

日本語テストの選択枝数の変更が 項目の正答率と識別力に与える影響

渡邊百里（新宿日本語学校）

要 旨

昨今、増加しているオンラインテストでは多枝選択問題が主流で、4枝選択問題が多く採用されている。日本語テストの研究において、多枝選択問題の選択枝の数の変更が項目の正答率や識別力に与える影響について言及しているものは見当たらない。そこで、本研究では、選択枝の数を4枝から3枝にするという変更が項目の正答率と識別力にどのような影響を与えるかをA日本語学校のオンラインプレースメントテストを用いて検証し考察する。4枝から3枝にした結果、正答率はやや上がる傾向、また識別力はやや下がる傾向が見られた。しかし、4枝選択と3枝選択の正答率や識別力の差の数値からみると、選択枝の数の変更が、正答率や識別力に大きい影響を与えるとは言い難い。さらに、選択枝の数と解答の正誤の間に関連性があるかカイ二乗検定を行った結果、10項目中8項目は、選択枝が4枝か3枝かということは正誤と関連性がないことがわかった。

【キーワード】 多枝選択問題 選択枝数 日本語オンラインプレースメントテスト
項目分析

1. 背景

さまざまな分野で増えているオンラインテストでは多枝¹⁾選択問題が主流となっている。また、日本語能力試験や日本留学試験等の日本語テストにおいても4枝選択問題が多く採用されている。テストの作問者として、選択枝を作成する際に、正答以外に、受験者を惑わせる3つの誤答を考えるのに時間と労力がかなりかかった経験がある。多枝選択問題の選択枝の最適な数はいくつなのだろうか。そのような疑問から本研究が始まった。

2. 先行研究

多枝選択問題における最適な選択枝数を探る研究は、テスト分析の分野では長く行われており、Rodriguez (2005) によって、まとめられている。Rodriguez (2005) は、同じ試験時間の場合、3枝選択の方が、4枝選択や5枝選択よりも、より多くの項目が実施でき、項目の精度が悪くなることはないとして述べており、3枝選択を支持している。

英語テストの研究では、Shizuka, Takeuchi, Yashima & Yoshizawa (2006) が、大学入試の英語読解試験において、項目ごとの選択枝数を4枝選択から3枝選択に削減した際のテスト得点への影響について調査している。その結果、4枝選択の中でほとんど選ばれていない選択枝を削除し3枝選択にしても、正答率や識別力にはほとんど影響がないことが明らかとなったとしている。また、言語テストの分野ではないが、Sidick, Barrett & Doverspike (1994) は、ある大きな自治体の新人警察官を対象とした読解力、レポート作成能力、論理的推論能力の測定において、選択枝数を5枝選択から3枝選択に削除し、両テストの結果を比較している。その結果、5枝から選択率の低い選択枝を削除し3枝にしても、2つのテストの平均値の差に大きな違いがないことを実証している。さらに、Delgado & Prieto (1998) は、心理学研究法入門を履修している大学1年生に対して、通常授業の評価試験において、選択枝の数が及ぼす得点への影響を検証している。その結果、4枝選択枝の中で選択率の低い選択枝を削除し、3枝選択枝に変更しても、正答率がわずかに上がるものの、識別力が低くなることはなかったと述べている。これらの研究の共通点は、選択枝の数を減らす方法として、過去のテスト結果において、選択率の低かった選択枝を削除している点である。そして、いずれの研究においても、選択枝の数を減らしても、テスト結果に大きな影響がないという結果を得ており、開発コストや作業時間の削減を考えると、テストの目的によっては、3枝選択の方が望ましいと主張している。一方、Haladyna & Downing (1993) は、3つの異なる分野のテスト結果を使用し、効果的な錯乱枝の数を探る研究を行っている。その結果、4つまたは5つの選択枝の中で、実際に機能している錯乱枝の数は、2つ以下であるという実験結果が得られたため、2つから3つの選択枝を作成することを勧めている。

これらの先行研究では、効率的な選択枝の数は3つ、あるいは2つから3つであると述べている。しかし、日本語テストの研究において、多枝選択問題の選択枝の数の変更が項目の正答率や識別力に与える影響について言及しているものは、現在のところ筆者の知る限り見当たらない。ゆえに、日本語テストにおいても同様の結果が得られるか検証することは、有用であると思われる。そこで、本研究では、選択枝の数を4枝から3枝にするという変更が項目の正答率と識別力にどのような影響を与えるかを検証し考察する。

3. 研究方法

3.1 オンラインプレースメントテストの概要

本研究では、A日本語学校において2020年10月、2021年4月に実施したオンラインプレースメントテスト（以下、オンラインPTとする）を用いて実験を行った。表1は、「オンラインPTの問題構成（2020年10月実施）」を示したものである。オンラ

インPTは全120項目で配点は各1点、問題形式は、2つから5つの選択枝から正答を1つ選ぶ多枝選択式である。項目の内容は、A日本語学校の初級Ⅰ、Ⅱ、中級Ⅰ、Ⅱの4つのレベルで学習する文法や語彙が理解できるかを問うものである。使用しているプラットフォームには時間制限機能がないため、受験者には、オンラインPTの受験案内の際に、60分以内に解答し提出するように伝えている。

表1 オンラインPTの問題構成（2020年10月実施）

レベル	問題	内容	形式	項目数
初級Ⅰ	1	助詞	4枝選択	10
	2	文法、接続詞、表現	4枝選択	10
	3	い形容詞、な形容詞	4枝選択	5
	4	動詞	4枝選択	10
			小計	
初級Ⅱ	1	文法	4枝選択	10
	2	補助動詞	4枝選択	5
	3	授受補助動詞	4枝選択	5
	4	動詞	4枝選択	5
			小計	
中級Ⅰ	1	副詞、接続詞	4枝選択	4
	2	機能語	5枝選択	5
	3	接続（動詞、形容詞）	4枝選択	10
	4	動詞	4枝選択	5
	5	文完成	2枝選択	5
	6	読解	3枝選択	1
			小計	
中級Ⅱ	1	動詞	4枝選択	10
	2	副詞、接続詞、機能語	4枝選択	10
	3	応答	3枝選択	5
	4	文完成	2枝選択	4
	5	読解	3枝選択	1
			小計	
合計				120

3.2 手順

本節では、研究の手順を述べる。

- ① 2020年10月に実施したオンラインPTの結果の4枝選択の項目において選択率が5%以下であった誤答選択枝を削除し、4枝から3枝に変更する。表2は、4枝選択から3枝選択に変更した10項目（項目002、017、018、019、023、024、027、028、089、097）の問題文と選択枝を示したものである。例えば、項目002は、「コーヒー（ ）おねがいます。」という問題文の（ ）に入るのに適切なものを4つの選択枝「が」、「を」、「は」、「と」の中から1つ選ぶというものである。選択枝「と」の選択率が1.71%であったため、削除対象とし、表では下線を施している。また、項目017は、「このりょうりはとても（ ）そうですね。」という問題文の（ ）に入るのに適切なものを、4つの選択枝「おいしく」、「おいし」、「おいしい」、「おいしいです」の中から1つ選ぶというもので、選択枝「おいしいです」の選択率が2.16%であったため、削除対象とし、表では下線を施している。これらの例のように、削除する選択枝を決定し、3枝選択に変更する。対象となったのは10項目であった。
 - ② 3枝選択に変更した項目を含むオンラインPTを2021年4月に実施し、テスト結果の項目分析を行う。項目分析については、各項目の正答率と識別力を求める。識別力は、各項目（4枝選択から3枝選択に変更した10項目）と2020年10月と2021年4月の両テストに共通の100項目の合計点との間の相関を計算する。本研究における項目分析結果の数値の解釈について、正答率は、0.25以下を「難しすぎる項目」、0.85以上を「やさしすぎる項目」とする。また、識別力は、0.3未満を「識別力が低い項目」とし、これらの数値の項目は、今後、項目改訂の検討を行う必要があると判断する。
 - ③ 選択枝の数の変更による影響を検討する前に、2020年10月と2021年4月の両テストに共通の項目（100項目）の得点について、平均値の差の検定を行う。各テストの120項目のうち、10項目は4枝選択を3枝選択に変更した項目、100項目は何も変更をしなかった共通項目、残る10項目はその他の変更を加えた項目となっている。
 - ④ 選択枝数変更の影響を検討するため、4枝選択と3枝選択の項目の正答率の比較を行う。
 - ⑤ 選択枝数変更の影響を検討するため、4枝選択と3枝選択の項目の識別力の比較を行う。
 - ⑥ 選択枝数変更の影響を検討するため、選択枝の数（3枝か4枝か）と、結果（項目に正答したか誤答したか）との間に関連性があるかを検討する。
- 以上の①から⑥の手順で研究を行う。

表2 4枝選択から3枝選択に変更した10項目の問題文と選択枝

項目002				
問題文	コーヒー（ ）おねがいます。			
4枝	が(6.41%)	を*(80.77%)	は(11.11%)	と(1.71%)
3枝	が(8.70%)	を*(79.71%)	は(11.59%)	
項目017				
問題文	このりょうりは とても（ ）そうですね。			
4枝	おいしく(4.76%)	おいし*(51.52%)	おいしい(41.56%)	おいしいです(2.16%)
3枝	おいしく(5.84%)	おいし*(45.99%)	おいしい(48.18%)	
項目018				
問題文	ひらがなを べんきょうします。()、カタカナを べんきょうします。			
4枝	それから*(73.89%)	それとも(18.14%)	だから(5.31%)	しかし(2.65%)
3枝	それから*(79.41%)	それとも(14.71%)	だから(5.88%)	
項目019				
問題文	()スポーツが すきですか。			
4枝	どこ(4.82%)	なに(18.86%)	いつ(2.19%)	どんな*(74.12%)
3枝	どこ(2.92%)	なに(12.41%)	どんな*(84.67%)	
項目023				
問題文	もうすこし（ ）あるいてください。			
4枝	しずか(12.00%)	しずかに*(68.00%)	しずかで(16.44%)	しずかだ(3.56%)
3枝	しずか(12.59%)	しずかに*(68.89%)	しずかで(18.52%)	
項目024				
問題文	()あかるいへやですね。			
4枝	ひろ(3.57%)	ひろい(16.07%)	ひろいで(14.29%)	ひろくて*(66.07%)
3枝	ひろい(18.94%)	ひろいで(12.12%)	ひろくて*(68.94%)	
項目027				
問題文	みずを（ ）たいんです。			
4枝	のみ*(75.55%)	のむ(13.54%)	のんで(9.61%)	のみて(1.31%)
3枝	のみ*(81.02%)	のむ(13.87%)	のんで(5.11%)	
項目028				
問題文	うるさいから、テレビを（ ）ください。			
4枝	けし(5.75%)	けす(3.98%)	けすて(9.29%)	けして*(80.97%)
3枝	けし(3.70%)	けすて(6.67%)	けして*(89.63%)	
項目089				
問題文	新人のころは、上司にいろいろなことを（ ）。			
4枝	させられました*(45.81%)	させました(27.37%)	しました(22.91%)	させよう(3.91%)
3枝	させられました*(56.57%)	させました(21.21%)	しました(22.22%)	
項目097				
問題文	今度大きな地震が起きたら、この建物は（ ）かねない。			
4枝	くずれる(27.65%)	くずれて(39.41%)	くずれ*(30.00%)	くずれます(2.94%)
3枝	くずれる(25.00%)	くずれて(38.04%)	くずれ*(36.96%)	

※選択枝の横の（ ）は、選択率、「*」は正答、下線は削除対象となった選択枝を示す。

4. 結果

4.1 基本統計量

2回のテスト結果の基本統計量を表3にまとめた。2020年10月は2021年4月と比べて受験者数が100名ほど多く、平均値、中央値は若干低いという結果であった。一方、標準偏差、最高値、最低値、信頼性係数の推定値については、大きな差は見られなかった。信頼性係数の推定値はクロンバックの α 係数を用いて算出している。日本語教育では、 α 係数は0.850あれば信頼できると言われている（村上 2013 p.212）が、本研究では、各回とも0.900以上の数値になっており、極めて信頼性の高いテストであることが確認できた。

表3 オンラインPTの基本統計量 (N=120)

	受験者数	平均値	標準偏差	最高点	最低点	中央値	α 係数
2020年10月	235	52.94	27.71	116	2	48.0	0.974
2021年4月	138	55.14	27.87	120	3	54.0	0.975

※表のタイトル横のNは項目数、信頼性係数の推定値は、クロンバック α 係数

4.2 2020年10月と2021年4月の受験者の日本語能力の差の確認

選択枝の数の変更による影響を検討する前に、2020年10月と2021年4月の両テストに共通の項目（100項目）の得点について、平均値の差の検定を行った。表4は、両テストに共通の100項目の受験者数、平均値、標準偏差をまとめたものである。

表4 2020年10月と2021年4月の共通100項目の基本統計量

実施時期	受験者数	平均値	標準偏差
2020年10月	235	43.9	23.6
2021年4月	138	45.6	23.4

共通項目における2020年10月受験者と2021年4月受験者の間に、得点の平均値に差があるかを検証するため対応のない t 検定を行った。その結果、両テストの受験者の共通項目得点の平均値の差は有意ではなかった ($t(289) = 0.649, p > .05$)。これにより、2020年10月受験者と2021年4月受験者間で能力水準に違いが見られないと考えられる。

4.3 選択枝の数の変更による正答率への影響

2020年10月受験者と2021年4月受験者の間に能力の差がないことが確認できたので、4枝選択と3枝選択の項目の正答率の比較を行う。表5は、4枝選択から3枝選択へ変更した10項目の正答率の比較をまとめたものである。また、図1は、4枝選択から3枝選択に変更した10項目の正答率と識別力の変化を視覚化したものである。4枝選択を3枝選択にした結果、正答率が上がったものは8項目（項目018、019、023、024、027、028、089、097）であった。その差をみると、8項目中6項目は、0.007から0.072の間であった。しかし、8項目中2項目については、項目019が0.122の差、項目028が0.098の差となっており、0.100、つまり10%程度の差があった。また、項目019は、3枝にすることで、正答率が0.841、項目028は0.877となり、やさしすぎる項目となった。一方、4枝選択を3枝選択にした結果、正答率が下がったものは2項目（項目002、017）で、両項目ともその差は、0.050以下であった。これらの数値から、4枝選択を3枝選択にすることで、正答率がやや上がる傾向がみられ、中には10%程度上がる項目もあった。

4.4 選択枝の数の変更による識別力への影響

次に、選択枝の数の変更による識別力への影響について述べる。4枝選択と3枝選択の項目の識別力の比較を行った。表6は、4枝選択から3枝選択へ変更した10項目の識別力の比較をまとめたものである。また、図1は、4枝選択から3枝選択に変更した10項目の正答率と識別力の変化を視覚化したものである。4枝選択を3枝選択にした結果、識別力が上がったものは、5項目（項目019、024、027、089、097）であった。その差をみると、識別力が上がった項目については、どの項目も0.070以下の差であった。一方、識別力が下がったものは5項目（項目002、017、018、023、028）であった。その差をみると、5項目中4項目は、0.015から0.081の間であった。しかし、項目028は0.104と10%程度の差があった。これらの数値から、選択枝を減らすことで、識別力がやや下がる傾向があるが、その差は0.081以下程度におさまっている。ただ、項目002、018は、3枝選択のとき、識別力0.300未満となり、識別力が低い項目となったため、選択枝の数を減らしたことが原因かをさらに検討する必要がある。

表5 10項目の正答率の比較

項目	正答率		
	4枝選択	3枝選択	差
002	0.804	0.797 ↓	0.007
017	0.506	0.457 ↓	0.049
018	0.711	0.783 ↑	0.072
019	0.719	0.841 ↑	0.122
023	0.651	0.674 ↑	0.023
024	0.630	0.659 ↑	0.029
027	0.736	0.804 ↑	0.068
028	0.779	0.877* ↑	0.098
089	0.349	0.406 ↑	0.057
097	0.217	0.246 ↑	0.029

表6 10項目の識別力の比較

項目	識別力		
	4枝選択	3枝選択	差
002	0.325	0.250* ↓	0.074
017	0.508	0.442 ↓	0.066
018	0.295*	0.214* ↓	0.081
019	0.485	0.504 ↑	0.018
023	0.440	0.426 ↓	0.015
024	0.593	0.654 ↑	0.062
027	0.419	0.448 ↑	0.029
028	0.470	0.366 ↓	0.104
089	0.514	0.552 ↑	0.038
097	0.407	0.420 ↑	0.013

※正答率の後ろの「*」は0.850以上

※正答率の後ろの「↑」は3枝選択にして正答率が上がったことを、「↓」は下がったことを示す。

※識別力の後ろの「*」は0.300未満

※識別力の後ろの「↑」は3枝選択にして識別力が上がったことを、「↓」は下がったことを示す。

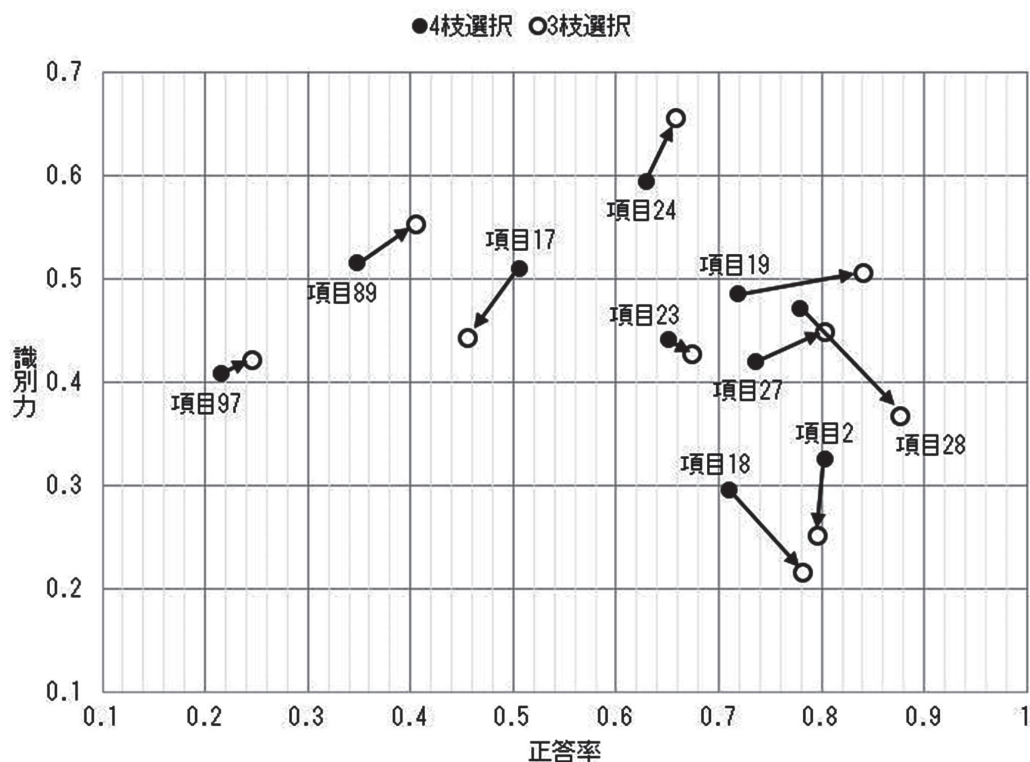


図1 4枝選択から3枝選択に変更した10項目の正答率と識別力の軌跡

4.5 選択枝の数と解答の正誤の関連性について

選択枝の数（3枝か4枝か）と、結果（項目に正答したか誤答したか）との間に関連があるかを検討するため、4枝選択から3枝選択へ変更した10項目において、カイ二乗検定を行った。表7は、10項目における結果を示したものである。4枝選択と3枝選択における正答と誤答の人数、（ ）内は期待度数を表している。

カイ二乗検定の結果、10項目中8項目は、3枝か4枝かは、正答したか否かに影響しないことがわかった。しかし、項目019と項目028においては、4枝は誤答が多く、3枝は正答が多いことが明らかとなった（項目019： $\chi^2(1)=4.965, p<.05$ 項目028： $\chi^2(1)=4.134, p<.05$ ）。そこで、この2項目の4枝と3枝の正答率、識別力を比較する。表8は、項目019と項目028における4枝と3枝の正答率、識別力を比較した表である。3枝選択の場合、項目019は正答率0.841、識別力0.504、項目028は正答率0.877、識別力0.366であった。

表7 10項目における選択枝の数と正誤のクロス表（人数）

項目	4枝		3枝		χ^2	df	p
	正答	誤答	正答	誤答			
002	189 (188.1)	45 (45.9)	110 (110.9)	28 (27.1)	0.013	1	ns
017	119 (114.2)	112 (116.8)	63 (67.8)	74 (69.2)	0.842	1	ns
018	167 (171.7)	59 (54.3)	108 (103.3)	28 (32.7)	1.130	1	ns
019	169 (178.0)	59 (50.0)	116 (107.0)	21 (30.0)	4.965	1	<.05
023	153 (153.8)	72 (71.3)	93 (92.3)	42 (42.8)	0.003	1	ns
024	148 (150.4)	76 (73.6)	91 (88.6)	41 (43.4)	0.193	1	ns
027	173 (177.7)	56 (51.3)	111 (106.3)	26 (30.7)	1.180	1	ns
028	183 (190.3)	43 (35.7)	121 (113.7)	14 (21.3)	4.134	1	<.05
089	82 (88.9)	97 (90.1)	56 (49.1)	43 (49.9)	2.535	1	ns
097	51 (55.2)	119 (114.8)	34 (29.9)	58 (62.2)	1.020	1	ns

()内は期待度数

表8 4枝と3枝の正答率と識別力の比較（項目019と028）

項目	4枝選択		3枝選択	
	正答率	識別力	正答率	識別力
019	0.719	0.485	0.841	0.504
028	0.779	0.470	0.877	0.366

両項目とも、識別力は特に問題はないと思われるが、正答率は高く、3枝で正答が多くなっていた。そこで、表9「4枝選択から3枝選択に変更した10項目の問題文と選択枝」(表2一部再掲)の項目019と028において、各選択枝の選択率を確認したところ、項目019の選択枝「どこ」は4枝選択の場合、4.82%、3枝選択の場合、2.92%であった。また、項目028の選択枝「けし」は4枝選択の場合、5.75%、3枝選択の場合、3.70%であり、いずれの場合も選択率が5%程度、あるいは、それ以下であることがわかった。これらの選択枝は、誤答選択枝として十分に機能しておらず、これが正答率を高くしている要因となっている可能性がある。

表9 4枝選択から3枝選択に変更した10項目の問題文と選択枝(表2一部再掲)

項目019				
問題文	() スポーツが すきですか。			
4枝	どこ(4.82%)	なに(18.86%)	<u>いつ</u> (2.19%)	どんな*(74.12%)
3枝	どこ(2.92%)	なに(12.41%)	どんな*(84.67%)	
項目028				
問題文	うるさいから、テレビを () ください。			
4枝	けし(5.75%)	<u>けす</u> (3.98%)	けすて(9.29%)	けして*(80.97%)
3枝	けし(3.70%)	けすて(6.67%)	けして*(89.63%)	

※選択枝の横の()は、選択率、「*」は正答、下線は削除対象となった選択枝を示す。

5. 考察

本研究では、日本語テストにおいて選択枝の数を4枝から3枝にするという変更が項目の正答率と識別力にどのような影響を与えるかを検証した。その結果、正答率が上がったものは10項目中8項目あり、その差をみると、8項目中6項目は0.080以下であった。しかし、8項目中2項目については10%程度の差があった。一方、正答率が下がったものは2項目で、その差は0.050以下であった。4枝選択を3枝選択にすることで、正答率がやや上がる傾向がみられた。正答率の差について、Delgado & Prieto (1998)は、4枝選択と3枝選択の項目の正答率を比較し、各項目において、わずかに上がる傾向があると述べている。そして、4枝選択と3枝選択の項目の正答率の平均値に差があるかをt検定を行い確認し、その差は有意ではなかったとしている。本研究においては、10項目中8項目の正答率が上がったが、その中の6項目の上がった数値は0.080以下であったため、正答率に大きな影響を与えているとは言い難い。

また、選択枝の数の変更が識別力に与える影響についての検証では、4枝選択を3枝選択にすることで、識別力が上がったものは、10項目中5項目であった。その差をみると、識別力が上がった項目については、どの項目も0.070以下の差であった。一方、識別力が下がったものは5項目であり、その差をみると、5項目中4項目は、

0.015から0.081の間であった。しかし、項目028は0.104と10%程度の差があった。これらの結果から、選択枝を減らすことで、識別力が下がる傾向があるが、その差は0.080程度であった。この数値からみると、識別力に大きな影響を与えているとは言い難い。

さらに、選択枝の数（3枝か4枝か）と、結果（項目に正答したか誤答したか）との間に関連があるかを検討するため、4枝選択から3枝選択へ変更した10項目において、カイ二乗検定を行った。その結果、10項目中8項目は、3枝か4枝かは、正答したか否かに影響しないことがわかった。しかし、項目019と項目028においては、4枝選択は誤答が多く、3枝選択は正答が多いことが明らかとなった。また、3枝選択の場合、正答率も高くなっていた。当該2項目の選択枝の中に、4枝選択のときも、3枝選択のときも、選択率が5%程度、あるいはそれ以下である選択枝があり、それらは誤答選択枝として十分に機能しておらず、正答率を高くしている要因となっている可能性がある。

6. まとめと今後の課題

本研究では、選択枝を4枝選択から3枝選択にすると、正答率は上がる傾向、識別力は下がる傾向があることが明らかとなった。しかし、4枝選択と3枝選択の正答率や識別力の差の数値からみると、選択枝の数の変更が、正答率や識別力に大きい影響を与えるとは言い難い。ただし、この結果は10項目という限られた項目数によるものである。そのため、本研究の結果をもって一般性の高い結論を導くことには限界がある点に注意が必要である。

また、今後の課題として、

- 1) 4枝選択と3枝選択の場合の解答時間の変化や、無答の項目の有無に関する分析など、受験者の解答行動の変化についても明らかにすること、
 - 2) 本研究では、日本語の文法や語彙の理解を問う問題が中心のオンラインPTの結果をデータとして扱ったが、聴解テストや読解テスト等を実施した結果から、技能による違いについて検討すること、
- などが挙げられる。

さらに、今回は、選択枝の数の変更を行う際、4枝選択の項目の各選択枝の選択率が既知で、最も選択率が低い選択枝を除いて3枝選択にしたが、実際のテスト作成場面では、どの誤答選択枝の選択率が低いかは作成時点では分からないはずである。そこで、より現実のテスト作成場面に即して考えるならば、PTの作成時に、4枝選択の誤答選択枝の作成順序を分かるようにしておくことで、選択率との関係から、無理をして、誤答選択枝を3つ作成する必要のないことが明らかにできる可能性がある。

このような研究手続きに関する検討も必要であるが、この点も含めて今後の研究課

題として示しておく。

注

- 1) 日本テスト学会（編）（2007）『テスト・スタンダード 日本のテストの将来に向けて』では、「多肢選択」や「選択肢」を、「多枝選択」や「選択枝」と表記している。本研究においても『テスト・スタンダード』の表記に準じて「多枝選択」や「選択枝」を用いることとする。

参考文献

- 日本テスト学会（編）（2007）『テスト・スタンダード 日本のテストの将来に向けて』金子書房
- 村上京子（2013）「項目分析と改善」関正昭・平高史也（編）『テストを作る（日本語教育叢書「つくる」）』スリーエーネットワーク、199-209
- Delgado, A.R. & Prieto, G. (1998). Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment, 14*, 197-201.
- Haladyna, T.M. & Downing, S.M. (1993). How many options is enough for a multiple-choice test items? *Educational and Psychological Measurement, 53*, 999-1010.
- Rodriguez, M. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research, *Educational Measurement: Issues and Practice, 24*:2.
- Shizuka, T., Takeuchi, O., Yashima, T. & Yoshizawa, K. (2006). A comparison of three-and four-option English tests for university entrance selection purposes in Japan. *Language Testing, 23*, 35-57.
- Sidick, J.T., Barrett, G.V. & Doverspike, D. (1994). Three-alternative multiple-choice tests: an attractive option. *Personnel Psychology, 47*, 829-835.

謝辞

本稿は、日本大学大学院総合社会情報研究科で提出した修士論文の一部をまとめたものです。本稿の執筆にあたり、始終熱心に指導して下さった島田めぐみ教授、大変貴重なご助言をくださった2名の査読者の先生方、多くのご意見をくださったゼミの皆様、調査に協力して下さった日本語学校の皆様に心より感謝申し上げます。また、保坂敏子教授からは、様々な機会とご助言を賜りましたことを、ここに厚く御礼申し上げます。