


# 博士論文審査及び最終試験の結果

審査委員（主査） 根岸雅史 

学位申請者 曲明

論文名 「スピーキングテストとしてのグループ・オーラルテストの妥当性について  
—インタビューテストとの比較を中心に—」

## 結論

曲明氏から提出された博士学位請求論文「スピーキングテストとしてのグループ・オーラルテストの妥当性について—インタビューテストとの比較を中心に—」について、論文審査と口述による最終試験の結果、審査委員会は全員一致して博士（学術）の学位を授与するにふさわしい研究であるとの結論に達した。

なお、審査委員会は、根岸雅史を主査に、副査として、田島信元教授（白百合女子大学）、高島英幸教授、三宅登之教授、吉富朝子教授を加えた5名で構成された。

## 論文の概要

本論文は、中国語教育におけるスピーキングテストに関する論文である。対面式スピーキングテストでは、従来インタビューテストが用いられるのが一般的であったが、3人以上の受検者同士で会話をさせるグループ・オーラステスト(以下、グループテストとする)が近年提唱され、利用され始めている。そこで、本論文では、この新しいグループテストの特性をインタビューテストと対比することで明らかにしようとしている。検証の方法は、大学生の中国語学習者にインタビューテストとグループテストを実施し、それらのテスト得点や発話にどのような違いがあるかを質的・量的に比較分析することを通して、その原因を探っている。最後に、分析結果とその考察をもとに、スピーキングテストへの示唆が述べられている。

本論文は全6章で構成されている。

第1章では研究背景、問題意識、研究の構成を説明している。言語教育の指導とテストは表裏一体の関係である。仮にスピーキングの指導だけ行っても、それに対する評価が行われなければ、学習者は話すことの練習をしなくなってしまう。その一方で、スピーキングテストで真っ先にイメージされるインタビューテストは、時間や費用がかかりすぎるために、実用性が低く、特に大人数の教室などでは、やりにくいという問題がある。そこで、こうした問題点を克服するために、英語教育では、グループテストが行われるよう

になってきている。日本の中国語教育分野においては、スピーキングテストに関する研究はほとんど行われていないため、本研究では、英語教育分野におけるインタビューテスト研究を参照しながら、日本人大学生の中国語学習者を対象にしたグループテストの妥当性について分析および考察を行う。

第2章は主に2つの先行研究の概観を行っている。1つはスピーキングテストで測る能力、いわゆる言語能力とは何かについて時系列的に言語の能力観概念を考察し、両テストのパフォーマンスを比較する際に均等に測れる能力の側面を決めている。その後、本研究の研究対象であるグループテストとインタビューテストの形式上の違いを考察し、テストの形式の違いが受験者のパフォーマンスに及ぼす影響について研究する意義を確認した。グループテストとインタビューテストの大きな違いは、受験者の「対話者」にあるとしている。つまり、インタビューテストの評価対象は母語話者または母語話者に近い言語熟達者である試験官と受験者との一対一の会話であるのに対して、グループテストは受験者同士の会話が評価対象である。この「対話者」の違いがテストパフォーマンスに何らかの影響を与える可能性があるとするれば、その違いを量的、及び質的に検証していく必要があるとしている。第2章で行ったもう1つ先行研究の概観はグループテストについての実証的な研究結果のまとめである。1980年代以降、グループテストについて、実証的な研究はいくつか行われたが、その多くはテスト得点に関する研究であった。テスト得点は、採点手続き、つまり、評価尺度や採点者の主観的な判断による人為的な要素などの影響を含んだものであり、誰がどのような評価尺度を用い、どのような評価方法で評価したかによって受験者の得点は変化し得るものである。そこで、本研究では、テスト形式の影響を明らかにするために、受験者の得点のみならず、彼らの発話内容についての研究も必要であると判断している。

第3章では、本研究の研究方法、テストの概要、両テストの採点方法、得点および発話の分析方法について説明している。本研究は、中国語を学ぶ日本人の大学生96人を対象にグループテストとインタビューテストを実施した。テストの形式の違いが受験者のパフォーマンスに及ぼす影響を明らかにするために、「両テストの形式の違いが得点にどのような影響を及ぼすか」と「両テストの形式の違いが発話にどのような影響を及ぼすか」と二つの研究課題を立てた。採点は分析的採点にし、発音、語彙、文法、流暢さ、コミュニケーションスキルと5つの下位項目を決めている。また、発話は、①発話の発話量、ターンの長さ、ターンの数、②発話の正確さ、流暢さ、複雑さ、③発話におけるコミュニケーション・ストラテジーの使用という三つの観点から分析するとしている。

本研究では、「得点の分析」を通して、テスト形式の影響の大きな傾向を明らかにし、「発話の分析」の部分では問いを深め、より細かく分析するために、テストの形式の効果のみならず、テストの形式と言語習熟度の組み合わせによる効果も分析している。なお、得点の分析では統計手法として相関分析とt検定を用い、発話の分析は二元配置分散分析

とカイ二乗検定を用いている。

第4章では、研究結果について述べている。まず、両テストの形式の違いが受験者の得点にどのような影響を与えているのかを、両テストの総合得点及び下位項目得点の相関分析と平均値の差の検定(t検定)を行い検証した結果、すべての項目において有意な相関が見られた。「総合得点」と「発音」では両テストの得点には高い相関があるが、他の各下位項目では中程度の相関が見られた。また、t検定の結果では、「発音」、「語彙」、「文法」の3項目は、両テストの得点の平均に有意差がなく、一方「流暢さ」では、インタビューテストよりグループテストの方が平均得点が有意に高く、「コミュニケーションスキル」項目では、インタビューテストよりグループテストの方が平均得点が有意に低いことが分かった。

次に、両テストの形式の違いが受験者の発話に及ぼす影響を検証した。まず両テストの形式の違いが発話量、ターンの数、ターンの長さに及ぼす影響を検証したところ、テストの形式の違いは発話量に影響しないが、ターンの数に影響を及ぼし、インタビューテストより、グループテストのターンの数の方が有意に多いことが分かった。1ターンの長さについては、テストの形式と言語習熟度の間交互作用の効果が認められ、成績の上位群にのみ有意傾向が見られ、グループテストよりインタビューテストの方が1ターンの長さが長かった。

更に両テストの形式の違いが発話の正確さ、複雑さ、流暢さに及ぼす影響を検証したところ、テストの形式の違いは発話の正確さの全指標、複雑さの全指標、流暢さの5指標のうちの3指標(「繰り返しの数」、「自己訂正の数」、「出だしの言い間違いの数」)においては影響を及ぼさなかった。テストの形式の違いが有意な影響を及ぼしたのは流暢さの二指標である「有声ポーズ数」と「無声ポーズ数」であった。テスト形式と言語習熟度の組み合わせの効果は「有声ポーズ数」に影響を及ぼし、成績の下位群にのみ有意傾向が見られ、グループテストよりインタビューテストの方が多かった。「無声ポーズ数」の項目において、テスト形式と言語習熟度要因の組み合わせの効果による影響は有意ではないが、テスト形式要因のみの効果は「無声ポーズ数」に影響を及ぼし、グループテストよりインタビューテストの方が「無声ポーズ数」が多かった。

最後に両テストの形式の違いがコミュニケーション・ストラテジーの使用に及ぼす影響を検証した結果、成績の下位群にのみ使用頻度に有意差があった。「直接アピール」の項目に関して、グループテストでの使用頻度は有意に低く、インタビューテストでは有意に高かった。「話題回避」項目においては、インタビューテストでの使用頻度は有意に低く、グループテストでは有意に高かった。

これらの研究結果から、両テストで測っている内容に同じ部分と異なる部分があったことが分かったとしている。スピーキングテストで使うテスト形式によりパフォーマンスが異なっていれば、そのテスト形式はスピーキングテストの種類に加えるべきであり、本研

究の研究結果はグループテストの実施意義を示したと主張している。

第5章では上述の研究結果をもたらした要因について考察している。まず得点分析の結果から両テストで測っている5つの下位能力のうち、発音、語彙、文法の3項目は、両テストの得点の平均に有意差がなく、一方、流暢さとコミュニケーションスキルの2項目においては、得点の平均に有意差が見られたため、テストの形式の違いによってこの2項目の難易度が異なると考えられる。インタビューテストよりグループテストでの1ターンの長さが短いことから、ターンが極端に短い場合には、文法の複雑さや談話レベルの評価が難しくなる可能性があり、グループテストで文法の複雑さや談話構成能力などを評価対象とする場合には、発話量、ターンの長さなどの発話特徴を考慮した評価尺度を作る必要があると主張している。

グループテストにおいて受検者がインタビューテストより流暢である理由として、次の理由を挙げている。インタビューテストでは、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は予測することが難しい。それに対してグループテストでは、受験者3人の協働で会話を作り上げていく中で、受験者自身が話題を提供したり、相手に質問したりするため、相手の発話内容を予測しやすいと考えられる。また、隠れた準備時間の確保も流暢さに繋がったと考えられる。

最後に、下位群の受験者は、インタビューテストでは「直接アピール」というコミュニケーション・ストラテジーを有意に多く使う一方で、グループテストでは「話題回避」を有意に多く使うことに関して、両テストにおける対話者の支援の有無がコミュニケーション・ストラテジーの使用状況の違いに繋がったと考えられる。インタビューテストでは、会話の相手は言語能力が高い母語話者であり、受験者はコミュニケーション挫折に遭遇する際に、「アピール」さえすれば、ほぼすべてのケースにおいてインタビュアーからの助けがあったため、「話題回避」をせずに会話を続けることができたと思われる。一方、グループテストを受ける時、「間接アピール」で言語運用能力の限界を示しても、助けしてもらえない時「話題回避」または「伝達回避」になるケースが多かったとしている。この問題を解決するには、成績の下位群の学生たちにさまざまな発話文脈においても支援をもらえるように「直接アピール」の言い方を始め、また「対話者に対して理解の可否の確認」や「簡単な言い換え」などさまざまなコミュニケーション・ストラテジーの使用を学生たちに指導することを提案している。言語習熟度の低い学習者だからこそ、このような教育上の工夫がスピーキング能力の向上に繋がると主張している。

第6章では、これまでの研究のまとめを行い、本稿の示唆と限界について述べている。特に、スピーキングテストの改善、指導と評価の一貫性、中国語教育におけるコミュニケーション教育への示唆について詳細に論じている。最後に、インフォームド・アセスメントという概念を提唱し、教師と学習者の間で、評価の意図や目的を共有することの重要性を強調している。本研究はグループテストとインタビューテストの形式が受験者のパフォ

パフォーマンスに及ぼす影響を明らかにすることによって、テストの目的に合わせて適切なテストの形式を選択することの意義と根拠を示している。

## 審査の概要及び評価

高い評価を与えられる点は以下の4点である。①中国語教育におけるスピーキングテスト研究は発展途上にある中で、先駆的な研究を行った点。特に、グループでのスピーキングテストは採用可能であれば興味深いので、このような検証を行なったのは意義があった。②スピーキングテストにおける、インタビューテストとグループテストの違い、とりわけ、構成概念に関わる違いを実証的に示した点。③研究目的に応じて様々な統計手法を適切に使っただけでなく、首尾一貫した全体像を描きながら、その分析結果を解釈している点。④中国語教育におけるスピーキングテストにおけるテスト方法の選択に関して、有意義な示唆が得られた点。

各審査委員より疑問もしくは批判として指摘のあった改善の余地のある点は以下の諸点に集約できる。(1)日本語の表記上の誤りや不自然さ、同様の内容が複数箇所で繰り返されているなど、主に外形的な問題があった、(2)受験者についての中国語学習履歴や中国語の熟達語レベルなどが充分には記述されていなかった、(3)中国語の分析的採点に用いられた「リズムとイントネーション」は、単に「中国語らしさ」というようなものであり、言語学的な厳密さに欠けるのではないか、(4)コミュニケーション能力モデルとしてBachmanモデルを引き合いに出しているが、分析的採点の評価基準はそれを反映したものになっていない、という点。

(1)と(2)の点については、最終版の提出にあたり、必要な修正を施すこととした。(3)の点には、具体例を挙げて、「リズムとイントネーション」上の問題がどのように「中国語らしさ」を損なうかについての説明があった。また、(4)の点に対しては、Bachmanモデルを参照しながらも、インタビューテストとグループテストを共通に比較できる観点のみを採用した、との説明があった。口述試問では、審査員からのこうした疑問や批判点に対して、若干の課題はあるもののおおむね的確に答えており、それらの課題を大きく上回る形で、研究分野に対する貢献が高く評価された。よって審査委員会は全員一致して冒頭に述べた結論に達した。