

日本語会話テストにおいてテスト形式が受験者の発話に与える影響

—グループ形式とインタビュー形式の比較—

堀川有美*

Effects of Test Format on Student Performance in Japanese Oral Tests :

A Comparison between Group Format and Interview Format

HORIKAWA Yumi

abstract

This paper investigates the effects of test format on student performance in Japanese oral tests by analyzing student discourses during tests. In a high-intermediate Japanese course, 6 Korean students took two types of oral tests: the group format and the interview format. Student performance was analyzed in terms of fluency, accuracy, and complexity based on Skehan's (1996, 1998) approach. The results showed students spoke more accurately in a group than in an interview. The result of fluency was not clear because only one measure showed significance. Complexities were not different between the two formats. Although the problem of small size was admitted, the importance of considering the nature of test format in test design and score inference was implicated.

Keywords : oral test, group format, interview format, performance, interlocutor

1. はじめに

外国語・第二言語教育における会話テストの形式として、試験官と受験者が一対一で対話するインタビュー形式のテストが一般的に用いられている。対面式のインタビューテストは、受験者との実際のやりとりを評価するため、真正性 (authenticity) の高いテスト方法であると考えられてきた。しかし近年、インタビューテストは、試験官が受験者に質問をする形で会話が進められるという、日常会話とは異なる談話的特徴を持つことが指摘され (van Lier 1989; Lazaraton 1992 等)、唯一の会話テスト形式として用いられることに批判の声が上がっている (Swain, 2001 等)。

このようなインタビューテストへの批判を受けて注目されているテスト形式の1つが、学習者同士の話し合いを評価するグループ形式やペア形式の会話テストである。インタビューに比べ参加者間の力関係が対等になり、日常的なコミュニケーションに近い会話が実現すると言われている⁽¹⁾。

言語テストにおいて、テスト形式がテスト結果に影響を及ぼすことは多くの言語テスト研究者が述べていることであるが (Bachman, 1990)、インタビュー形式とグループ/ペア形式に注目し、受験者の発話を比較したものは、管見の限り、後述の Taylor (2000, 2001) のみである。Taylor の研究は、発話の多様性という観点から

キーワード: 会話テスト、グループ形式、インタビュー形式、パフォーマンス、対話者

*平成16年度生 国際日本学専攻

両テストにおける受験者の言語行動の違いを明らかにしたものだが、外国語・第二言語の会話テストにおいては、「流暢であったか」「正確であったか」等、受験者の発話自体の特徴についても評価がなされることが普通である。そこで、本研究では、こうした評価の観点を視野に入れて、グループ形式とインタビュー形式の会話テストにおける受験者の発話の特徴を調べる。具体的には、Skehan (1996, 1998) のアプローチに基づき、流暢さ・正確さ・複雑さの3つの観点から、グループ形式とインタビュー形式の会話テストにおける受験者の発話の特徴を分析する。両テストにおける受験者のパフォーマンス⁽²⁾の異同を明らかにすることによって、テスト形式の選択やテスト得点の解釈のために、より実用的な示唆を得ることを目指したい。

2. 先行研究

2-1 スピーキングテストのモデル

まずはじめに、スピーキングテストにおけるパフォーマンスのモデルを示す。Skehan (2001) は、McNamara (1995) に基づいて、スピーキングテストにおけるパフォーマンスのモデルを提示した(図1)。モデルの中心にパフォーマンスが据えられているが、このパフォーマンスは採点手続きを通して得点に変換されていることから、「パフォーマンス=テスト得点」ではないことがわかる。また、パフォーマンスに影響を与える要因として対話者、受験者の能力、タスクの3つが挙げられており、受験者の言語能力以外の様々な要因が直接・間接的に得点に影響を与えることが示されている。

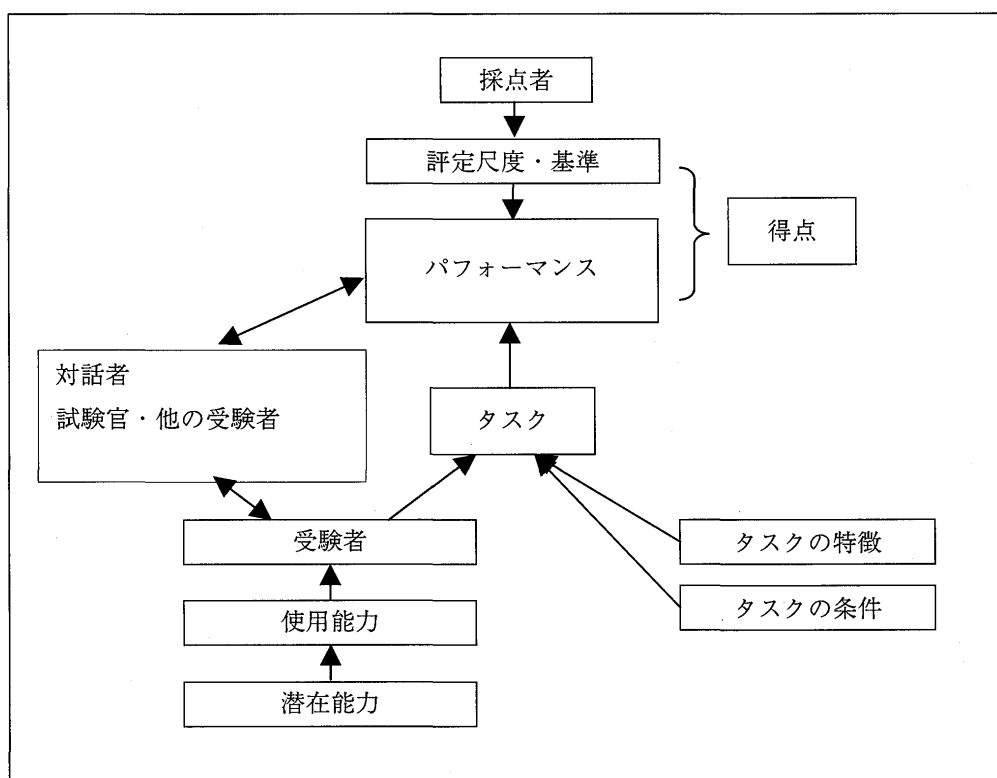


図1 スピーキングテストにおけるパフォーマンスのモデル (A model of oral test performance)
(Skehan, 2001 : 18 より翻訳)

2-2 グループ形式とインタビュー形式の比較

本研究で比較するインタビューテストとグループテストの大きな違いは、Skehan のモデルにおける「対話者」にあると考えられる。インタビューテストは日本語母語話者または日本語熟達者である試験官と受験者との1対1の会話であるのに対し、グループテストでは3名以上の受験者同士が会話を行う。つまり、対話者との関係、

及び会話参加者の人数が異なっていると言える。

グループテストと他のスピーキングテストの比較は、主にテスト得点の比較という形で行われてきた⁽³⁾。しかし、上述の Skehan のモデルで確認したように、テスト得点は採点手続きの影響を含んだものである。対話者およびタスクの影響に焦点を当てるためには、採点手続きの影響を受ける前のパフォーマンスそのものを分析する必要がある。

受験者のパフォーマンスを分析したものとしては、Taylor (2000, 2001) が挙げられる。Taylor は、ケンブリッジ英語検定試験で行われているインタビュー形式とペア形式のスピーキングテストを比較分析し、ペア形式のほうが、ターンが多い・ターンの長さが多様である・発話機能の種類が多い、という結果を報告している。Taylor (2001:15) は、これらの結果から、伝統的な1対1のインタビュー形式よりもペア形式のほうが、より豊富 (richer) でより多様な (more varied) 受験者の言語サンプルを抽出できるとしている。

先に述べたように、これまでグループテストが注目されてきた主な理由の1つは、伝統的な1対1のインタビューテストに比べて、受験者同士が対等な関係で会話に参加でき、現実場面の日常的コミュニケーションに近い会話を実現できるという点である (例: Reves 1982; Fulcher 1996; Bonk & Ockey 2003)。しかし、多くの会話テストにおいては、対等な関係で会話に参加できたかどうかだけでなく、「流暢であったか」「正確であったか」等、発話の特徴の観点からもパフォーマンスが評価される。このような観点からの分析は、まだ行われていない。

2-3 スピーキングパフォーマンスの分析の観点

受験者のパフォーマンスを発話の特徴という観点から分析するために、本研究では、Skehan (1996, 1998) による流暢さ・正確さ・複雑さの3つの観点到に注目した。

Skehan は、言語学習に関する認知的アプローチに理論的基盤を置き、言語学習の目標として、この3つを区別することを主張した。学習者は、言語を使用する際に、意味 (流暢さ) と形式 (正確さ・複雑さ) の両方に同時に注意を払うことはできないため、どちらかを犠牲にしてもう片方に焦点を当てるといふ。例えば、意味を重視した発話では流暢さが増す一方で、形式への注意が疎かになり、正確さや複雑さが低下するという。また、形式への注意はさらに、自分が知っている確実な形式を用いようとする場合 (正確さ) と、誤用を恐れず新しい形式を用いようとする場合 (複雑さ) の2つに分けられる。

3つの観点は、以下のように定義されている (Skehan & Foster, 1999: 96-97)。

流暢さ (fluency) : 語彙化されたシステムを使いながら、リアルタイムで言語を使用し、意味を強調する力。

正確さ (accuracy) : 言語をコントロールしながら、誤用を防ぐ能力。誤用を引き起こし得る挑戦的 (challenging) な構造は使わない。

複雑さ (complexity) : 十分にコントロールできない可能性がありながらも、より上級の言語を使う力。

この3つの観点は、タスク重視型 (task-based) 教授法におけるタスクの研究 (Foster & Skehan 1996; Ortega 1999; Skehan & Foster 1999 等) やテスト場面におけるタスクの分析 (Wigglesworth 1997; Iwashita, McNamara, & Elder 2001 等) において、タスクの特性や実施条件⁽⁴⁾が学習者のパフォーマンスに与える影響を分析する際に用いられてきた。

このように、理論的根拠を持ち、テスト研究を含む複数の研究において用いられていることから、流暢さ・正確さ・複雑さの3つの観点で分析を行うことにより、実用的な示唆を得るだけでなく、先行研究との比較も可能になると考えられる。

3. 研究目的

本研究では、グループテストとインタビューテストを、対話者の関係や人数が異なる2種類のテスト形式と捉え、このテスト形式が受験者のパフォーマンスにどのような影響を与えるかを探る。

研究課題: グループ形式テストとインタビュー形式テストでは、受験者のパフォーマンスの流暢さ・正確さ・複雑さに違いがあるか。

4. 研究方法

4-1 対象者とテストの位置づけ

2003年6月～8月に発音・会話学習を目的として設置された日本語クラスの参加者のうち、6名を分析対象とした。全員、日本に滞在している韓国人学習者であり、職業は主婦・学生・教師・アルバイトであった。日本語学習歴・滞日期間にはばらつきがあるが、コース開始前に実施したSPOT (A版) の点数⁽⁵⁾および事前面談⁽⁶⁾における日本語でのやりとりから、中級以上の日本語レベルと判断された。

クラスは週1回2時間行われ、毎回テーマを決めて、ディスカッションやロールプレイなどの活動を行った。コース後半の時期に、学習者の会話能力診断およびフィードバックを目的として、グループテストとインタビューテストを実施した。後日、各受験者の長所と弱点に関する情報をフィードバックしたが、本クラスが研究の一環として実施されたクラスであったことから、テスト結果を成績決定やコース修了認定のために用いることはなかった。したがって、本テストは、結果が何か重大な決定に使われることはない、利害関係の小さいテスト (low-stakes test) であったと言える。

4-2 テスト実施手続き

両テストとも、「相手とのやりとりにおいて、自分の考えや意見を日本語で述べる力」を測定するためのテストとした。話題は、①「仕事を選ぶ基準」、②「住むところを選ぶ基準」の2つである。6名の受験者のうち、3名は話題①をインタビュー、話題②をグループで受験し、他の3名は、話題②をインタビュー、話題①をグループで受験した (表1参照)。

表1 テストの話題と形式の組み合わせ

	話題①仕事	話題②住むところ
受験者 A、B、C	インタビュー	グループ
受験者 D、E、F	グループ	インタビュー

グループテストは1グループ3名で実施した。組み合わせは、コース中の学習者の様子も考慮した上で、レベルや学習者の背景が比較的近い者同士が同じグループになるよう、教師が決定した⁽⁷⁾。話題は、質問を書いたカードによって提示し、開始と終了の合図以外は学習者に進行を任せた。もしも会話が停滞したり発話が極端に少ない受験者がいたりした場合には教師が介入するという対処方法を予め決めておいた。時間は、Fulcher (1996) を参考に、約15分間と設定した。

インタビューテストは、インタビュアーと受験者の1対1で行われた。インタビュアーは、2名の女性日本語教師であった。インタビュアーには、事前にテストの流れを説明し、他の学習者を対象に練習を行った。テストの流れは、「ウォームアップ→メインの質問→終結部」とし、時間は、一人当たりの時間がグループテストと同程度になるよう、約7分間とした⁽⁸⁾。

4-3 分析方法

グループテストとインタビューテストの発話を文字化し、その文字化データを後述のASユニット (Foster, Tonkyn, & Wigglesworth, 2000) に分けた後、受験者の発話を対象に流暢さ・正確さ・複雑さを表す各指標の値を算出した。その後、グループ形式とインタビュー形式の平均値または中央値に差があるかどうか検定を行った。統計分析は、全て統計処理ソフト「Statcel97.xls」(柳井, 1998) を用いて行った。

なお、文字化は、全ての発話を書き起こした「完全版」と、フィラー・自己訂正・単純なくり返し・出だしの言い誤り・相づちを除いた「簡潔版」の2種類を作成した。正確さ及び複雑さの分析には簡潔版を使用した。流暢さの分析では、くり返しや言い換えなどの分析が必須であることから、完全版を用いた。

以下では、ASユニットと、流暢さ・正確さ・複雑さの各指標について説明する。

4-3-1 ASユニット

ASユニット (the Analysis of Speech Unit) は、「従属節⁹⁾を含む、1つの独立節または副節単位 (sub-clausal unit) から成る一人の話者の発話」と定義される (Foster, Tonkyn, & Wigglesworth, 2000)。独立節とは定型動詞 (finite verb) を含む最小限の節である。日本語における例を例1に示す。

- 例1 A: 日本語はどこで勉強したんですか
→ B: 韓国で勉強しました。

副節単位とは、省略された要素を談話や状況の文脈から補うことによって完全な節にすることができる1つ以上の句 (例2)、または「不規則文や「非文」に分類される最小限の発話 (例3) のどちらかである。

- 例2 A: 日本にいつ来ましたか。
→ B: 1年前。
例3 A: ありがとう。

ASユニットは、話し言葉の特徴を量的に分析するために、考案されたものである。類似の単位としてTユニット、Cユニット、アイデア・ユニット等があるが、ASユニットは、話し言葉の特徴 (介入、共話、言い誤り、等) も考慮した詳細な定義がされている点、音声的特徴やターン交替ではなく統語構造を第1の判断基準にしている点が特徴である。

本研究では、半数のデータ (インタビューテストおよびグループテスト各々3名分) について、協力者1名と筆者の間で一致率を求めたところ、91.0%であった。残りの半数については筆者が1名で行った。

4-3-2 分析指標

(1) 流暢さ

流暢さの指標としては、沈黙やポーズの長さ・頻度、言い換え等の非流暢性を示す特徴、発話速度などが用いられている。本研究では、1秒以上の沈黙の回数・秒数、および以下4つの非流暢性を示す特徴の出現率 (1ASユニット当たり) を求めた。

言い換え (reformulations): 統語、形態素、語順を修正してくり返された節または句。

代替 (replacements): すぐに他と交替された語彙。

出だしでの言い誤り (false starts): 完成前に放棄された発話。言い換えられる場合もそうでない場合も含む。

くり返し (repetitions): 統語、形態素、語順が修正されずくり返された単語、句、節。

沈黙は、ターン間で生じた場合など、どの話者の沈黙なのか判断できない場合がある。そのため、基本的にASユニットの途中で生じた沈黙のみを数えた。ただし、ターン間の沈黙であっても質問などによって次の話者が明らかな場合は、次の話者の沈黙とした。

(2) 正確さ

正確さは、助詞や複数形など特定の文法項目の誤りを測定する方法 (specific measures) と、全体的な正確さを測定する方法 (general measures) の2つに大きく分けられるが、後者の全体的な正確さを測定する方法のほうが特定の項目に偏らずにデータの特徴を把握できるため、適切な測定方法であると言われている (Foster & Skehan, 1996: 304)。本研究でも後者を採用し、具体的には、1ASユニット当たりの誤用数、及び全節数に占める誤用のない節の割合を算出した。誤用は、自己訂正が行われなかった語彙・文法の誤用をカウントした¹⁰⁾。また、一つのデータの中に同じ誤用がくり返し現れた場合は、まとめて1回として計算した。

(3) 複雑さ

複雑さは、通常、1Tユニット当たりの節数、すなわち従属節の割合によって測定される。これに加えて、語彙の多様性も複雑さに加えられることがある。本研究では、構文の複雑さを1ASユニット当たりの節数によって測定し、語彙の多様性を総語数に対する異なり語数の割合 (type-token ratio) によって測定する。

本研究で用いる指標を表2にまとめる。

表2 本研究で用いる分析指標と文字化資料の種類

	指標	文字化資料
流暢さ	・ 1秒以上の沈黙の回数	完全版
	・ 1秒以上の沈黙の合計秒数	
	・ 1ASユニット当たりの言い換え・代替・くり返し・出だしでの言い誤りの数	
正確さ	・ 誤用のないASユニットの割合	簡潔版
	・ 1ASユニット当たりの誤用数	
複雑さ	・ 1ASユニット当たりの節数	簡潔版
	・ 総語数に占める異なり語数の割合	

5. 結果

5-1 流暢さ

流暢さの6つの指標について、それぞれの記述統計および検定結果を表3・表4に示す⁽¹⁾。全ての指標において、インタビューのほうが平均値が高いことがわかり、グループテストのほうが受験者の発話が流暢である傾向が見られた。しかし、検定の結果、有意差が得られたのは、出だしでの言い誤りの1指標のみであった。

表3 非流暢性を表す特徴の出現率 (1ASユニット当り) の記述統計およびt検定の結果

	インタビュー (n=6)		グループ (n=6)		t (5)	
	平均	標準偏差	平均	標準偏差	t	p
言い換え	0.16	0.07	0.07	0.04	2.27	0.07
代替	0.06	0.06	0.01	0.01	2.12	0.09
くり返し	0.10	0.05	0.08	0.05	0.94	0.39
1秒以上の沈黙回数	0.26	0.22	0.12	0.10	1.29	0.25
沈黙秒数	3.69	3.46	1.92	1.66	1.08	0.33

表4 出だしでの言い誤りの記述統計およびウィルコクソン符号付順位和検定の結果

	インタビュー (n=6)		グループ (n=6)		T
	平均	標準偏差	平均	標準偏差	
出だしでの言い誤り	0.37	0.05	0.23	0.10	1

※ T 値の有意点は、 $T < 2$ 、 $19 < T$ ($p < 0.1$)

5-2 正確さ

正確さは、2つの指標についてインタビューとグループテストの結果を比較し、検討した。1ASユニット当たりの誤用数は、対応のあるt検定により検討した(表5)。誤用のない節の割合は、正規性を仮定できなかったため、ウィルコクソン符号付順位和検定を行った(表6)。分析の結果、1ASユニット当たりの誤用数はインタビューのほうが多く、誤用のない節の割合はグループテストのほうが高かった。両方とも、有意差が得られた。正確さの2つの指標とも、インタビューよりグループテストのほうが受験者の発話が正確であるという結果を示した。

表5 1ASユニット当たりの誤用数の記述統計およびt検定の結果

	インタビュー (n=6)		グループ (n=6)		t (5)	
	平均	標準偏差	平均	標準偏差	t	p
1ASユニット当たりの誤用数	0.24	0.11	0.16	0.08	2.88	0.03

表6 誤用のない節の割合の記述統計およびウィルコクソン符号付順位和検定の結果

	インタビュー (n=6)		グループ (n=6)		T
	平均	標準偏差	平均	標準偏差	
誤用のない節の割合	88.59	5.93	92.38	4.73	1

※ T 値の有意点は、 $T < 2$ 、 $19 < T$ ($p < 0.1$)

5-3 複雑さ

1ASユニット当たりの節数(従属節の割合)により統語的複雑さを調べ、総語数に占める異なり語数の割合により語彙的複雑さを調べた。2つの指標とも、平均値はグループテストのほうが若干高かったが、対応のあるt検定を行ったところ、有意差は得られなかった(表7)。

さらに、標準偏差を見ると、グループテストのほうが大きいことが分かる。特に総語数に占める異なり語数の割合では、両者の標準偏差に大きな差がある。このことから、グループテストにおける受験者発話の複雑さは、インタビューテストに比べて受験者個人によるばらつきが大きいことが推測される。

表7 複雑さの指標の記述統計およびt検定の結果

	インタビュー (n=6)		グループ (n=6)		t (5)	
	平均	標準偏差	平均	標準偏差	t	p
1ASユニット当たりの節数	1.93	0.21	2.14	0.42	-1.43	0.21
総語数に占める異なり語数の割合	37.72	3.85	37.80	10.33	-0.03	0.98

表8 結果のまとめ

結果	
流暢さ	グループテストのほうが発話が若干流暢である(1指標でのみ有意差あり)
正確さ	グループテストのほうが発話が正確である
複雑さ	差はない

6. 考察

本研究では、グループテストとインタビューテストにおける受験者の発話を、流暢さ・正確さ・複雑さの3つの観点から比較分析した。それぞれの分析結果の要因を考察する。

流暢さは、5つの指標のうち1指標のみ（出だしでの言い誤り）でグループテストのほうが発話が流暢であるという結果が得られた。明確な結果が得られなかったものの、グループテストのほうが若干流暢さが増す傾向があると考えられる。Skehan (2001) は、対話者のいる対話タスクと独話タスクを比較し、対話タスクのほうが流暢さが低いことを明らかにしている。その理由として、対話タスクでは相手の発話内容を予測できないために、リアルタイムでの言語処理に負荷がかかるという説明がされている。グループテストとインタビューには両方とも対話者が存在するが、相手発話の予測可能性という点では違いがあると推測される。インタビューでは、インタビュアーが発話内容の大部分をコントロールしており、次に何が話されるかを受験者が予測することは難しい。それに対して、グループテストでは、受験者同士が協働で会話を作り上げていくなかで、受験者自身が話題を提供したり、相手に質問したりする。そのため、グループテストのほうが相手の発話内容を予測しやすいと考えられる。ただし、データ数を増やし、他要因との相互作用も考慮に入れて、さらに検証する必要がある。

正確さに関しては、分析に用いた2つの指標ともグループテストにおける発話のほうが正確であるという結果を示した。一般に、会話においては、相手が使った語彙や表現を自分の発話に取り入れて話すということが行われているが、インタビューでは日本語母語話者であるインタビュアーが対話者であったことを考えると、インタビューのほうが正確な発話となってもおかしくない。実際に今回のデータでも、受験者がインタビュアーの使用した語彙を自らの発話に取り入れた例が見られたが、しかし、全体の誤用数を比較したところ、グループテストにおける誤用のほうが少なかった。この理由として考えられることの1つは、グループテストでは学習者自身に会話のコントロールが任されているため、誤用の起きない確実なレベルの日本語で話すことができたという解釈である。それに対してインタビューでは、インタビュアーによって導入された話題に応じて日本語のレベルが高くなる場合があると考えられ、その結果、誤用が増えた可能性がある。しかし、5-3で述べたように複雑さには両テスト間で差が見られなかったことから、日本語のレベル以外にも要因があると考えられる。上述のSkehan (2001) では、独話タスクよりも対話者のいるタスクのほうが、相手の発話中に準備ができるため、発話が正確になることが明らかにされている。インタビューは対話タスクではあるが、受験者の発話を引き出すことが目的であるため、インタビュアーよりも受験者の発話のほうが圧倒的に多くなる。しかし、グループテストでは、他の受験者が発話している間に自分は何を言うか考え、準備することが可能である。このように、受験者自らが発話内容や発話機会を決定できることが、グループテストにおける受験者発話の正確さにつながったと推測される。

複雑さについては、両テスト間で違いは見られなかった。しかし、グループテストにおいて個人差が大きかったことから、受験者によっては、どちらかのテストの発話のほうが複雑さが増していたと考えられる。つまり、何らかの個人要因との相互作用があると考えられる。複雑な発話ができるかどうかは、言語能力の熟達度に加えて、「誤用を恐れずに新しい形式にチャレンジする」という意欲も関わっている。したがって、受験者の性格・評価に対する動機づけなどの個人要因の影響も、大きいと考えられる。また、語彙的複雑さについては、話題の影響もあるだろう。受験者がよく知っている話題であれば、使用する語彙の種類も増えると考えられるためである。

以上の結果から、会話テストにおける結果の解釈にどのような示唆が得られるかを考えたい。Bachman & Palmer (1996: 46) は、「(テストで) 用いられたタスクの特性は、常にある程度テスト得点に影響を与える可能性があり、測定しようとする能力の情報だけを提供するようなテストは実際には存在しない」と述べている。本研究では、タスクの特性の影響が発話の正確さで顕著に現れたと言える。インタビューテストよりもグループテストのほうが正確な発話がなされていたことは、インタビューテストにおいて正確さの評価が低かったとしても、グループテストでは高い評価を得る可能性があることを意味する。会話能力を評価する際には、対話者との関係や参加者の人数によって、パフォーマンスの正確さが変わる可能性があることを認識する必要がある。特定の対話者との会話における正確さであることを認識して得点を解釈するか、あるいは対話者の属性や人数が異なる複数のテスト形式を組み合わせる評価を行うことが望ましいだろう。

7. おわりに

本研究では、会話テストにおける受験者の発話を複数の指標を用いて分析したが、データ数が少なく結果を一般化できない点は本研究の限界である。今後、データ数を増やし、検証を重ねていきたい。また、テスト形式の影響は、受験者の日本語レベル・性格・動機など、様々な個人要因との相互作用を伴うものである。このような相互作用アプローチから、テスト形式の研究を行っていくことが望まれる。さらに、本研究では、受験者のパフォーマンスに焦点を当てて客観的な分析を行ったが、評定者が主観的に行う判定結果でも同じ結果が出るかどうかを確かめる必要があるだろう。

また、グループテスト実施においては、受験者の組み合わせも大きな問題の1つである。本研究では、コース中の学習者の様子を把握した上で組み合わせを決定したが、常にそのような機会が事前にあるとは限らない。グループテストの適切な使用を可能にするために、様々な要因から受験者の組み合わせの要因を検討することも、今後の課題である。

注

- (1) この他に、グループ/ペア形式のテストには、「教育実践を反映できる」「複数の受験者を同時に評価できる」「受験者のテスト不が小さい」等の利点があるとされている (Swain 2001; Fulcher 2003 等)。
- (2) 本研究では、テストの成績ではなく、会話テストで実際に受験者が行った言語使用という意味で「パフォーマンス」を用いる。
- (3) グループテストを含む4種類のスピーキングテストの得点間の相関を調べた Shohamy, Reves, & Bejarano (1986) では、グループテストの得点はその他のテスト (インタビュー、ロールプレイ、報告タスク) の得点と比較的相関が低いことが明らかにされた。また、Fulcher (1996) では、1種類のグループテストと2種類のインタビュー (絵の説明を行うものと、配布されたテキストについての議論) が分析され、グループテストのほうがインタビューよりも得点が高いという結果が出されている。
- (4) タスクの特性 (task characteristic) とは、タスク自体に備わっている特徴のこと。よく知っている話題かどうか、対話者がいるかどうか、等。実施条件 (task condition) とは、タスクが実施される方法のこと。準備時間の有無、視覚的サポートの有無、言語伝達の媒体、等。(Skehan, 1998: 174-177)
- (5) 52点以上であった。
- (6) 事前面談では、コース実施教師2名が、コース内容やスケジュールの説明を行い、コース応募動機、学習歴、学習ニーズなどを聞いた。
- (7) 受験者の組み合わせに関しては、ペア形式の会話テストにおいて受験者が既知関係であるほうが得点が高いこと (O'Sullivan, 2002)、性格 (内向的/外向的) が同質の者同士のほうが得点が高いこと (Berry, 2004) が示されている。
- (8) インタビューでは、インタビュアーの発話時間よりも受験者の発話時間のほうが多くなると予想されるため、テストの実施時間を10分ではなく約7分とした。
- (9) 日本語の従属節は、益岡 (1997) が名詞節 (補足語や主語)・連体節・連用節・並列節の4つに分類している。本研究での「節」の認定においては、この4分類を用いた。
- (10) ここでの正確さとは統語的正確さのことを指しており、相手や文脈に合った適切な発話かどうかという視点は含まれていない。本研究では先行研究との比較可能性を優先したが、今後、日本語に独特な待遇性の視点を取り入れた分析も必要であろう。
- (11) 対応のあるt検定を用いたが、出だしでの言い誤りのみ、正規性を仮定できなかったためウィルコクソン符号付順位和検定を用いた。

参考文献

- (1) Bachman, L.F. & Palmer, A.S. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- (2) Berry, V. (2004) *A Study of the interaction between individual personality differences and oral performance test facets*. Unpublished PhD thesis. London: Kings College, University of London.
- (3) Bonk, W.J., & Ockey, G.J. (2003) A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20 (1), 89-110.
- (4) Brown, A. (2003) Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25.
- (5) Foster, P., & Skehan, P. (1996) The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, 18 (3), 299-323.

- (6) Foster, P., Tonkyn, A., & Wigglesworth, G. (2000) Measuring Spoken Language: A Unit for All Reasons. *Applied Linguistics*, 21(3), 354-375.
- (7) Fulcher, G. (1996) *Testing tasks: Issues in task design and the group oral*. *Language Testing*, 13(1), 23-51.
- (8) Fulcher, G. (2003) *Testing Second Language Speaking*. Edinburgh: Pearson Education.
- (9) Iwashita, N., McNamara, T., and Elder, C. (2001) Can We Predict Task Difficulty in an Oral Proficiency Test? Exploring the Potential of an Information-Processing Approach to Task Design. *Language Learning*, 51(3), 401-436.
- (10) Lazaraton, A. (1992) The Structural Organization of a Language Interview: A Conversation Analytic Perspective. *System*, 20(3), 373-386.
- (11) 益岡隆志 (1997) 『新日本語文法選書 2 複文』 くろしお出版.
- (12) McNamara, T. (1995) Modelling Performance: Opening Pandora's Box. *Applied Linguistics*, 16(2), 159-179.
- (13) Ortega, L. (1999) Planning and Focus on Form in L2 Oral Performance. *Studies in Second Language Acquisition*, 21, 109-148.
- (14) O'Sullivan, B. (2002) *Learner acquaintanceship and oral proficiency test pair-task performance*. *Language Testing*, 19(3), 277-295.
- (15) Reves, T. (1982) The group-oral examination: a field experiment. *World Language English*, 1(4), 259-262.
- (16) Shohamy, E., Reves, T., & Bejarano, Y. (1986) Introducing a new comprehensive test of oral proficiency. *ELT Journal* 40(3), 212-220.
- (17) Skehan, P. (1996) A Framework for the Implementation of Task-based Instruction. *Applied Linguistics*, 17(1), 38-62.
- (18) Skehan, P. (1998) *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- (19) Skehan, P. (2001) TAsks and language performance ASsessment. In Bygate, M., Skehan, P., & Swain, M. (eds) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. London: Longman, 167-185.
- (20) Skehan, P., & Foster, P. (1999) The Influence of Task Structure and Processing Conditions on Narrative Retellings. *Language Learning*, 49(1), 93-120.
- (21) Swain, M. (2001) Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18(3), 275-302.
- (22) Taylor, L. (2000) Investigating the paired speaking test format. *University of Cambridge Local Examinations Syndicate Research Notes*, 2, 14-15.
- (23) Taylor, L. (2001) The paired speaking test format: recent studies. *University of Cambridge Local Examinations Syndicate Research Notes*, 6, 15-17.
- (24) 柳井久江 (1998) 『4Steps エクセル統計』 オーエムエス出版.
- (25) van Lier, L. (1989) Reeling, Wtihin, Drawling, Stretching, and Fainting in Coils: Oral Proficiency Interview as Conversation. *TESOL Quarterly*, 23(3), 489-508.
- (26) Wigglesworth, G. (1997) An investigation of planning tie and proficiency level on oral test discourse. *Language Testing*, 14(1), 85-106.

(2006年12月1日受理)