは、このテストの中で「難しい項目」とした C レベルの項目（70 項目）のみを受験者が受験したと想定した場合の、信頼性を表すクロンバックの α 係数と識別力を表す点双列相関係数を求めた。その結果、C レベルの項目 70 項目のクロンバックの α 係数は、 $\alpha = 0.973$ であった。A レベルと B レベルも含めた計 220 項目を対象とした場合の α 係数は $\alpha = 0.983$ で、ほぼ同様の信頼性が確保されることが分かった。また C レベルの項目のみの点双列相関係数の平均も 0.574 であり、基準とされる 0.3 を超えることが分かった。つまり、漢字テストは現行のブレースメントテストの C 項目のみを使用した場合でも、十分な信頼性と識別力があると言える。

以上のように、文法テスト、漢字テストともに現行のプレースメントテストのうち難易度の高い項目を厳選することで、信頼性を確保しつつ、よりプレースメント受験者層の能力を識別できるテストへの改訂が可能であることが分かった。

7. まとめ

本研究の目的は、古典的テスト理論を援用して、現行の文法・漢字プレースメントテストを項目分析し、その有用性を検討することであった。具体的には、現行のテストが受験者層の能力を識別できる信頼性の高いテストであるのか、プレースメントテストをどのように改善することができるか、データを基に検討した。以下、本研究で明らかになった点である。

本研究で分析の対象としたプレースメントテストは、診断的評価に必要な、多様なレベルの項目を数多く出題すること、受験者のレベルを正確に判定することなどの条件を備えていることが分かった。文法テスト、漢字テストともに信頼性係数も高く、信頼性の高いプレースメントテストであると言える。各項目の識別力を表す点双列相関係数も基準とされることが多い0.3を超える項目が多く、受験者の水準の違いを反映することができるということが分かった。また、G-P分析を行ったところ、特に受験者層の中位群以上と下位群を識別するのに機能していることが分かった。さらに、項目の識別力が低い項目であっても、項目での問い方や選択肢構成に問題があるものは見当たらなかった。

しかし、このプレースメントテストに改善の必要がないわけではない。文法テストと漢字テストのそれぞれの項目困難度を算出したところ、受験者にとって著しく易しい項目が多数みられた。この傾向は、特に漢字テストに強く表れていた。このような傾向が見られた理由は、本研究で対象としたプレースメントテスト開発当時の受験者層と現在の受験者層の日本語能力が異なっているからであると考えられる。著しく易しい項目がプレースメントテストで使用されることは、項目として識別しないなどの原因になりうる。そこで今後は、項目困難度に問題があった項目については削除をするなどの改善も必要である。本稿では、その1つの方法として、このプレースメントテストで比較的難易度が高い項目群のみに厳選してプレースメントテストを実施することを提案した。もちろん、難易度が

高い項目でも識別力がない項目については選択肢構成を変更するなどの改善は必要である。

最後に、今後の課題を述べる。本研究では、古典的テスト理論に基づいて試験を分析した。しかし、古典的テスト理論による分析では、受験者集団が変わると分析結果に影響を及ぼしてしまうという問題点が指摘されている（野口・大隅、2012）。したがって、このブレースメントテストの受験者層の日本語レベルが今後大きく変わることがあれば、テストの信頼性係数や識別力なども異なってくる可能性があるため、再度分析が必要となる。また、本稿では、このブレースメントテスト結果が実際のレベル判定の結果をどの程度説明ができるのか、ということについては言及できなかった。それは、大阪大学日本語日本文化教育センターでの日本語能力のレベル判定には、本稿で取り上げた文法・漢字ブレースメントテストのみが使用されるわけではなく、他のパフォーマンステストの結果も合わせてレベル判定されるためである。今後は、文法・漢字テスト以外のテストについての信頼性と妥当性を分析した上で、レベル判定の結果を文法・漢字テストとその他のテストがどの程度説明（予測）できるのかさらに分析する必要がある。

参考文献

- 石田敏子（1992）『入門日本語テスト法』大修館書店
- 伊東祐郎（2008）『日本語教師のためのテスト作成マニュアル』アルク
- 今村和宏（2001）「ブレースメント・テスト改良のための統計分析」『一橋大学留学生センター紀要』4:19-37.
- 小森和子（2011）「ブレースメントテストのオンライン化の試みと問題項目の分析評価」『九州大学留学生センター紀要』19:89-106.
- 三枝令子（1986）「ブレースメント・テストの統計的処理の試み」『筑波大学留学生日本語教育センター日本語論集』2:
- 三枝令子（1987）「ブレースメント・テストの妥当性と今後の展望」『筑波大学留学生日本語教育センター日本語論集』3:
- 酒井たか子（1990）「ブレースメント・テスト 文字問題に関する一考察」『筑波大学留学生日本語教育センター日本語論集』6:
- 中村洋一（2002）『テストで言語能力は測れるか～言語テスト分析入門～』桐原書店

野口裕之・大隅敦子 (2012) 「テストニング・評価」 近藤安月子・小森和子編『研究社日本語教育事典』 pp.337-360. 研究社

渡部倫子「日本語プレースメントテストにおける聴解テスト項目の分析」『留学生教育』17:125-132.

J.D. ブラウン (1999) 『言語テストの基礎知識』 大修館書店

付記：本稿は、大阪大学日本語日本文化教育センターにおける評価基準の改定を検討する過程で行った分析の一部である。ワーキンググループのメンバーには貴重なコメントをいただいた。ここに記して感謝申し上げる。

Reliability and Item Discrimination of Grammar and Kanji Placement Test

Yuko YAMATO

The present study examined the reliability and the item classifying viability in a *CBP* test (a *Computer-Based Placement* test for evaluating grammar and kanji) in order to identify the points for further improvements. The *CBP* test played a major role in allocating foreign students to the Japanese Language classes accordingly. However, due to the paradigm shifts of screening process in language ability level, it is considered that a reconfirmation of the test must be needed. Thus, this study conducted an evaluation survey to confirm whether *CBP* test still can be fit to the new screening system. A total of 164 items (which contains 3 or 4 multiple choice answers) for grammar tests, and a total of 220 items (which contains 5 multiple choice answers) for kanji tests were examined. Both the grammar ($\alpha=0.967$) and the kanji tests ($\alpha=0.983$) showed a high reliability in the analysis. However, the item classification indicated that, the test questions in which the degree of difficulty is higher, possess high efficiency in classifying, with comparison to the questions with lower difficulty. In conclusion, this study re-confirmed that, the *CBP* test can still be used to determine the levels of Japanese language ability of the foreign students which in turn allocate them into classes such as beginner, intermediate or post-intermediate.