

日本語プレースメントテストにおける 文法テスト項目の改訂

渡部 倫子

(2012年10月2日受理)

Refining of the Japanese Grammar Placement Test

Tomoko Watanabe

Abstract: This study examines and refines a grammar placement test using the classical test analysis. The participants were 44 international students who were studying Japanese at Okayama University in Japan. After revisions were carried out on grammar test items that are high in item discrimination and low in item facility. The new version was found to be more efficient than the old one, since it make the overall test more difficult while keeping the same reliability level as before. When the result of dividing classes was predicted, it was found that new grammar test items were significant predictors.

Key words: Placement test, Japanese language, Grammar, Classical Test Theory
キーワード：プレースメントテスト、日本語、文法、古典的テスト理論

1. はじめに

岡山大学全学日本語コースのプレースメントテスト（以下、プレースメントテスト）は2005年後期に開発され、これまで、文法テスト項目の分析と改訂（坂野, 2009; 渡部, 2011）、聴解テスト項目の分析（渡部, 2012）が行われてきた。坂野（2009）は項目応答理論の一つとして考えられている Rasch モデル（Rasch, 1960/1980）を用いて、文法テストの各項目を分析した。分析の結果、以下の改訂すべき点が挙げられた。

1) テスト全体の難易度を上げる。2) 全ての項目を受験者能力が高い者の方が低い者より正答できる項目とする。3) 予測される解答パターンとのずれや受験者能力値と正答との相関に問題がある選択肢をなくす。これらの点を参考に文法テスト項目が改訂された。渡部（2011）は、2010年と2011年に実施されたプレースメントテスト（改訂された文法テスト項目）を分析し、聴解テスト項目、文法テスト項目共に、高い信頼性係数を得た。また、聴解テストと文法テストに

よってクラス分け結果を十分に予測できることも分かった。一方で、依然テスト全体の難易度を上げるといふ課題が残されていることが指摘された。本稿では、これまでの分析結果をふまえ、文法テスト項目を再度改訂し、プレースメントテストとして使用するにあたって適切な項目であるかどうかを検証する。なお、聴解テスト項目の改訂は今後の課題とする。

2. 旧文法テスト項目の概要

プレースメントテストはオンライン化されており、聴解テスト、文法テストの順に実施される（坂野他, 2010）。文法テストは、全80問からなっている。第1問 - 第20問は初級1で学習する文法から出題され、第21問 - 第40問は初級2、第41問 - 第60問は中級入門、第61問 - 第80問は中級1で学習する文法となっている。以上の問題は大きく2つのセクションに分けられており、セクション1（第1問 - 第40問）の終了後、自動採点が行われ、セクション1及び聴解テストの合

計点が一定以上あれば、次のセクション2（第41問～第80問）に進む。点数が足りなければ、これで終了となる。文法テストの問題については、旧日本語能力試験2級、3級、4級の問題から選んだもの、旧日本語能力試験の問題から語彙等を部分的に変えたもの、岡山大学で独自に作成したものがある。問題は多肢選択問題で、回答は4つの選択肢からなる⁽¹⁾。

被験者は聴解テストと文法テストの総合点によって、初級1、初級2、中級入門、中級1、中級2、上級の6レベルのいずれかに振り分けられる。旧分割点を表1に示す。

3. 新文法テスト項目における改訂点

最も大きい変更点は上級レベルが増設されたことである。被験者は聴解テストと文法テストの総合点によって、初級1、初級2、中級入門、中級1、中級2、上級1、上級2の7レベルのいずれかに振り分けられる。新分割点を表1に示す。

セクション1においては、項目容易度が高く、項目弁別力が低い2問を削除し、全38問に変更した。また、旧文法テストでは、問題文と選択肢における全ての漢字にルビを付していたが、旧日本語能力試験3・4級の漢字のみルビを付し、その他の漢字はひらがなで表示することにした。

セクション2においては、旧日本語能力試験の1級に相当する7問を新たに追加し、全47問に変更した。

この7問の選定にあたっては、21名の被験者によるパイロットテストを実施した。その分析の結果、試作問題15問の中から項目容易度が低く、項目弁別力が高い7問が精選された。新問題の追加によって、プレースメントテストは、文法テスト項目（セクション1、セクション2）85問、聴解テスト項目15問の計100問で構成されることとなった。また、セクション2の問題文と選択肢の漢字については、問題番号39から58までは旧日本語能力試験3・4級の漢字のみルビを削除し、問題番号59以降は全てのルビを削除することにした。

以上のように、難易度が高い文法問題の追加とルビ無しの漢字による提示によって、テスト全体の難易度を上げることを目指した。

4. 被験者と分析方法

新文法テスト項目の試行は、2012年4月に実施された岡山大学全学日本語コースのプレースメントテストの一部として行った。被験者は岡山大学に在籍する留学生44名である。そのうち、漢字圏学習者は23名、非漢字圏学習者は21名であった。

本稿では古典的テスト理論の観点から、文法テスト項目の信頼性の検討ならびに項目分析を行う。分析にはR (2.14.0)を用いた。Rは統計計算とグラフィックスのための言語・環境である (RjpWiki, 2004)。なお、Raschモデルを用いた分析は次稿に譲ることとする。

表1 新旧分割点

| 聴解テスト+文法テストの合計点 | | レベルほか |
|-----------------|--------------|-----------|
| 旧分割点 (95 問) | 新分割点 (100 問) | |
| 0-18 点 | 0-16 点 | 初級1 |
| 19-31 点 | 17-29 点 | 初級2 |
| 32 点以上 | 30 点以上 | セクション2受験可 |
| 32-57 点 | 30-55 点 | 中級入門 |
| 58-69 点 | 56-66 点 | 中級1 |
| 70-82 点 | 67-77 点 | 中級2 |
| 83 点以上 | 78-88 点 | 上級 (上級1) |
| | 89 点以上 | 上級2 |

表2 記述統計量と内的一貫性による信頼性の推定 (n = 44)

| | 平均値 | 標準偏差 | α 係数 |
|---------------|-------|-------|-------------|
| 聴解テスト (15 問) | 9.93 | 3.81 | 0.85 |
| 文法テスト (85 問) | 47.68 | 21.32 | 0.97 |
| セクション1 (38 問) | 27.93 | 7.83 | 0.92 |
| セクション2 (47 問) | 19.75 | 14.19 | 0.97 |
| 総合得点 (100 問) | 57.61 | 24.73 | 0.98 |

表3 項目容易度, 項目弁別指数, 点双列相関係数

| | | | | | | | | | | | | | | | | | |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|------|------|
| | G1 | G2 | G3 | G4 | G5 | G6 | G7 | G8 | G9 | G10 | G11 | G12 | G13 | G14 | G15 | G16 | G17 |
| 項目容易度 | 0.84 | 0.82 | 0.66 | 0.48 | 0.86 | 0.80 | 0.93 | 0.80 | 0.39 | 0.89 | 0.93 | 0.84 | 0.80 | 0.82 | 0.91 | 0.84 | 0.86 |
| 項目弁別指数 | 0.42 | 0.50 | 0.50 | 1.00 | 0.50 | 0.58 | 0.17 | 0.58 | 0.75 | 0.33 | 0.00 | 0.50 | 0.50 | 0.50 | -0.08 | 0.25 | 0.50 |
| 点双列相関係数 | 0.95 | 0.45 | 0.64 | 0.65 | 0.37 | 0.32 | 0.91 | 0.47 | 0.90 | 0.79 | 0.43 | 0.94 | 0.76 | 0.57 | 0.50 | 0.15 | 0.12 |
| | G18 | G19 | G20 | G21 | G22 | G23 | G24 | G25 | G26 | G27 | G28 | G29 | G30 | G31 | G32 | G33 | G34 |
| 項目容易度 | 0.77 | 0.73 | 0.86 | 0.82 | 0.84 | 0.73 | 0.77 | 0.39 | 0.50 | 0.70 | 0.80 | 0.66 | 0.75 | 0.50 | 0.43 | 0.66 | 0.41 |
| 項目弁別指数 | 0.50 | 0.50 | 0.33 | 0.33 | 0.08 | 0.75 | 0.67 | 0.08 | 0.33 | 0.50 | 0.58 | 0.75 | 0.83 | 0.58 | 0.42 | 0.67 | 0.00 |
| 点双列相関係数 | 0.20 | 0.99 | 0.45 | 0.65 | 0.66 | 0.89 | 0.58 | 0.02 | 0.23 | 0.67 | 0.25 | 0.87 | 0.93 | 0.82 | 0.81 | 0.81 | 0.41 |
| | G35 | G36 | G37 | G38 | G39 | G40 | G41 | G42 | G43 | G44 | G45 | G46 | G47 | G48 | G49 | G50 | G51 |
| 項目容易度 | 0.89 | 0.86 | 0.75 | 0.66 | 0.73 | 0.30 | 0.50 | 0.43 | 0.59 | 0.73 | 0.41 | 0.64 | 0.39 | 0.73 | 0.45 | 0.68 | 0.32 |
| 項目弁別指数 | 0.42 | 0.42 | 0.58 | 0.83 | 0.92 | 0.58 | 1.00 | 0.75 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.42 |
| 点双列相関係数 | 0.81 | 0.84 | 0.60 | 0.95 | 0.60 | 0.84 | 0.47 | 0.70 | 0.67 | 0.33 | 0.30 | 0.66 | 0.41 | 0.47 | 0.94 | 0.98 | 0.54 |
| | G52 | G53 | G54 | G55 | G56 | G57 | G58 | G59 | G60 | G61 | G62 | G63 | G64 | G65 | G66 | G67 | G68 |
| 項目容易度 | 0.64 | 0.50 | 0.55 | 0.48 | 0.48 | 0.55 | 0.27 | 0.18 | 0.30 | 0.45 | 0.43 | 0.20 | 0.34 | 0.57 | 0.39 | 0.48 | 0.48 |
| 項目弁別指数 | 0.92 | 0.92 | 1.00 | 0.83 | 1.00 | 0.83 | 0.83 | 0.50 | 0.50 | 0.83 | 0.67 | 0.58 | 0.83 | 0.92 | 0.75 | 0.92 | 0.75 |
| 点双列相関係数 | 0.22 | 0.22 | 0.62 | 0.08 | 0.87 | 0.92 | 0.56 | 0.85 | 0.71 | 0.72 | 0.31 | 0.32 | 0.32 | 0.10 | 0.37 | 0.38 | 0.58 |
| | G69 | G70 | G71 | G72 | G73 | G74 | G75 | G76 | G77 | G78 | G79 | G80 | G81 | G82 | G83 | G84 | G85 |
| 項目容易度 | 0.41 | 0.39 | 0.36 | 0.50 | 0.32 | 0.43 | 0.50 | 0.48 | 0.25 | 0.48 | 0.25 | 0.27 | 0.14 | 0.18 | 0.20 | 0.39 | 0.05 |
| 項目弁別指数 | 0.67 | 0.83 | 0.75 | 1.00 | 0.75 | 0.92 | 0.92 | 0.92 | 0.50 | 0.83 | 0.67 | 0.42 | 0.25 | 0.50 | 0.50 | 0.83 | 0.08 |
| 点双列相関係数 | 0.86 | 0.64 | 0.17 | 0.31 | 0.03 | 0.84 | 0.40 | 0.26 | 0.97 | 0.77 | 0.71 | 0.17 | 0.72 | 0.43 | 0.59 | 0.26 | 0.23 |

5. 結果と考察

プレースメントテスト結果の要約とテスト項目の内的一貫性(クロンバックの α 係数)を表2に示す。この表から、プレースメントテストの信頼性が高いことが明らかになった。

続いて文法テスト項目の分析を行った。まず、文法テスト項目(以下、G1~G85)の良し悪しを判断するために、項目容易度と項目弁別力(項目弁別指数, 点双列相関係数)を算出した(表3)。項目容易度は「(その項目に正答した人数) ÷ (受験した人数)」で算出され、最適な範囲は.30~.70であるといわれている。項目弁別指数(UL指数)は、得点の上位27%の正答率と下位27%の正答率の差である。項目弁別指数は、.40以上がとてよい項目、.30から.39がよい項目であるが改良が必要かもしれない項目、.20から.29は改良が必要な項目、.19以下はよくない項目で削除するか作り直すことが必要な項目とされている(Brown, 2005: p.75)。

もう一つ、項目弁別力を示すものとして点双列相関係数がある。これは、テストの各項目がテストの総合得点とどの程度関連があるかを示す。一般的には、.00~±.20をほとんど相関がない(.00は無相関)、±

.20~±.40を低い(弱い)相関がある、±.40~±.70をかなり(比較的強い)相関がある、±.70~±1.00を高い(強い)相関がある(+1.00は完全な正の相関、-1.00は完全な負の相関)と解釈する。この点双列相関係数と先に求めた項目容易度との散布図を図1に示す。

表3と図1から、文法テスト項目の多くが適正であ

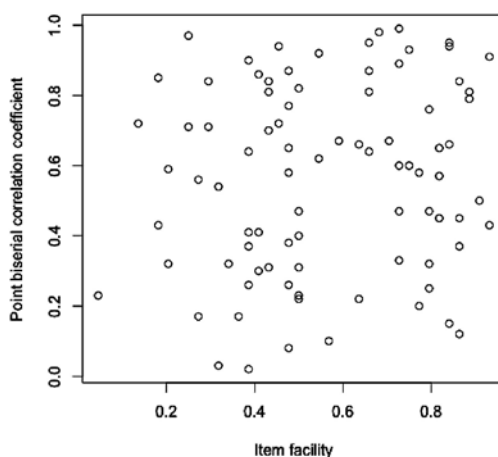


図1 点双列相関係数と項目容易度の散布図

るが、改訂を検討すべき項目も少なくないことが分かる。特に、項目容易度が高く項目弁別指数が低い項目 (G 7, G11, G15), 項目弁別指数が低い項目 (G22, G25, G34), 新たに追加した7問の中では項目容易度と項目弁別指数が低い (G85) について、今後、改良を検討する。

次に、文法テスト項目全体の難易度が今回の改訂によって改善されたかを検討するため、セクション別の項目容易度を図2に示した。坂野 (2009) で検討された Section 1 - 2 が本論 (改訂後) のセクション 1, Section 3 - 4 がセクション 2 にあたる。また, Section 5 は新たに追加した7問である。坂野 (2009) の分析結果では、上に行くほど Section 間の差が小さくなっており, Section 2 と Section 3, Section 3 と

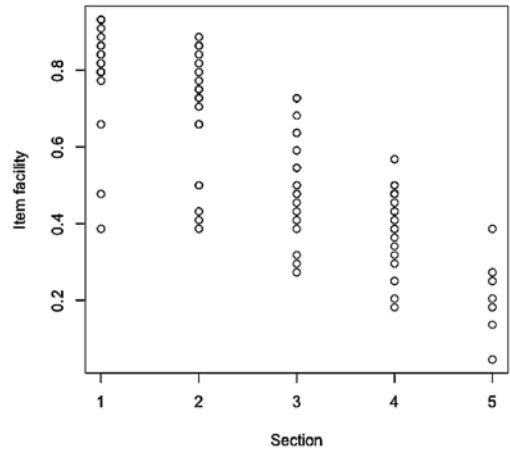


図2 セクション別の項目容易度

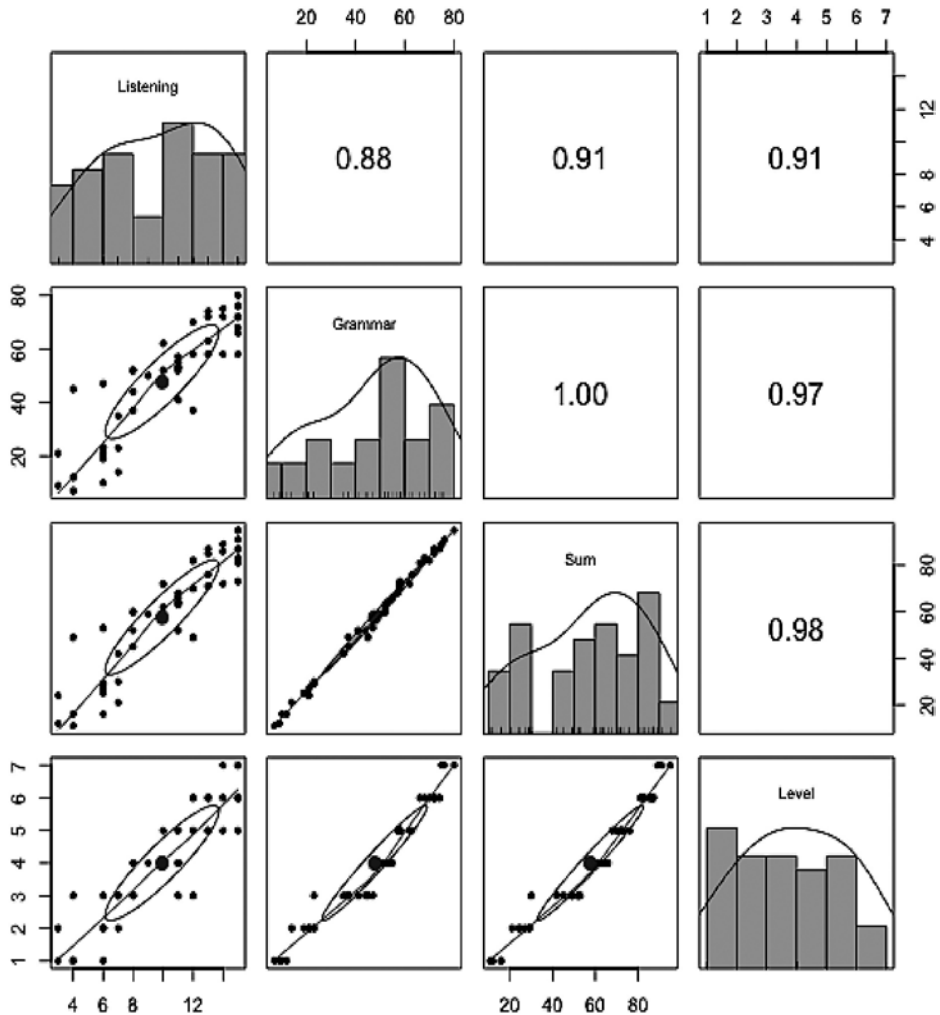


図3 散布図行列

Section 4の項目容易度の差がはっきりとはあらわれていないことが指摘されていた。今回の改訂でSection 3とSection 4の漢字ルビを削除し、新たにSection 5を加えたことで、各Section間の難易度の差がより明確になったことが分かる(図2)。

さらに、得点ごとの人数の分布を視覚化するため、聴解テスト(Listening)、文法テスト(Grammar)、総合得点(Sum)、クラス分け結果(Level)の4つの変数間の関係を図3に示した。図3の左下左端のy軸(1-7)は、クラス分け結果(Level)を意味し、初級1、初級2、中級入門、中級1、中級2、上級1、上級2レベルに対応している。聴解テスト、文法テスト、総合得点、クラス分け結果の4つの変数間の関係を解明するため、相関係数を求め、図3中の右上の数値および左下の散布図に示した。各図のx軸とy軸には棒グラフ(ヒストグラム)中に記載されている変数名が対応している。例えば、1.00は文法テストと総合得点の相関係数を示している。また、左下左端の散布図のx軸は聴解テスト(Listening)、y軸はクラス分け結果(Level)である。なお、散布図中の楕円形は相関係数の95%信頼楕円であり、楕円形の中央にあるフィットライン上の点はx軸とy軸の母平均である。

図3から、4つの変数間に強い正の相関が認められた。また、棒グラフで示された度数分布から、文法テストはやや右寄りの分布となっているものの、上級2以外のレベルには、ほぼ同数の被験者がプレースされたことが分かる。問題数の増加に伴い変更された分割

点(表1)は適切であったと予測される。

最後に、被験者の属性の一つである漢字圏(kanji)／非漢字圏(non-kanji)による分布の違いを図4に示す。文法テストの平均値(標準偏差)は、漢字圏が53.57(19.97)、非漢字圏が41.24(21.34)であった。図4の左から、初級1は非漢字圏の被験者のみ、上級2は漢字圏の被験者のみであったが、その他のレベルでは偏りが見られなかったことが分かる。今回の改訂で漢字のルビを削除したことによって、難易度が上がり過ぎること、非漢字圏の被験者が不利になることといった懸念があった。しかし、今回のサンプルでは目立った問題はなかった。また、聴解テストの平均値(標準偏差)は、漢字圏が11.22(3.45)、非漢字圏が8.52(3.75)であった。文法テスト項目と聴解テスト項目における分布をみると、全体的に漢字圏の被験者のほうが非漢字圏の被験者よりも得点が高いことが分かる(図4の右)。これまで、日本語コース担当教員から度々指摘があった「漢字圏の学習者は文法に比べて聴解の成績が低い傾向であり、非漢字圏の学習者はその逆の傾向である」という経験的な予測は、今回のサンプルには当てはまらなかった。

6. おわりに

本稿では、岡山大学全学日本語コースのプレースメントテストにおける文法テスト項目改訂後の試行結果を報告した。項目容易度が高い問題2問を削除し、難易度が高い新たな問題7問を加え、セクション2の漢字ルビを削除したことによって、文法テスト全体の難易度を上げることが出来た。また、聴解テスト、文法

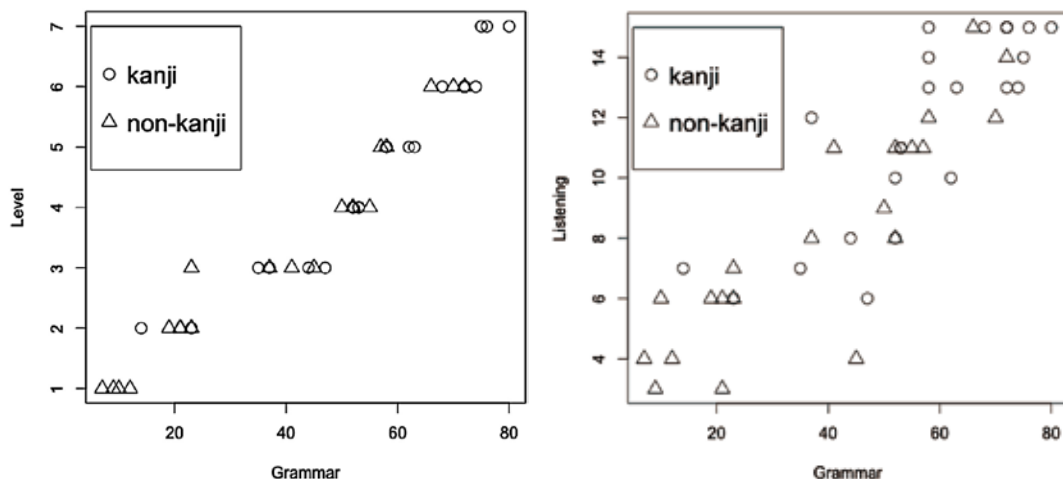


図4 漢字圏 / 非漢字圏による分布の違い

テスト共に、高い信頼性係数が得られた。各テストとレベルが強い共変関係であったことから、改訂後の文法テスト項目によってクラス分け結果を十分に予測できると考える。

しかし、試行時のサンプル数が少なく統計的検討ができないという課題が残された。今後も試行を積み重ね、より適切な項目からなるテストに向けた改訂を進めていきたい。

【謝 辞】

貴重なデータを提供してくださった岡山大学言語教育センターに心から御礼申し上げる。

【注】

(1) 問題例として改訂時に削除された2問を示す。

問1 今 岡山に_____。

- 1) 住みます 2) 住みます 3) 住んでいます
4) 住んでいます

問2 A: あした、いっしょに昼ごはんを_____。

B: いいですね。

- 1) 食べません 2) 食べませんか 3) 食べました
4) 食べましたか

【参考文献】

- 青木繁伸 (2009) 『Rによる統計解析』オーム社。
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』第4号, pp.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, pp.89-106.
近藤ブラウン妃美 (2012) 『日本語教師のための評価入門』くろしお出版。
静哲人 (2007) 『基礎から深く理解するラッシュモデ

リング 項目応答理論とは似て非なる測定のパラダイム』関西大学出版部。

坂野永理 (2006) 「Rasch モデルによる漢字ブレースメントテストの改良」『日本テスト学会誌』2 (1), pp.91-100.

坂野永理 (2009) 「日本語コース文法ブレースメントテストの分析」『大学教育研究紀要』第5号, 岡山大学国際センター, pp.33-42.

坂野永理・渡部倫子・大久保理恵 (2010) 「オンライン日本語ブレースメントテストの開発」『大学教育研究紀要』第6号, 岡山大学国際センター, pp.107-117.

渡部倫子 (2011) 「Rを利用した日本語ブレースメントテストの分析—古典的テスト理論による分析—」『広島大学日本語教育研究』第21号, 広島大学大学院教育学研究科/教育学部日本語教育学講座, pp.25-32.

渡部倫子 (2012) 「日本語ブ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』第17号, 印刷中.

Bachman, L. and Palmer, A. (2010) *Language Assessment in Practice*. Oxford Applied Linguistics.

Brown, J. D. (2005) *Testing in language programs: A comprehensive guide to English language assessment* (New ed.). New York: McGraw Hill.

Rasch, G. (1960/80) *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research.) Expanded edition with foreword and afterword by B. J. Wright. (1980). Chicago: The University of Chicago Press.

RjpWiki (2004) 「What is R? (Rとは?)」 < <http://www.okada.jp.org/RWiki/?RjpWiki> > (2012年9月9日閲覧).

Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing*, 11 (3), 321-344.