

教室内評価としての グループ・オーラル・テストの実施報告

— 評定者間信頼性と受験者の反応 —

堀川 有美、徳間 望

1. はじめに

日本語教育の重要な目標のひとつに他者とのコミュニケーション能力の育成が挙げられる。教室活動としてペアワークやグループワークが取り入れられてきており、そのような日本語コースの評価方法のひとつとして、学習者同士の話し合いを評価するグループ・オーラル・テスト（以下、グループテスト）がある。

本発表では、他者とのコミュニケーション能力育成を目標とした日本語会話クラスにおいて、到達度テストとしてグループテストを実施した結果について、評定者間信頼性と受験者の反応の2つの観点から報告する。

2. 先行研究

グループテストは、相互作用や会話管理の機会をテスト受験者に提供することができ、肯定的波及効果 (positive backwash) につながると考えられている。

実際にテストで受験者間の相互作用が引き出されていたかどうかを調べた研究に He & Dai (2006) がある。He & Dai は、大規模テスト (中国の大学生向けの英語スピーキングテスト CET-SET¹⁾) のディスカッション・セクションの談話を分析し、受験者間の相互作用を表す発話機能の出現頻度を調べた。その結果、受験者間の談話は相互作用的ではなく、各々がスピーチを行っているような会話になっていたことが報告されている。

それに対して、堀川 (2007) は、教室内テストとして実施されたグループテストを He & Dai (2006) と同じ発話機能を用いて調べた結果、受験者間の相互作用がより豊富に引き出されていたことを明らかにした。既知関係の受験者を対象とする教室内テストのほうが、受験者間の相互作用を引き出

しやすい可能性が示唆された。

しかし、堀川 (2007) では発話の分析を行ったにとどまっている。相互作用が受験者の口頭能力を反映していたかどうかを明らかにするためには、アンケートやインタビューにより受験者の反応を確認する必要がある。また、テスト得点は、評定尺度に基づいて評定者が決定するものであり、引き出された発話だけではなく評定尺度・評定者の影響も含めて分析する必要がある。教室内テストの場合、ふだんから学生と接し、学生に関する多くの情報を持っているコース担当教師が評定を行うが、そのことは評定結果に何らかの影響を与えるのだろうか。

受験者の反応および評定尺度・評定者の分析は、実施上の手続きに問題がなかったかどうかの確認にもつながり、教室内テストとしてのグループテストの実用性向上に貢献すると考えられる。

3. 研究目的

本研究は、教室内テストとしてのグループ・オーラル・テストの実施上の問題点を探るために、受験者の反応と評定者間信頼性を調べることを目的とする。研究課題は以下の2つとする。

研究課題 1 グループテスト受験時の受験者の反応はどのようなものか

研究課題 2 コース担当教師と、他の日本語教師の評定結果にちがいはあるか

4. 研究方法

4.1 対象者

韓国の大学の中級日本語会話 2 クラスの学生計 25 名を対象とした。専攻は日本語²⁾で、学年は 2~4 年生であった。この 2 クラス³⁾では、「自然なイントネーションを身につけ、相手や場面に応じた表現ができること、また、ロールプレイやストーリーテリ

ング、ディスカッションなどを通じて、自分の意見や考えをわかりやすく、まとまった長さで話せること」を学習目標としていた。授業の形態は、ペアワーク、または3人でのグループワークを基本とし、学生同士で話す活動が多く取り入れられていた。

4.2 テスト手続き

上記2クラスにおいて、2007年12月に期末試験としてグループテストを実施した。試験の1週間前にテストの方法と評価項目をクラスで説明し、練習を行った。

テストは、3～4人のグループで行った。学生の組み合わせは、学年・性別が偏らないように教師が事前に決定し、学生に知らせた。テストのテーマは、①学歴の影響、②田舎か都会か、③結婚の是非、④海外で働くことの4つの中から教師が選び、当日テストの際に学生に提示した。話し合いの時間は15分間とし、話し合いが始まってからは学生のみで会話を進行することとした。テストは録音録画された。

評価は、コース担当教師1名が評価尺度（稿末資料1）にもとづき行った。評価尺度の項目はコースの目標にもとづき選定され、尺度レベルはコースで目標とする到達度を基準に記述された。

4.3 データと分析方法

研究課題1では、テスト終了直後に実施したアンケートの5件法回答部分をデータとした。アンケートは、口頭能力テストの受験者の反応を調査したFulcher (1996) を韓国語に訳したものである（項目は表1を参照）。回答を1～5の数値に置き換え、平均値・標準偏差を求め分析を行った。考察の際には、アンケートの自由回答部分及びインタビュー資

料も補助資料とした。

研究課題2では、コース担当教師1名の評価結果とコース担当外の評価者7名の評価結果をデータとした。コース担当外の評価者はコースと関わりのない日本語教師である。テストの録画資料を見て評価をしてもらった。評価尺度の説明と練習を行った後、7グループ中2グループ（8名）分について評価者7名全員で評価を行った結果、 α 係数0.93という高い信頼性が確認されたため、残りの5グループ17名の評価は、評価者2名ずつで評価を行った。ずれがあった場合には平均値をコース担当外教師の得点とした。コース担当教師の評価結果と比較するために、平均の差の検定と相関分析を行った⁵。

5. 結果

5.1 研究課題1：受験者の反応

テストの全体的な感想に関する質問（2、12）で平均値が高く、標準偏差が小さいことから、大半の受験者がテストを肯定的に受け止めていたことがわかった。一方で、緊張するかどうか（質問4、5）は標準偏差が大きく、個人差が大きいことがうかがえる。

テストの手続きに関する質問では、グループの組み合わせ（質問9）やテストの時間配分（質問7）には特に問題がなかったが、話題の話しやすさ（質問9）は個人差、あるいは話題による差があったと推測される。

表1 受験者の反応に関するアンケート結果（n=25、質問6のみ n=24）

質問	平均	標準偏差	最小値	最大値
1 今日のテストでよくできたと思う	3.24	0.83	2	5
2 今日のテストは面白かった	4.48	0.51	4	5
3 今日のテストは難しかった	3.48	0.92	2	5
4 テスト前に緊張した	3.40	1.29	1	5
5 テストをするとき緊張した	3.32	1.14	1	5
6 テストで何をすればいいかわかった	3.75	0.74	2	5
7 テストの時間は短かった	2.96	0.89	1	4
8 今日のテストの話題は話しやすかった	3.20	0.87	2	4
9 グループの組み合わせは適切だった	4.44	0.65	3	5
10 他の日にテストを受けても同じ結果だったと思う	4.24	0.78	2	5
11 今日のテストは私の日本語の話す能力を正確に評価できると思う	3.88	0.93	2	5
12 今日のテストが気に入った	4.16	0.69	3	5

表2 コース担当教師と担当外評定者の評定結果

	平均	標準偏差	最小値	最大値
コース担当教師	3.8	0.5	2	4
コース担当外評定者	3.5	0.6	1.7	4

表3 項目別相関

発音	語彙	文法	流暢さ	ターンテーキング
.281	.735*	.754*	.663*	.735*

*p<.01

5.2 研究課題2：評定者間信頼性

コース担当教師の出した得点と、コース担当外評定者の出した得点を比較した。平均値はコース担当教師のほうが若干高かったが、平均の差の検定を行ったところ、有意差は得られなかった。つまり、評定の厳しさにはちがいがなかったと言える。

次に、相関分析を行ったところ、1%水準で有意に相関があることがわかった ($r_s=0.582$)。項目別の相関を確認したところ(表3)⁶、発音だけが有意差が得られず、他は比較的高い相関があることが確認された。したがって、発音以外の項目では両者が同じような評定を行っていたと言える。

6. 考察

全体的な反応は、「面白かった」「気に入った」など肯定的であり、グループテストによって話す能力を評価することを大半の受験者が肯定的にとらえていた。グループの組み合わせ、時間の長さ、テストの進め方については、特に不満や疑問は見られず、これらの面ではテスト実施手続きに問題がなかったと考えられる。

一方で、話題の話しやすさは、話題によって差があったことがわかった。自由記述やインタビューでも多くの受験者が話題の話しやすさについて言及しており、受験者が力を発揮できたと感じるかどうか、話題が影響していると考えられる。今回は、連続するグループの話題を変えるために、教師が話題を決定したが、受験者が話題を選べるようにしたり、話し始める前の準備時間をもう少し長くしたりする等の工夫が必要だと考えられる。

評定については、コース担当教師とコース担当外教師の評定の厳しさには違いがなく、また、両者には相関があったことから、評定の信頼性が確認された。コース担当教師は、1名でも信頼性の高い評価を行うことができていたと言える。それを可能にし

た要因として、ひとつには評定尺度を使用して基準を明示化したこと、もう1つは教室内テストの特徴と考えられるが、教師がふだんから学生と接しているため、その学生に関する多くの情報をもっているため、短時間でも的確に評価を行うことができていたことが考えられる。

ただし、項目別相関では、発音で有意な相関がみられなかった。その理由として、コース担当教師は学生の発音に慣れていたために、担当外の評定者とは異なる聞き方をしていただろう可能性が考えられる。また、コース担当外評定者は、録音・録画資料を用いて評価を行ったが、録画資料のほうの音質が悪く聞き取りにくい箇所もあったため、コース担当教師の評定とは異なってしまった可能性もある。

7. 今後の課題

今後の課題としては、話しやすさの違いが、受験者個人の違いなのか、話題の違いなのかを調べ、さらに、その話しやすさの違いが、テスト時のパフォーマンスやテスト得点に影響を与えるかどうかを調べることで、考えられる。

また、発音の評定において、コース担当教師と担当外の教師の間に相関がみられなかった理由を探ることも今後の課題とする。

注

1. College English Test-Speaking English Test の略。
2. 1名のみイタリア語専攻であった。
3. 2つのクラスは曜日時間が異なるが、同じ科目として設定されており、学習目標やシラバスは同じである。
4. ディスカッションにおいては意見内容の論理性や独自性等も評価のポイントになり得るが、今回のクラスの目標及びレベルでは評定範囲外のため、評定項目に含めなかった。
5. データの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と検定およびスピアマン

の順位相関係数の検定を行った。

6. 関連性はデータの特性により、分析を行えなかった。

参考文献

堀川有美 (2007). 「グループ・オーラル・テストにおける相互作用の発話機能」『言語文化と日本語教育』第 33

号, 21-30.

Fulcher, G (1996). Testing tasks: issues in task design and the group oral. *Language Testing*, 13(1), 23-51.

He, L. & Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, 23(3), 370-401.

ほりかわ ゆみ/国際交流基金日本語試験センター

yumih@mub.biglobe.ne.jp

とくま のぞみ/韓国外国語大学校

n.tokuma@gmail.com

稿末資料1 評価尺度

	発音	語彙	文法	流暢さ	関連性	ターン・テーキング
4	全体的に発音が明瞭で聞き取りやすく、理解を妨げない。 〔明瞭さ◎、理解◎〕	話題内容や文脈に合わせて、様々な語彙を使用している。〔豊富さ◎、適切さ◎〕	発話内容に合わせて、文法を効果的に使用している。軽微な誤用が見られることもあるが、意味の理解を全く妨げない。〔複雑さ○、正確さ◎〕	全体的に発話がスムーズで、発話が中断する際にはフィラーを適切に使用している。	他者の発話やディスカッションの流れに合った発話ができる。それまでの話と自分の発言内容を関連づけることができる。	適切なタイミングで自分から発言したり、他者の発話を促したりできる。〔自分◎、他者◎〕
3	発音が不明瞭な部分もあるが、会話の流れの中で理解を妨げない。 〔明瞭さ○、理解○〕	上級語彙などの不適切な使用が見られるが、多くの語彙は適切に使用している。〔豊富さ○、適切さ○〕	基本的にシンプルな文法を使用している。誤用があるが、ディスカッションを妨げない。〔複雑さ△、正確さ○〕	発話途中でポーズが入ったり、言いよどんだりすることがあるが、コミュニケーションの流れを阻害しない。	全体的に、他者の発話やディスカッションの流れに合った発話ができるが、ときどきポイントがずれていることがある。	積極的に自らターンをとることは少ないが、ターンをふられたときにきちんと応答できる。〔自分○、他者○〕
2	発音の不明瞭が目立ち、会話進行中に理解を妨げる。 〔明瞭さ△、理解△〕	基礎的な語彙を中心に、発話している。〔豊富さ△〕	シンプルな文法のみ用いて、表現している。意味の理解を妨げるような誤用が目立つ。〔複雑さ△、正確さ△〕	不自然に長いポーズや、言いよどみ等によって、コミュニケーションが阻害されることがある。	賛成・反対のような短い発話で、他者の発言に反応する。	ターン取得に失敗したり、他者からターンをふられたときに柔軟に対応できないことがある。他の話者の会話参加に対する配慮がほとんど見られない。〔自分△、他者△〕
1	全体的に発音が不明瞭で、理解困難。〔明瞭さ×、理解×〕	少数の基礎的な語彙のみを使用しており、意見を述べるのに語彙が不足している。〔豊富さ×〕	ごく基礎的な文法知識をもっているが、運用できない。〔複雑さ×〕	途切れ途切れになる話し方。ポーズや言いよどみが極端に多く、発話が途中で止まってしまう。	他者の発話やディスカッションの流れを考慮していない。	ターンをふられないと、ほとんど会話に参加できない。〔自分×、他者△〕