

2つの日本語プレースメント・テストの等質性の検証

吉川 達¹

A Study of Equality between Two Batteries of Japanese Placement Test.

Toru YOSHIKAWA

要 旨

本稿では、佐賀大学で行われたプレースメント・テストの別版の作成過程を説明し、それが現行版のプレースメント・テストの代替テストとなりうるか検証した結果を報告する。プレースメント・テスト別版は、旧日本語能力試験の過去問題を引用して作成した。使用する項目は、旧日本語能力試験の分析結果をもとに項目困難度、識別力の観点から基準を設けて抽出した。完成した別版プレースメント・テストと、構成を同じくする現行版プレースメント・テストのプレテストを行い、その結果を分析して両版の比較を行った。プレテストの結果、両版共に高い信頼性が得られ、さらに合計点においても非常に高い相関が得られたことから、プレースメント・テスト別版を現行版の代替テストと使用することができるという結論に至った。ただし、レベル判定については版によって若干のずれが生じたため、レベル判定規準の改善が示唆された。

【キーワード】プレースメント・テスト 古典的テスト理論 困難度 識別力 レベル判定

1. はじめに

佐賀大学では、交換留学生や研究生等外国人留学生を対象とした日本語授業を行っている（以下「日本語コース」）。日本語コース受講者は、最初にプレースメント・テストを受験してレベル判定された後に、各自のレベルに合った授業を受講する。プレースメント・テストは佐賀大学の事情に合わせて作られたものであるが、これまで1種類しかなかった。そのため、日本語コース受講生から再受験の希望があった場合も、対応できなかった。

そこで現行のプレースメント・テスト（以下「現行版」）を基にして、別版のプレースメント・テスト（以下「別版」）を作成することになった。本稿では、現行版・別版共通のプレースメント・テストの構成及び別版の作成過程を示し、現行版・別版のプレテストを行った結果を報告する。さらに、プレテストの分析結果から別版を現行版と同質の代替テストとして使用することができるか検討する。

1 全学教育機構

2. 佐賀大学「日本語コース」概要

プレースメント・テストは、それを実施する機関のカリキュラムやコースと密接に関係している。ここでは、佐賀大学の日本語コースについて概説する。

日本語コースは、初級前半レベルから上級レベルまで6レベルで構成される。各レベルの目安は、次の通りである。

図1. 佐賀大学日本語コースレベルの目安

上級	
↑	— 日本語が自由に使えるレベル
中上級	
↑	— 日本語能力試験N2合格レベル
中級	
↑	— 日本語能力試験N3合格レベル
初中級	
↑	— 日本語能力試験N4合格レベル
初級後半	
↑	— 日本語能力試験N5合格レベル
初級前半	

日本語コース受講者はプレースメント・テストでレベル判定され、各レベルで開講される授業から受講したい授業を選ぶ。留学身分によって開講される授業が異なるわけではなく、レベルによって受講できる授業が決まる。到達目標は各授業で設定されているが、上位レベルに行くほど大学での学術活動に必要な、いわゆるアカデミック・ジャパニーズを意識した授業内容になっている。

受講者は1学期6か月の授業が終わった後に当該レベルの到達基準に達していれば、翌学期には1段階上位レベルの授業を受講することができる。

プレースメント・テストの再受験は認めていないが、受講者の中には授業時間外に自習で日本語学習を進め、学期終わりに飛び級を希望する者もいる。また、学期の始めと学期の終わりで日本語能力が変化したのか、レベルチェックのためにプレースメント・テストの再受験を希望する受講者もいる。これまでは、このような受講者に対応できなかった。J-CAT等の大規模試験でプレースメント・テストを代用することも検討されたが、学内のレベル判定基準との相関がまとまった数でとれないことなどの理由から導入には至っていない。

3. プレースメント・テストの構成とレベル判定規準

日本語コースのレベルの目安に合わせて、プレースメント・テストは現行版、別版ともに表1に示す6つの問題群から構成される。

プレースメント・テストの作文を除いた客観テスト部分は、主に平成21年以前の旧日本語能力試験の過去問題を用いている。伊東（2005）は「日能試（日本語能力試験）をクラス配置の目的に利用するには、得点の解釈において、日本語プログラムの授業レベルと整合性をもたせなければならない」としながらも、日本語能力試験をプレースメント・テストに利用することの有用性を述べている。

客観テスト部分の解答は、全てマークシートの解答用紙に記入する。テスト後解答用紙をマークシートリーダーによって読み取った後、CSVファイルとして出力し、エクセルで処理する。解答用紙だけでなく、問題用紙も試験後全て回収される。

出力された客観テスト部分の得点は、問題による配点の重みづけはせず、正答数によって得点を計算する。算出された得点を基に、表2の基準に従って受験者を初級前半レベルから上級レベルまでに判定する。各レベルの基準に合わない条件が一つでもあった場合、1段階下位のレベルに判定される。

正答数によるレベル判定を行った後、Part6の作文を教員2、3名が読んで、レベルの調整を行う。さらに授業開始後に一定期間を設け、授業担当教員もしくは受講者からレベルが適当ではないと申し出があった場合は、プレースメント・テストの結果と受講者の状況を考え合わせてレベル変更を行う場合がある。しかし、レベル変更の申し出は毎学期1例程度である。

表1 プレースメント・テストの問題構成

項目群	分野	項目数	構成	時間
Part1	聴解	10問	旧日本語能力試験 4級絵あり1問・絵なし1問 3級絵あり2問・絵なし1問 2級絵あり2問・絵なし1問 1級絵あり1問・絵なし1問	約15分
Part2	文法	20問	4級文法	30分
Part3		20問	3級文法	
Part4		20問	2級文法	
Part5	読解	8問	2級短文単問形式 5問 1級短文単問形式 3問	25分
Part6	作文	1問	自作課題	20分
計 78問+作文1問				90分

表2 レベル判定規準

上 級	Part1-Part5	正答数71 (90%) 以上
	かつPart5 (読解)	6 以上
中 上 級	Part1-Part5	63 (80%) ~70
	かつPart4 (2級文法)	14 (70%) 以上
	かつPart6 (読解)	5 以上
中 級	Part1-Part5	47 (60%) ~62
	かつPart4 (2級文法)	10 (50%) 以上
	かつPart3 (3級文法)	12 (60%) 以上
	かつPart5 (読解)	3 以上
初 中 級	Part1-Part5	46以下
	かつPart3 (3級文法)	12 (60%) 以上
初級後半	Part1-Part5	46以下
	Part3 (3級文法)	11以下
	かつPart2 (4級文法)	12 (60%) 以上
初級前半	Part1-Part5	46以下
	Part2 (4級文法)	11以下

4. プレースメント・テスト別版の作成過程

別版を作成するにあたり、旧来の現行版も表1の構成に合わせる形で改定を行った。旧来の現行版も、以下に述べる別版の作成過程と基本的に同じ方法で作成している。現行版で使用しているのは、平成13年度から16年度の旧日本語能力試験の過去問題である。ただし読解問題8問中4問は自作問題である。改訂の際には、これまで使用していた項目を取捨選択し、平成13年度から16年度の日本語能力試験過去問題から引用した項目を加え、必要に応じて自作問題も加えた。

別版の作成においては、作文を除く、聴解、文法、読解分野の項目を平成19年度から平成21年度第2回目までの旧日本語能力試験の過去問題から引用した。

過去問題から項目を引用する際は、日本語能力試験実施委員会・日本語能力試験企画小委員会(2009、2010、2011)『日本語能力試験分析評価に関する報告書』を参照し、識別力が0.400以上で困難度が0.600程度の項目を中心に候補を選んだ。引用した項目の中から必要数を選別した時点で、同機関で働く他の日本語授業担当専任教員4名に項目のチェックを依頼し、指摘があった項目の入れ換えや、部分的な修正を行った。

項目を引用する際、困難度を0.600程度とした理由は、次の通りである。旧日本語能力試験は、「集団基準準拠テストの代表(伊東 2005)」である。中村(2002)は「集団基準準拠テストでは、項目困難度としては、0.500が最も理想」としており、さらに「当て推量に

よってそれぞれの選択肢が選ばれる確率も考慮に入れなければならないので、(中略) 選択肢が4つの場合、最適困難度は、 $0.5+0.5 \times 1/4=0.625$ となる」としている。このようなことから、困難度に関しては0.600を中心として項目を採用した。

また、識別力に関して旧日本語能力試験の識別力は、点双列相関係数が用いられている。点双列相関係数で識別力を示す場合、一般的に0.30以上が識別力の高い項目とされている(中村2002)。日本語能力試験においては、0.300以上の項目は多数存在するため、より条件を厳しくした0.400を採用の基準とした。

正答数をもとに結果分析を行う古典的テスト理論の特徴として、受験者が変われば得られる項目の特性も変わるという問題、つまり受験者依存の問題がある。その意味では、旧日本語能力試験の過去問題の分析数値を用いて項目を選定しても、別の受験者集団を対象に実施した場合は、全く異なる特性を示す可能性はある。さらに、日本語能力試験は級別に受験者が分かれており、一部の項目を除いてそれぞれの級で異なる問題を解く。その点は、全ての受験者が同一のテストを受験するというブレースメント・テストの性質と異なる。

しかし、旧日本語能力試験の分析結果は日本語大規模テストで唯一公表されている数値データであり、毎回数万から数十万という大規模な受験者が受験している。その中で数値的に良問と判断される項目は、項目の内容を見ても妥当性が高いと判断されるものが多い。公表されている数値の限界を認識しつつも、実用面では十分に利用する価値があるものとする。伊東(2011)や中村(2011)が言及している良質な項目の「再利用」という観点からも、分析結果が開示されている過去問題を利用することは意味がある。

以下、各分野の作成過程について説明する。

4-1 聴解分野

識別力が0.400以上で、困難度が0.500から0.700のものを候補として引用し、トピックや使用されている語彙等から適切と思われる項目を選別した。

表3に聴解分野で使用した問題を示す。表中の困難度、識別力は日本語能力試験実施委員会・日本語能力試験企画小委員会(2009、2010、2011)から引用した数値である。なお、年度の①は当該年度の第一回目の試験を、②は第二回目の試験を表す。

表3 聴解分野の構成

	級	年度	問番号	困難度	識別力
絵のある問題					
1	4級	H21②	I-5	0.620	0.422
2	3級	H20	I-10	0.637	0.497
3	3級	H21②	I-6	0.563	0.564
4	2級	H19	I-3	0.621	0.435
5	2級	H21②	I-6	0.575	0.423
6	1級	H20	I-11	0.653	0.478
絵のない問題					
7	4級	H20	II-3	0.561	0.561
8	3級	H20	II-1	0.563	0.564
9	2級	H20	II-7	0.561	0.453
10	1級	H19	II-13	0.601	0.563

4-2 文法分野

文法分野で使用対象とした項目は、旧日本語能力試験「読解・文法」分野の文法部分である。3級と4級は問題Ⅰ、問題Ⅱ、問題Ⅲの中から、2級は問題Ⅳと問題Ⅴの中から引用した。引用する際には以下の4つの段階を設けて項目を選んだ。

段階Ⅰ：識別力0.400以上 かつ 困難度0.550-0.650

段階Ⅱ：識別力0.390以上 かつ 困難度0.540-0.660

段階Ⅲ：識別力0.400以上で段階Ⅰ・Ⅱの範囲を除く困難度0.500台もしくは0.600台

段階Ⅳ：その他必要がある項目

上記の条件に合う項目を候補として挙げ、選択肢や質問対象の文法事項の重複などを考慮して各級20問を選別した。ただし、2級項目においては、1問だけ必要な問題を自作した。

使用した項目の一覧を表4から表6に示す。なお、表中の困難度、識別力は日本語能力試験実施委員会・日本語能力試験企画小委員会（2009、2010、2011）から引用した数値である。

4-3 読解分野

読解分野の結果は中級以上の判定に使用するため、1級と2級の項目から引用した。引用した項目は、1級2級ともに短文を読んで一つの質問に答える短文単問形式で、1、2級いずれも「読解・文法」分野の問題Ⅲの問題である。短文単問形式の問題のみ使用したのは、長文問題を使用した場合、万が一受験者によってトピックの有利・不利、得意・不

表4 4級文法分野の構成

	年度	問番号	困難度	識別力		年度	問番号	困難度	識別力
1	H19	6	0.624	0.465	11	H19	24	0.573	0.483
2	H21①	16	0.624	0.539	12	H19	25	0.550	0.466
3	H19	12	0.635	0.393	13	H20	17	0.546	0.536
4	H19	22	0.548	0.455	14	H20	20	0.594	0.510
5	H19	19	0.559	0.478	15	H20	24	0.555	0.395
6	H20	8	0.566	0.461	16	H20	26	0.593	0.428
7	H20	10	0.554	0.407	17	H21①	22	0.590	0.413
8	H21①	15	0.616	0.545	18	H19	16	0.643	0.448
9	H21①	23	0.644	0.534	19	H21①	24	0.592	0.474
10	H19	21	0.576	0.505	20	H21①	29	0.511	0.412

表5 3級文法分野の構成

	年度	問番号	困難度	識別力		年度	問番号	困難度	識別力
1	H19	3	0.545	0.444	11	H20	2	0.582	0.518
2	H19	12	0.549	0.409	12	H20	7	0.576	0.575
3	H19	15	0.567	0.481	13	H20	9	0.536	0.535
4	H19	16	0.535	0.508	14	H20	16	0.636	0.555
5	H19	21	0.649	0.427	15	H20	19	0.536	0.559
6	H19	25	0.550	0.490	16	H20	28	0.696	0.528
7	H19	29	0.617	0.400	17	H21①	10	0.610	0.442
8	H19	33	0.667	0.478	18	H21①	19	0.642	0.455
9	H19	34	0.636	0.573	19	H21①	32	0.559	0.572
10	H19	36	0.601	0.398	20	H21①	37	0.615	0.478

表6 2級文法分野の構成

	年度	問番号	困難度	識別力		年度	問番号	困難度	識別力
1	—	—	—	—	11	H20	41	0.559	0.457
2	H19	51	0.271	0.118	12	H20	42	0.546	0.457
3	H20	26	0.572	0.471	13	H20	43	0.585	0.483
4	H20	27	0.615	0.405	14	H20	49	0.605	0.477
5	H20	32	0.679	0.436	15	H20	50	0.576	0.432
6	H21①	52	0.585	0.443	16	H20	53	0.528	0.453
7	H20	34	0.543	0.423	17	H21①	37	0.654	0.444
8	H20	35	0.582	0.404	18	H21②	45	0.585	0.431
9	H20	36	0.694	0.409	19	H20	33	0.613	0.395
10	H20	38	0.588	0.431	20	H21②	46	0.565	0.468

得意があった場合には、その長文に関する全ての項目に影響を及ぼすという理由からである。さらにテストの信頼性の観点から、限られた紙幅、時間の中で少しでも多くの項目数を確保しようとするなら、短文単問形式の方が適当であると判断した。それでも8問という項目数は充分ではないが、中級以上の判定のためという限られた目的であるので、この項目数は最低限の数を確保していると考ええる。

以下、使用した項目の一覧を示す。項目を選定するにあたっては、0.400以上という識別力の条件を優先させ、その中から困難度が0.400から0.750の範囲内にあるものを選んだ。なお、表中の困難度、識別力は日本語能力試験実施委員会・日本語能力試験企画小委員会(2009、2010、2011)から引用した数値である。

表7 読解分野の構成

	年度	問番号	困難度	識別力
2級項目				
1	H19	2	0.527	0.472
2	H19	3	0.439	0.491
3	H19	4	0.450	0.459
4	H20	2	0.600	0.412
5	H21①	3	0.431	0.412
1級項目				
6	H20	3	0.735	0.448
7	H21①	3	0.555	0.413
8	H21②	3	0.733	0.421

5. プレテストの実施と結果

2014年1月の2日間、本学留学生33名を対象に現行版・別版のプレテストを行った。プレテストに参加した学生は、4か月前もしくは10か月前に一度プレースメント・テストを受験しているが、今回別版を作成するにあたり現行版も改定を行っているため、リハーサル効果の影響は小さいと考える。

プレテスト実施時には、解けない問題は解かず、空白にしておくよう受験者に注意をした。

実施したプレテストの結果の基本統計量を表8、表9に示す。

全ての分野において両版で概ね近い数値が出ているが、平均値に関しては、特に近い値を示した。合計点の平均値は0.18の差であるが、この差が有意であるか t 検定を行ったところ、有意差はなかった($t(32)=0.828$, ns)。

表8 現行版のプレテスト結果 (n=33)

	聴解 (10)	4級文法 (20)	3級文法 (20)	2級文法 (20)	読解 (8)	合計 (78)
平均	8.61	17.85	14.27	11.58	5.30	57.61
標準偏差	1.64	2.60	4.69	7.41	2.17	16.26
最大値	10	20	20	20	8	77
最小値	5	11	3	0	0	25

表9 別版のプレテスト結果 (n=33)

	聴解 (10)	4級文法 (20)	3級文法 (20)	2級文法 (20)	読解 (8)	合計 (78)
平均	8.09	17.39	16.03	11.15	5.12	57.79
標準偏差	1.49	3.14	5.11	6.61	2.69	17.48
最大値	10	20	20	20	8	78
最小値	4	9	4	0	0	20

次に各テストの信頼性係数として、クロンバックの α 係数を算出した。 α 係数は「一般のテストでは0.8以上の値を示すことが望ましい(日本語能力試験実施委員会・日本語能力試験企画小委員会 2009:41)」とされている。両版共に非常に高い値となっていることから、高い信頼性が確保されていると言える。

表10 現行版・別版の α 係数

	現行版	別版
α 係数	0.962	0.969

続いて各分野の現行版と別版のピアソンの積率相関、及び散布図を示す。読解に関しては、両版共に無回答の受験者が1名いたため、それを省いた数値を示す。

表11 現行版と別版の各分野の相関

聴解	文法4級	文法3級	文法2級	読解	合計
.849	.899	.853	.940	.714	.962

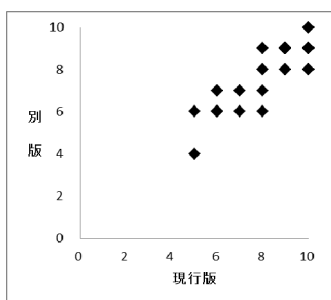


図2 聴解散布図

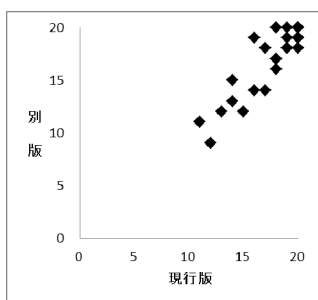


図3 4級文法散布図

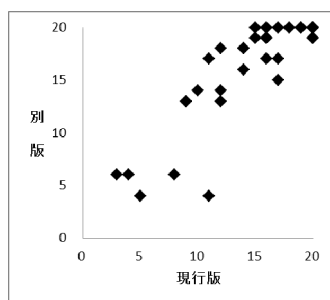


図4 3級文法散布図

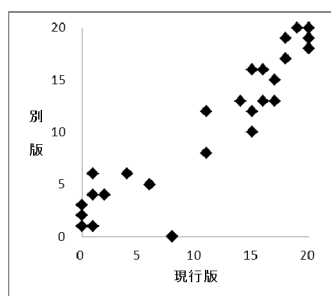


図5 2級文法散布図

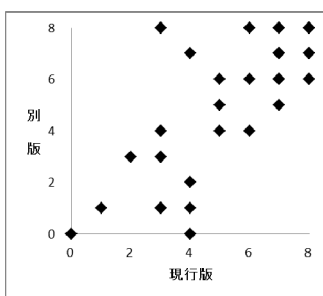


図6 読解散布図

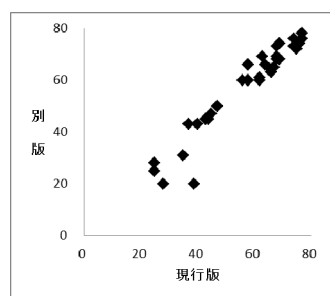


図7 合計散布図

それぞれの分野の中で、読解の相関が最も低い。しかし、8問という多くはない問題数で $r=.700$ 以上の値を示しているのは、注目すべき点である。吉川（2010）は、日本留学試験の読解分野について模擬試験の結果と本試験の結果の相関を求めているが、受験者が100名以上で問題数が20問であるにもかかわらず、7回あった模擬試験で最大でも $r=.460$ の相関しか得られていない。その例から見ても、本テストの読解の相関が強いことがわかる。

合計の散布図を見ると、得点下位層の分布にばらつきが見られる。中・上位層に比べると下位層においてはテスト毎に得点が揺れる可能性があると考えられる。本テストの解答形式が四肢選択であるため、実力がない受験者が確信のない状態で答える、または当て推量で答えることは想像に難くない。

6. 項目分析

プレテストで得られた結果を中村（2002）のテスト処理ソフト「TDAP」を使用して処理し、古典的テスト理論による項目分析を行った。TDAPでは、項目応答理論による1パラメータ・ラッシュ・モデルでの分析を行うことも可能であるが、村上（2013）は項目応答理論による分析は非常に大規模なデータが必要になるとし、「クラスルーム・テストのような場合、項目応答理論によるソフトを使っても正答率、識別力以上の情報はほとんど得られない」としている。また、実際に吉川（2011）は、111名分のテスト結果を1パラメータ・ラッシュ・モデルで処理しているが、正答数による処理と差のない結果となったとし

ている。本稿のプレテストは受験者数が33であり、項目応答理論による分析は適さないことから、正答数をもとにした古典的テスト理論による分析を行う。

別版を作成する際には、日本語能力試験の困難度と識別力を基準に項目の抽出を行った。プレテストの結果においても、困難度と識別力を中心に項目の分析を行う。なお、表12から表16で示す識別力は、旧日本語能力試験と同様に点双列相関係数によるものである。

聴解は項目1から6が絵のある問題で、7から10が絵のない問題である。また、1と7が4級、2、3、8が3級、4、5、9が2級、6と10が1級の項目である。絵のある問題も、絵のない問題も概ね級が上がるごとに困難度も下がっていく傾向がみられる。正常にレベル分けに機能していると言えるであろう。

表12 聴解分析			表13 読解分析			表14 4級文法分析					
困難度	識別力		困難度	識別力		困難度	識別力		困難度	識別力	
1	0.939	0.477	1	0.909	0.449	1	0.879	0.432	11	0.970	0.152
2	0.909	0.266	2	0.545	0.487	2	0.970	0.152	12	0.818	0.821
3	0.848	0.348	3	0.727	0.684	3	0.788	0.398	13	0.970	0.275
4	0.788	0.536	4	0.758	0.597	4	0.727	0.439	14	1.000	0.000
5	0.848	0.064	5	0.667	0.817	5	0.848	0.618	15	0.909	0.645
6	0.576	0.406	6	0.545	0.696	6	0.667	0.238	16	0.848	0.638
7	1.000	0.000	7	0.455	0.538	7	0.758	0.733	17	0.848	0.496
8	0.848	0.564	8	0.515	0.661	8	0.909	0.664	18	0.909	0.278
9	0.848	0.334				9	1.000	0.000	19	0.909	0.615
10	0.485	0.523				10	0.848	0.393	20	0.818	0.332

表15 3級文法分析						表16 2級文法分析					
困難度	識別力		困難度	識別力		困難度	識別力		困難度	識別力	
1	0.727	0.586	11	0.848	0.108	1	0.667	0.529	11	0.576	0.791
2	0.939	0.558	12	0.636	0.521	2	0.545	0.413	12	0.636	0.554
3	0.697	0.555	13	0.727	0.862	3	0.576	0.713	13	0.515	0.672
4	0.788	0.545	14	0.848	0.741	4	0.545	0.512	14	0.455	0.683
5	0.788	0.467	15	0.818	0.702	5	0.697	0.544	15	0.424	0.442
6	0.788	0.765	16	0.909	0.333	6	0.545	0.555	16	0.515	0.574
7	0.818	0.821	17	0.879	0.594	7	0.545	0.586	17	0.606	0.841
8	0.939	0.462	18	0.788	0.829	8	0.333	0.244	18	0.576	0.823
9	0.788	0.752	19	0.788	0.812	9	0.576	0.716	19	0.636	0.866
10	0.818	0.140	20	0.697	0.735	10	0.545	0.491	20	0.636	0.803

読解は1から5が2級の項目で、6から8が1級の項目である。全ての項目で高い識別力を示しており、また1級項目のほうが困難度が低く、期待した通りの結果となっている。項目がレベル分けに機能していると言えるであろう。ただし、項目2に関しては、困難度が他の項目よりも低い。受験者の選択した選択肢を見ると、正答以外の選択肢すべてに誤答者の選択が分散している。実質選択肢数も3.165を示しており、迷わしの選択肢が機能した結果困難度が低くなったと考えられる。

文法に関して、各項目の個別の分析はここでは行わないが、級が上がるにつれて、困難度も下がる傾向にある。また、困難度が非常に高い、言い換えれば正答率が高い4級項目では、識別力はうまく発揮されていないが、2級項目においては、ほぼすべての項目で高い識別力を示している。これらを勘案すると、文法においてもレベル分けがうまく機能していると言えるであろう。

7. 現行版・別版のレベル判定結果

プレテストの結果を表2の基準でレベル分けを行った場合、両版とも同じようにレベル分けができるかという点が重要である。現行版、別版それぞれの結果をもとに全ての受験者のレベル判定を行った結果を以下に示す。なお、表中では初級前半レベルを1、初級後半レベルを2、初中級レベルを3、中級レベルを4、中上級レベルを5、上級レベルを6で示す。

33名のうち11名は現行版と別版で判定レベルに違いが生じた。2レベル以上異なる判定が出た受験者はいない。11名の判定結果が異なるが、いずれの受験者もいずれかの分野において1～3点の差で異なるレベルに判定されている。

いくつもの不確定要素があり、完全に判定を一致させることは現実には難しいであろう。誤差ともいえるこの差の影響が少なくなるよう、今回分析対象としなかった作文問題を含めた最終的な評価を適切に行う必要がある。

表17 現行版・別版レベル判定結果

受験者	現行版	別版	受験者	現行版	別版	受験者	現行版	別版
A	4	4	L	2	2	W	6	6
B	<u>3</u>	<u>4</u>	M	1	1	X	<u>5</u>	<u>4</u>
C	4	4	N	<u>2</u>	<u>1</u>	Y	6	6
D	3	3	O	2	2	Z	6	6
E	<u>3</u>	<u>4</u>	P	2	2	AA	5	5
F	3	3	Q	<u>2</u>	<u>3</u>	AB	6	6
G	3	3	R	5	5	AC	<u>5</u>	<u>6</u>
H	3	3	S	<u>5</u>	<u>4</u>	AD	4	4
I	<u>4</u>	<u>3</u>	T	5	5	AE	6	6
J	<u>2</u>	<u>3</u>	U	5	5	AF	6	6
K	<u>5</u>	<u>6</u>	V	3	3	AG	<u>5</u>	<u>6</u>

8. まとめと今後の課題

これまで示したプレテストの結果をまとめると、以下の3つの点から、プレースメント・テスト別版を現行版の代替テストとして用いることができるということが出来る。

- ① 現行版と別版の相関が非常に高い
- ② 両版共に信頼性が高い
- ③ 級別の項目が困難度、識別力の点からレベル分けに機能している

村上(1991)は、二つのテスト得点の等価について述べ、「ここでもし、二つの試験が完全に同じ能力を測っているとすれば、二つの総点間の相関は、二つのテスト信頼係数の幾何平均である0.962に等しくなるはずである(p.212)」としている。本稿の現行版と別版の相関は $r = 0.962$ であり、両者が非常に近い能力を測定していると言えるであろう。

また、両版ともに非常に信頼性が高く、再テストを行った場合も同じ結果が得られるであろう。

さらに、それぞれの項目の特性から見ても、級が上がれば困難度も下がっており、正常にレベル分けに機能していると言える。

このような結果から、プレースメント・テスト別版を現行版に代えて使用することに問題ないと言えるであろう。

ただし、テスト結果からレベルを判定する際の基準は、さらに精度を高める必要がある。二つのプレースメント・テストを同質に扱うことができたとしても、その結果得られるレベルの判定結果に齟齬が生じれば、プレースメント・テストとしての意味をなさない。レベル判定の基準を再検討し、シミュレーションを重ねる、また詳細なルーブリックを作成し客観テスト部分で生じた誤差を作文の判定で修正するなど対策を考える必要がある。さらに、大学生のプレースメントを目的に開発されたJ-CAT(今井他 2010)と本稿で扱ったプレースメント・テストとの関係は、今後検討していく必要がある。テスト実施者の負担軽減などCAT使用による恩恵は大きいからである。これらを今後の課題としたい。

謝辞

本稿で扱ったプレースメント・テスト別版の作成及びプレテストは、平成25年度佐賀大学教育改善支援経費の助成により行われました。

参考文献

- (1) 伊東祐郎(2005)「プレースメント・テストの妥当性確認の試み」『東京外国語大学留学生日本語教育センター論集』第31号、161-174.
- (2) 伊東祐郎(2011)「項目バンクによって広がるテスト開発の可能性」『日本語教育』148号、57-71.

- (3) 今井新悟・伊東祐郎・中村洋一・菊池賢一・赤木彌生・中園博美・本田明子 (2010) 『J-CAT日本語能力をコンピュータで測る』 山口大学留学生センター
- (4) 中村洋一 (2002) 『テストで言語能力は測れるか』 桐原書店
- (5) 中村洋一 (2011) 「コンピュータ適応型テストの可能性」『日本語教育』148号、72-83.
- (6) 日本国際教育支援協会・国際交流基金編著 (2008) 『平成19年度日本語能力試験試験問題と正解』 凡人社
- (7) 日本国際教育支援協会・国際交流基金編著 (2009) 『平成20年度日本語能力試験試験問題と正解』 凡人社
- (8) 日本国際教育支援協会・国際交流基金編著 (2009) 『平成21年度第1回日本語能力試験試験1・2級問題と正解』 凡人社
- (9) 日本国際教育支援協会・国際交流基金編著 (2010) 『平成21年度第2回日本語能力試験試験問題と正解』 凡人社
- (10) 日本語能力試験実施委員会・日本語能力試験企画小委員会監修 (2009) 『平成19年度日本語能力試験分析評価に関する報告書』 アスク出版
- (11) 日本語能力試験実施委員会・日本語能力試験企画小委員会監修 (2010) 『平成20年度日本語能力試験分析評価に関する報告書』 アスク出版
- (12) 日本語能力試験実施委員会・日本語能力試験企画小委員会監修 (2011) 『平成21年度日本語能力試験(第1回・第2回)分析評価に関する報告書』 アスク出版
- (13) 村上京子 (2013) 「項目分析と改善」 関正昭・平高史也 (編) 『日本語教育叢書「つくる」テストを作る』 第3章2節、スリーエーネットワーク、pp. 199-209.
- (14) 村上隆 (1991) 「良いテストはどのような性質をもつか」、「大規模なテストデータの分析の一例」 日本語教育学会 (編) 『日本語テストハンドブック』 第2章、第5章、大修館書店、pp. 10-114、pp. 182-220.
- (15) 吉川達 (2010) 「日本留学試験の模擬試験を利用した本試験結果予測の可能性—古典的テスト理論を用いた分析—」 『佐賀大学留学生センター紀要』 第10号、55-66.
- (16) 吉川達 (2011) 「項目応答理論を使った日本留学試験の本試験結果予測の可能性」 『佐賀大学留学生センター紀要』 第11号、25-36.