

多肢選択式リスニングテストの問題文と選択肢の提示時期がテストパフォーマンスに与える影響

神奈川県／神奈川県立伊勢原高等学校 教諭 柳川 浩三
申請時：神奈川県／神奈川県立小田原城内高等学校 教諭

概要

多肢選択式リスニングテストの受験者のパフォーマンスは、問題文と選択肢の提示時期（項目様式）により影響されることが示された。問題文と選択肢の両方が本文を聞く前に提示される場合、受験者の正答数は多くなる場合があり、この傾向は受験者のリスニング力にかかわりなかった。また、問題文と選択肢の提示時期によりリスニング力測定の精度にも違いが見られた。おおむね精度の高い順に、選択肢のみが事前に提示される様式、問題文のみが事前に提示される様式、問題文と選択肢の両方が事前に提示される様式の順になった。

多肢選択式リスニングテストにおいてどの項目様式をとるのが最も適切であるのか、コミュニケーション能力の測定の観点を含めて教育的示唆を提示する。

1 はじめに

実践的コミュニケーション能力の養成を掲げた学習指導要領、「英語が使える日本人の育成のための戦略構想」、センター試験でのリスニング試験の導入など、矢継ぎ早に打ち出される対策は英語による口頭コミュニケーション能力の育成が国家的な緊急の課題であることを示している。適切な語学教育指導は、しかし、正しい測定に基づいた評価があってはじめて可能になる。その意味では、英語コミュニケーション能力の基底をなす英語リスニング力の測定は、今まで十分に関心が払われてきたとは言い難い。その証拠に、テスト形式の違いがリスニングテストのパフォーマンスに及ぼす影響についての実証的研究は少ない。とりわけ、英検や TOEIC BRIDGE など

の大規模テストや教室でも多用されている多肢選択式（Multiple-choice Item Questions, 以下 MCQ）リスニングテストの項目様式（item format）の違いが、テストパフォーマンスに及ぼす影響については研究が少なく、解明しなければならない点が多い。

そこで、本研究では MCQ の項目様式である問題文と選択肢の提示時期がテストパフォーマンスに与える影響を、日本の高校生を受験者とした場合について明らかにする。それにより、テスト作成者や英語教師がリスニングテスト形式の違いがテストパフォーマンスに与える影響を広く認識し、それによって、妥当性の高いリスニングテストの作成に寄与できるものとする。

2 先行研究

2.1 テスト形式と特性

Bachman (1990) と Bachman and Palmer (1996) は、テスト形式（test method）をテストを作成する際の5つの構成要素（facet）の一つとしてとらえ、“characteristics (facets) of the expected response”としている。Brindley (1998) はテスト形式を、“nature of the assessment task”と呼び、Shohamy (1984) はある特性（trait）を測るための“specific procedure or technique”としている。

あるテスト形式上に現れる言語テスト得点は、「間接的で、不完全で、不正確で、主観的で、相対的である」（Bachman, 1990, p.50）。ある特定の言語テストの得点はその言語テストでの得点であって絶対的なものではない（Cohen, 1984; Bachman, 1990）。言い換えれば、テスト得点はテスト形式によって変

わるものである (Bachman and Palmer, 1996; Watanabe, 1996; Brindley, 1998; 根岸, 2002)。つまり、言語テストの得点は、「そのテスト形式とそのテストが測ろうとしている特性との関数である」(Bachman, 1990, p.113)。とりわけ、この傾向はリーディング能力やリスニング能力のような直接的に測定できない特性を測るときに顕著である。

そもそも、言語テストとは、「特定の特性が特定の手続きでもって測定される1つの統合体 (union) である」(Messick, 1989)。Shohamy (1996) は、特性を “normal performance variation” と呼び、テスト形式を “testing performance variation” とし、テストパフォーマンスはこの2つの変数の複合体であるとしている (p.149)。一方、Campbell and Fiske (1959) は、特性を “construct relevant variance”, テスト形式を “construct irrelevant variance” と呼んでいる。テスト形式による影響を少なくし、特性とテスト形式間に密接な関係を作り出すことが、妥当性の高いテスト作成に資するとしている (Cohen, 1984; Shohamy, 1984; Buck, 2001)。

2.2 テスト形式と間接的にしか測れない特性

本研究では、言語テストの中でもリスニングテスト形式とそのパフォーマンスの関係に関心があるが、リスニングとリーディングが類似した認知過程であり、一般にリスニングテストとリーディングテストの相関が高い (田辺, 2002) ことから、リーディングテスト形式がパフォーマンスに与える影響を検証することは、リスニングテスト形式がパフォーマンスに与える影響を考える時に参考になるだろう。

2.2.1 テスト形式とリーディング

中野 (1994) は、高校生を被験者として要約問題と多肢選択問題との相関を調査した。その結果、多肢選択問題と要約問題との相関は必ずしも高くなく、多肢選択問題で高得点の受験者は要約問題でも得点が高いとは限らないこと、更に、要約問題で高得点の受験者は多肢選択問題でも得点が高いことがわかった。Shohamy (1984) では、MCQ と記述式 (open-ended question) のパフォーマンスの相違を比較した結果、テスト形式効果が有意だった。Kobayashi (2002) は cloze (穴埋め式) テストと open-ended question, 要約の3つのテスト形式によるパフォーマンスの違いをテキストタイプの相違と

の関連で検証した。その結果、テスト形式、テキストタイプともテストパフォーマンスに対して有意な効果が見られ、テキストの構成が明確な場合 open-ended question や要約形式において習熟度の高い被験者ほど良い成績を収めた。

こうした、リーディングテスト形式によるパフォーマンスの変化についての実証研究は、リスニングテスト形式がパフォーマンスに与える影響にも有益な示唆を与えてくれる。しかし、リスニングでは余剰性 (redundancy) を多く含んだテキストを理解しなければならないこと (Freedle and Kostin, 1996, 1999他)、リスニングテストがリーディングテストよりも記憶の影響を大きく受けること (Freedle and Kostin, 1996, 1999)、リスニングはリーディングと同様の認知処理を行うがその方略的違いがあること (Lund, 1991) から、リスニングテスト形式によるリスニングテストパフォーマンスへの影響は、それ独自に検証する必要がある。

2.2.2 テスト形式とリスニング

Berne (1993) は、MCQ, open-ended question, cloze テストの3つのリスニングテスト形式によるパフォーマンスの相違をテキストタイプ (講義かインタビュー) と学習経験との関連で検証した。その結果、テスト形式によるパフォーマンスの相違は有意であり、また、テスト形式とテキストタイプとの間に交互作用が見られた。被験者は MCQ において最も高い得点を得たが、その理由として、正しい答えを自ら作り出すか思い出さねばならない open-ended question や cloze テストに比べ、MCQ では正解の選択肢を認識 (recognition) することさえできればよいからだとしている。一方、Yanagawa (2001) は、MCQ と短答式 (Short Answer Question, 以下 SAQ) の2つのリスニングテスト形式によるパフォーマンスの相違を検証した結果、MCQ の方が SAQ よりも有意に得点が高かったとしている。

2.3 テスト形式と受験者能力

テスト形式による影響は受験者の習熟度によっても変わる (Watanabe, 1996)。Shohamy (1984) と Cohen (1984) は、リーディングテストのパフォーマンスは被験者の習熟度が低いほどテスト形式に左右されるとしている。一方、Kobayashi (2002) は、

cloze テストと open-ended question, 要約の3つのテスト形式によるパフォーマンスの違いを検証した結果、リーディングテスト形式の影響は受験者の習熟度が高いほど大きいとしている。また、Wu (1998) は MCQ リスニングテストは習熟度の高い受験者には有利だが、低い受験者には不利だとしている。

2.4 問題文と選択肢の事前提示 (pre-view) の影響を示唆する関連研究

いくつかあるリスニングテスト形式の中でも、本研究は MCQ に関心がある。MCQ には、批判も多いが (Widdowson, 1978; Hughes, 1989; Weir, 1990; Thompson, 1995; Nissan, DeVincenzi, and Tang, 1996; Buck, 2001), 実用性と採点の信頼性の高さから多くのテストで用いられている (Berne, 1995; Buck 2001) からである。

MCQは、選択式解答形式 (selected response) (Bachman, 1990; Bachman and Palmer, 1996) であり、記述式解答形式 (constructed response/production response) と対をなす。MCQ 内でも項目様式 (item format) が複数存在する。項目様式とは、問題文 (stem) や選択肢の長さ、選択肢の種類や数などを具体化することである (Dunkel, Henning, and Chaudron, 1993) が、本研究では、テスト形式の下位要素としてそれを定義する。MCQ の場合、MCQ を構成する3要素は本文 (text)、問題文または幹 (stem)、そして選択肢 (answer option) であるが、本文を聞く前に問題文や選択肢が提示されているかどうかで項目様式が分かれる。そして、項目様式によってテストパフォーマンスが変わると考えられるのは、1) リスニング方略、2) リスニングテストの項目困難度、3) テスト受験方略の3つの関連分野研究からである。

2.4.1 リスニング方略

Oxford (1990) は、認知方略 (cognitive strategy)、メタ認知方略 (metacognitive strategy)、情動方略 (affective strategy) の3つの学習方略を提示したが、これはそのままリスニング方略にも当てはまる。認知方略はトップダウンとボトムアップの2つの認知処理過程に分かれ、メタ認知方略とは planning, monitoring, evaluating の3つに分かれる。情動方略とは、不安や緊張を和らげたり、励ますことである。

合理的な解釈 (“reasonable interpretation”) (Brown and Yule, 1984) に到達するためには、トップダウン処理とボトムアップ処理の統合的・相互補完的作用が欠かせない (Richards, 1990; Peterson, 1991他)。トップダウン処理とは、予備知識や文脈情報を活用して意味の理解をたどる過程であり、一方、ボトムアップ処理とは、文法や語彙に沿って逐語的に意味の理解をたどる過程のことである。そして、トップダウン処理が適切にできるかどうか、上手な聞き手とそうでない聞き手を分けるとされている (O'Malley, Chamot, and Küpper, 1989; Tsui and Fullilove, 1998; Taguchi, 2001)。このことは、上手な聞き手は前もって提示されている MCQ の問題文や選択肢から内容に関するヒントを得て、トップダウン処理を活性化できることをうかがわせる。一方、上手でない聞き手はこれができないのであろう。反対に、本文を聞く前に問題文を見ることが認知的な重荷になり、特定の命題だけにしか注意が向かなくなる結果、全体の理解が阻害されるというマイナスの影響を指摘する向きもある (Sherman, 1997)。いずれにしても、認知方略の存在は問題文と選択肢の提示時期がテストパフォーマンスに影響を与えることをうかがわせる。

次に、メタ認知方略の観点からは、pre-listening 活動 (Underwood, 1989) としてあらかじめ問題文や選択肢を提示しておけば、聞き手は焦点を絞って聞くことができるので事前の提示は必要かつ有効であるという主張が多い (Littlewood, 1981; Ur, 1984; Mendelsohn, 1995; Buck, 1995; Thompson, 1995; Vandergrift, 1999)。そして、それによってメタ認知方略の一つである planning が可能になる結果、パフォーマンスが上がるのではないかと推察される。

更に、情動方略の観点からは、本文を聞く前に問題文や選択肢を見ることは、どのような問題が出題されるかわからないという不確実性からくる不安や緊張を緩和するという立場と、反対に、問題文や選択肢を見ることで不安や緊張が増すとする相反する立場がある (Sherman, 1997)。いずれの立場も、情動方略がパフォーマンスに影響することを示唆する。

2.4.2 リスニングテストの項目困難度

Freedle and Fellbaum (1987) は、TOEFL (Test of English as a Foreign Language) 5回分の受験者10,325人について、リスニング問題中の一形

式である単文を聞いてそれと同じ意味を持つ英文を4つの選択肢の中から選ぶ問題（2004年時点ではこの問題はない）、計100問の解答を分析した。その結果、項目の難易度は単文と選択肢の関係に影響を受けることがわかった。具体的には、単文中の単語と正解の選択肢中の単語の重なり語数が少ない項目、または、単文中の単語と不正解の選択肢（錯乱肢）中の単語の重なり語数が多い項目では、項目難易度が上がる傾向にあることを指摘した。また、錯乱肢中に単文中の語彙から連想される語彙（例えば car-bumper, book-dictionary, doorbell-ringing など）が多いと、項目難易度は上がるとしている。更に、習熟度の低い受験者は単語の重なり具合の多い選択肢にまず目をつけ、その中から音韻的に単文に似た語句を含んだ選択肢を消去する方略をとる傾向にあるとしている。例えば、単文“He looked sick.”を聞いて、選択肢“He looked at the thickening fog.”という選択肢は選ばない、というようにである。

同様に、Freedle and Kostin（1996, 1999）は、1981年から1992年までに実施された TOEFL のうち、47回分のテストのリスニング問題中にある69のミニトーク（mini-talk）について出題された337項目について分析を行った。その結果、項目難易度に最も大きな影響を与える変数はテキストの内容が学術的かそうでないかというテキストの質と、選択肢の語彙的魅力度（lexical attractiveness）だった。選択肢の語彙的魅力度とは、本文と選択肢の内容語の重なり具合から、選択肢の正解らしさを数値化したものである。選択肢の語彙的魅力度と項目難易度との相関は、リスニングテストの場合、他の変数と比べて顕著に高かった（ $r=-.142$ ）。Freedle and Kostin（1996, 1999）は、リスニングテストにおいては、1つの変数、例えば選択肢の語彙的魅力度がその項目の難易度を大きく左右することは考えにくく、項目の難易度はむしろ、様々な変数間の相互作用であるとしている。しかし、項目難易度と選択肢の語彙的魅力度との高い相関は、選択肢の提示時期がリスニングテストのパフォーマンスに影響することを示唆している。

2.4.3 テスト受験方略

テスト形式や項目様式によって受験者がとるテスト受験方略（test-taking strategy）は変わる（Nevo, 1989）が、Cohen（1998）と Nevo（1989）は、受

験者が MCQ リーディングテストでとるテスト受験方略の一つとして、本文中や問題文中に出てきた語句を含む選択肢を選ぶ傾向にあるとしている。また、テスト受験方略を習得することで受験者のパフォーマンスは上がるという主張もある（Nevo, 1989; Banerjee and Clapham, 2003）。だとすると、MCQ リスニングテストにおいて問題文や選択肢の提示時期が変われば受験者がとるテスト受験方略も変わり、パフォーマンスも変わるであろうことは想像に難くない。そしてそれは、テストそのものの妥当性を脅かす要因にもなる（Cohen, 1984; Nevo, 1989; Farr, Pritchard, and Smitten, 1990）。

2.5 問題文と選択肢の事前提示効果の実証研究

2.5.1 短答式（SAQ）リスニングテスト

Buck（1990, 1991）と Sherman（1997）は、短答式（SAQ）リスニングテストの問題文の提示時期によるパフォーマンスの相違を検証した。しかし、どちらの実証研究によっても、本文を聞く前に問題文を見ることはパフォーマンスに有意な効果をもたらさなかった。

Buck（1990, 1991）が、英国留学中の日本人女性6人を被験者として、本文を聞く前に問題文を提示するグループと提示しないグループの2グループ各3人ずつに分け事前提示の影響を検証した結果、テスト結果に有意な差はなかった。この原因として Buck（1991）は、「聞き手は単に問題文をスキャンしているだけで、問題文が何を求めているかは意識していないか、問題文が提示されていてもその答えを探しながら聞くとは限らない」（p.84）からとしている。あるいは、説明的・論理的文章を聞くときは事前提示の効果はあっても、Buck（1990, 1991）が使用したようなストーリー性のある聞き手の興味を引くようなテキストを聞くときは効果が相当薄れるのではないかとし、聞くテキストの質に起因させている。

Sherman（1997）は、問題文の提示時期のみが異なる項目様式を4種類作成し、中レベルの習熟度を有するイタリア人大学生被験者78人を各様式に無作為に割り当てた。その4種類とは、本文を聞く前に問題文が提示されている様式（前版）、本文を聞いた後で問題文が提示される様式（後版）、1度本文を聞いた後で問題文が提示され、もう1度本文を聞いた様式（サンドイッチ版）、そして、問題文を提示する形式ではなく思い出すこと（recall）で理解度を測る様

式の4つである。その結果、サンドイッチ版が最も有意に成績が高かった。興味深いのは、アンケート調査から前版を後版よりも好む被験者が多かったにもかかわらず、2つの項目様式間にリスニングテストのパフォーマンスで有意な差はなかった点である。この結果から Sherman (1997) は、「問題文が聞き手に重要だと感じられるのは本文を聞いた後のことである」(in Buck, 2001, p.137) ので、事前提示はパフォーマンスに影響を与えないか、あるいは、2.4.1で述べたように、事前提示することのプラス面とマイナス面が相殺されたかのどちらかであろうとしている。この結果を受けて Buck (2001) は、「問題文を提示することは、テスト作成者が思っているほど聞き手を動機づけない」(p.137) としている。

2.5 Berne 多肢選択式 (MCQ) のリスニングテストの問題文と選択肢の事前提示がリスニングテストのパフォーマンスに及ぼす影響を検証した。 Berne (1995) は、スペイン語を学んでいる米国の大学生62人を、MCQの問題文と選択肢を事前提示する群、新出語彙の提示をする群、統制群の3つに分け、テストパフォーマンスに与える影響をMCQを使って検証した。その結果、問題文と選択肢を事前提示した群が統制群に比べて有意にテストの成績が良かった。また、問題文と選択肢を事前提示した群と、新出語彙を事前提示した群の理解度の差は有意傾向だった ($p < .054$)。Berne (1995) は、本文を聞く前に問題文と選択肢を見ることで聞き手が内容に関する仮説やスキーマを喚起でき、トップダウンの処理が活性化されたことを要因として挙げている (p.325)。また、Wu (1998) は、英国留学中の EFL (English as a Foreign Language) 中国人大学生10人を被験者として、MCQ リスニングテストを受ける時の被験者の解答プロセスを書き起こした。その結果、本文を聞く前に問題文と選択肢を見ることができ MCQ は、習熟度の高い受験者には有利だが習熟度の低い受験者には不利になるとしている (これは2.3でも述べた)。

Berne (1995) と Wu (1998) は、しかし、MCQの問題文と選択肢の提示時期によるパフォーマンスの相違を直接的に検証したわけではない。Berne (1995) では、使用したテストの問題数も10問と少なく、テストの信頼性も不明である。Wu (1998) は、

MCQ リスニングテストで受検者がどのような認知処理やリスニング方略を経て実際に解答に至っているかを明らかにしようとした点で興味深い質的研究であるが、回答者のプロトコルを提示しているにとどまっている。

Yanagawa (2003) は MCQ の問題文と選択肢の提示時期によるパフォーマンスの相違を高校3年生30人を被験者として被験者内 (repeated measure) 方式で検証した。問題文と選択肢の提示時期が異なる3つの項目様式とは、TOEIC タイプ (問題文も選択肢も本文を聞く前に提示)、英検タイプ (選択肢のみ本文を聞く前に提示し問題文は本文を聞いた後に音声で提示)、そして実験タイプ (問題文のみ本文を聞く前に提示し選択肢は本文を聞き終えた後に提示) の3種類である。その結果、平均点 (満点30点) は TOEIC タイプが最も高くなり、実験タイプが最も高い信頼性係数 (.87) を示した。また、項目様式がパフォーマンスに与える影響は受験者の習熟度が低いほど大きかった。更に、実験タイプが他の2タイプに比べて受験者のリスニング力をより適切に弁別していることが示された。

しかし、この研究では、3つの項目様式間でテストの難易度の均一性が立証されていなかったこと、TOEIC タイプと実験タイプでは問題文を文字で提示していながら英検タイプでは音声で提示したこと、TOEIC タイプの信頼性係数が低く ($\alpha = .56$) テスト得点の天上効果が出たと考えられること、サンプルが限定的だったこと、以上の方法上の問題点があった。本研究ではこの点を改良しつつ、分析に当たっては項目応答理論を併用することで、より一般化の可能な研究結果が得られるものと期待される。

3 | 研究課題と仮説

今まで見てきたように、多肢選択式 (MCQ) リスニングテストの問題文と選択肢の提示時期により、受験者のパフォーマンスは変わると考えられる。しかし、それについての実証研究は少なく、解明すべき点が多い。そこで、以下のような研究課題を立てた。

『多肢選択式リスニングテストの問題文と選択肢の提示時期は、日本人高校生 EFL 学習者のリスニングテストパフォーマンスに影響するか』

具体的には、以下の2点を検証する。

1. MCQ リスニングテストの問題文と選択肢の提示時期によって、リスニングテストのパフォーマンスに有意な差が生じるか。…研究課題1
2. もし有意な差が生じれば、その差は聞き手（受験者）のリスニング力によって変わるか（項目様式と受験者のリスニング力との間に交互作用があるか）。…研究課題2

上記の研究課題に対して、先行研究より以下の仮説を立てた。

研究課題1の仮説：有意な差がある

研究課題2の仮説：変わる（項目様式と受験者のリスニング力との間に交互作用がある）

4 | 方法

首都圏の4つの高校から高校2年生と3年生が被験者として参加し、実験は学校単位で既成のクラス間かグループ間の被験者間（Between-group）方式で行った。プリテストとして SLEP（Secondary Level English Proficiency）テスト（ETS: Educational Testing Service 作成）を用いて、各実験校のクラス間またはグループ間でリスニング力に有意な差がないことを確認した。その後、3つの異なった項目様式を各クラスまたは各グループに無作為に割り当てた。実験は2003年の12月から2004年の3月にかけて行った。

4.1 被験者 被験者均よりやや高い英語力を有していると思われる4つの高校から、2年生279名と3年生33名、計312名（男246名、女66名）である。内訳は、都内私立男子校普通科2年3クラス（男122名、女0名、計122名）、神奈川県立高校普通科外国語コース2年1クラス（男3名、女36名、計39名）、埼玉県内私立男子校普通科3年1クラス（男33名、女0名、計33名）、神奈川県立高校普通科2年3クラス（男55名、女63名、計118名）である。

4.2 プリテスト SLEP テストのリスニング問題の一部を実施した。SLEP テストを実施した理由は2つある。第1に、その結果を用いて同一学年から3クラス参加している実験校ではクラス間でリスニング

力に有意な差がないことを確認するため、1クラスのみでの参加の実験校ではリスニング力に有意な差のないグループを3つ作るため。第2に、SLEPテストの得点を被験者のリスニング能力の指標とし、それと本テストとの結果を比較するためである。SLEPテストを選んだ理由は、ETS 作成の標準化されたテストであることと、難易度が本研究の被験者に適当であることによる。

SLEP テストはリスニングテスト75問（45分）とリーディングテスト75問（45分）の2部構成になっているが、ここではリスニングテストのみ57問を利用した。75問全問を実施しなかった理由は、時間的な制約があったことと被験者の集中力の低下が懸念されたことによる。

57問のリスニングテストは、Part A から Part C の3つのセクションからなる。Part A は、単文を聞いてそれに合う絵を選ぶ問題（25問）、Part B は単文を聞いてそれと全く同じ英文を4つの選択肢の中から選ぶ問題（20問）、Part C は4台の車を配した地図を見ながら、発話者がどの車に乗りしているかを特定する問題（12問）である。

表1: 使用した本テスト

	問題文	選択肢
様式①	+	+
様式②	-	+
様式③	+	-
十事前提示		一事後提示

Yanagawa (2003) で使用した実験タイプのテスト ($\alpha = .87$) をもとに、問題文と選択肢の提示時期のみが異なる3種類のテストを作成した。問題文も選択肢も本文を聞く前に提示する様式（様式①）、選択肢のみ本文を聞く前に提示し問題文は本文を聞いた後に提示する様式（様式②）、問題文のみ本文を聞く前に提示し選択肢は本文を聞き終えた後に提示する様式（様式③）の3つである（表1及び資料参照）。

Yanagawa (2003) で使用した実験タイプのテストは、過去に英検準2級の本試験として実施されたリスニング問題を利用している。準2級を選んだ理由は、このレベルが高校2年生及び3年生を対象にしているからである（日本英語検定協会, 1994）。準2級のリスニングテストは Part 1 から Part 3 までの3

つのセクションに分かれており、Part 1は短文を聞いてそれへの応答として適切な英文を選ぶ問題（5問）、Part 2は短い会話を聞いてそれに関する質問の答えを選択肢の中から選ぶ問題（5問）、Part 3は短いトーク（モノログ）を聞いてそれに関する質問の答えを選択肢の中から選ぶ問題（10問）の計20問である（2004年2月現在、準2級のリスニング問題数は変更になっている）。このうち、Part 1は本研究で関心のある問題文（stem）が存在しないため使用しなかった。しかし、Part 1を除いた15問では項目数が少ないので、2回分の英検準2級のリスニング問題（1999年度第2回実施と2001年度第1回実施）を組み合わせ利用した。その結果、本実験で使うテストは、Part 2から計10問、Part 3から計20問の、計30問の項目で構成されることになった。1問につき1点、満点は30点とした。

様式①を作成するため、各項目の問題文は選択肢とともに問題冊子に印刷した。様式②を作成するため、問題文は選択肢が印刷されている紙の裏に選択肢と一緒に印刷し、本文の放送が終わるまで問題文はわからないようにした。様式③は Yanagawa (2003) の実験タイプ版テストと同一のものを利用した（資料参照）。選択肢は問題文が印刷されている紙の裏に問題文と一緒に印刷し、本文の放送が終わるまで選択肢はわからないようにした。様式②と③はテストの受け方に戸惑う被験者が出ることが予想されたため、項目1の前にサンプル問題を設けた。

4.1 手続き
同一実験校で1クラスのみでの参加の学校では、クラス内を SLEP テストの結果に基づきリスニング力に有意な差がないよう3グループに分け、様式①～③に無作為に割り当てた。一方、同一実験校で3クラスでの参加の学校では、SLEP テストの結果からクラス間でリスニング力に有意な差がないことを確認した後、様式①～③を無作為に割り当てた。実験は、2003年の12月上旬から2004年の3月上旬にかけて行い、各実験校単位でプリテスト実施後2日から1か月の間隔を置いて本テストを実施した。学校ごとによるプリテストから本テストまでの間隔の違いは、実験結果の妥当性に影響を与えないと判断した。

テスト時間は約25分とやや長く被験者が集中力をなくす心配があったため、項目15終了後に1分間の休憩を入れ、リラックスするための軽音楽を流して

被験者が集中力を維持・回復できるようにした。被験者は音楽が流れている間既に終了した問題の解答を書き直すことや指定されたページ以外を見ることは禁じられた。本文は全様式で一度だけしか流さなかった。

5 結果と分析
英語圏で1年以上の滞在経験のある被験者、プリテスト（SLEP テスト）と本テストのどちらか一方を受験していない被験者、受験態度からデータの信憑性が薄いとテスト監督者より報告のあった被験者、テスト冊子に誤植・落丁のあった被験者、得点7点以下の被験者（4肢択一なので満点30点のテストで偶然にでも取れる点数が7点）は分析より除外した。その結果、分析対象となる被験者は計253名となった。

テスト実施の都合上、実験校単位で既成のクラス間あるいはグループ間でリスニング力に有意な差のない群に対し実験を行ったが、データの分析は学校の枠を取り払って行った。その方が、各項目様式における被験者データが増えるのでより一般化の可能なデータを提示できるからである。学校の枠を取り払った各項目様式ごとの被験者のプリテストの平均点を1要因の分散分析にかけたところ、有意な差はなかった（ $F(2, 250) = .36$ ）。なお、プリテストの信頼性係数は様式①受験者で.75、様式②受験者で.76、様式③受験者では.70だった。

表2：研究課題1（満点30点）と信頼性係数

	平均（標準偏差）	α
様式① (N=83)	16.5 (4.7)	.73
様式② (N=82)	15.3 (4.3)	.66
様式③ (N=88)	14.9 (4.0)	.62

仮説1では、MCQ リスニングテストの問題文と選択肢の提示時期（項目様式）によってリスニングテストのパフォーマンスに有意な差が生じるだろうと推測した。項目様式ごとの平均点と標準偏差は表2

の通りである。この結果を1要因の分散分析にかけたところ、5%水準で項目様式間で平均点に有意な差があり ($F_{(2, 251)} = 3.1$)、最小有意差 (LSD) 法を用いた多重比較の結果、様式①が様式③よりも有意に平均点が高かった。様式①と様式②間及び様式②と様式③間では平均点に有意な差は見られなかった。よって、仮説1は部分的に支持された。

5.2 仮説2 研究課題2Q リスニングテストの問題文と選択肢の提示時期（項目様式）によってリスニングテストのパフォーマンスに有意な差が生じるとすれば、その差は聞き手（受験者）のリスニング力によって変わるだろうと予測した。つまり、項目様式と受験者のリスニング力との間には交互作用があるだろうと。プリテスト（SLEP テスト）の結果に基づいて各項目様式の被験者を上位27%、中位46%、下位27%の3つに分け、各々の平均点と標準偏差を算出した（表3参照）。

■ 表3 項目様式別及びリスニング力別平均点
(満点30点)

	平均 (標準偏差)	上位	中位	下位
様式① (N=83)	16.5 (4.7)	21.1 (4.1)	15.9 (3.5)	13.0 (3.1)
様式② (N=82)	15.3 (4.3)	19.5 (3.3)	14.7 (3.7)	12.2 (2.6)
様式③ (N=88)	14.9 (4.0)	18.0 (4.0)	14.5 (3.3)	12.5 (3.2)

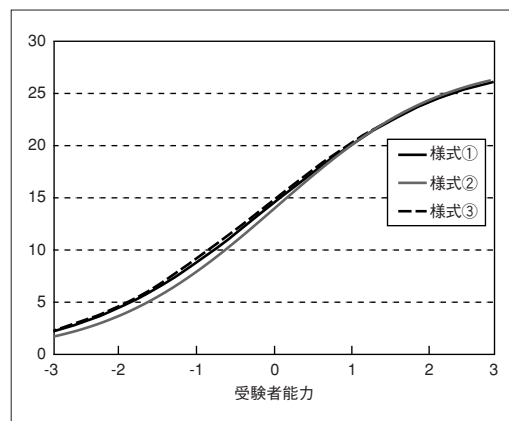
表3の結果を3×3の分散分析にかけた結果、項目様式とリスニング力の主効果は1%水準で有意であったが（各々 $F_{(2, 244)} = 4.7$, $F_{(2, 244)} = 78.2$ ）、項目様式とリスニング力の間には交互作用は見られなかった。よって、仮説2は支持されなかった。

次に、項目応答理論の1パラメータラッシュモデルを採用し、パラメータの推定には PROX 法（中村, 2002）を用いて項目様式間の困難度の相違を受験者のリスニング力との関連で分析した。具体的には、テスト特性曲線を描き各項目様式の困難度とリスニング能力の関係を示した。その際、モデルに適合していない（ミスフィット）項目4と項目19の2項目、及びミスフィットの受験者は分析対象としな

った。よって、項目数は28、様式①の分析対象者は68人、様式②は72人、様式③は76人となった。ラッシュモデルを採用した理由は、主として被験者数とそのモデルに適合していることとその簡便さによる（大友, 1996）。

その結果、特に平均的なリスニング力に達しない受験者（図1中の受験者能力が0に満たない者）にとっては、様式②が様式①及び③に比べてわずかに難しいと思われることを除けば、テストの困難度は受験者のリスニング力にかかわらず、項目様式間で同じであると言ってよいことが示された。これは、項目様式とリスニング力の間には交互作用がなかったとする前述の古典的テスト理論に基づく分散分析の結果と一致する。

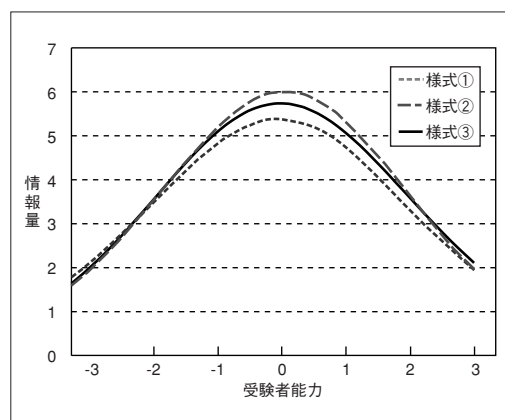
▼ 図1：テスト特性曲線



5.3 項目様式による測定精度の比較
あるテストの測定精度は受験者能力とに数値化したものがテスト情報関数である。したがって、ある能力を持つ受験者に対してどの項目様式が最も適切かをテスト情報関数は提示してくれる。図2は、様式①～③のテスト情報関数をグラフ化したものである。ただし、テスト特性曲線を描いた時と同様、ミスフィットとされた項目及び受験者を除いて再計算をした。

図2から、おおむね②が最も測定の精度が高く、次いで様式③、様式①の順であることがわかる。特に、この傾向は -1～1.5 のリスニング力を持つ受験者に対して顕著だった。

▼ 図2：テスト情報関数



■ 表4：SLEP テストとの相関

様式①	.66
様式②	.66
様式③	.44

6 考察と研究の限界

6.1 研究課題1

本実験の結果、項目様式①が様式③よりも正答数が多かったことから、問題文と選択肢の提示時期（項目様式）によってリスニングテストのパフォーマンスは異なる場合があることが示唆された。様式①では受験者が本文を聞く前に問題文と選択肢の両方を見ることで、多くの文脈情報を得ることができたため、受験者のトップダウン処理がより効果的に行われ理解が促進されたのではないと思われる。それに対して、様式③では受験者が問題文しか事前に見ることができないので、様式①の受験者ほどには文脈情報を得ることができずに、トップダウン処理が様式①ほどは効果的に行われなかったのではないかと。

しかし、そうだとするならば、様式①は様式③だけでなく様式②（受験者は選択肢のみ事前に見ることができる）よりも正答数が有意に多くなっているべきではない。しかし、結果はそうっていない（様式①と様式②とでは正答数の有意な差はない。表2参照）。このことは、2つのことを示唆している。第1に、問題文と選択肢を事前提示している様式①が、選択肢のみ事前提示している様式②よりも必ずしも提供する文脈情報が多いとは限らないということ。

言い換えれば、選択肢が事前提示されていれば（様式②）、問題文が添えられている（様式①）かどうかは、被験者が得る文脈情報の量にさほど影響しないのではないかとということである。

第2に、本文を聞く前に得られる文脈情報の豊富さとは別の要因が、テストパフォーマンスに関与していると考えられること。その別の要因とは、おそらく様式②にあって様式③にないもの、あるいは、様式③にあって様式②にないものだろう。具体的にはまず、様式に対する慣れがある。英検などで頻繁に用いられる様式②には慣れている受験者が多かったと思われるが、問題文だけが本文を聞く前に提示される様式③には慣れていない受験者が多かったのではないかと。このことは、テスト形式や項目様式に対する慣れが、テストパフォーマンスに影響することをうかがわせる。Shohamy（1984）や Cohen（1984）が指摘するように、特にいまだ習熟度の低い高校生については、テスト形式に対する慣れがパフォーマンスに与える影響は少なくないと思われる。

次に、認知負荷量の問題である。本文を聞く前から選択肢の中に正解が提示されている様式②では、正解の選択肢を選ぶことだけに集中すればよいのに対し、様式③では、少なくとも本文を聞く前と聞いている間は、受験者は自分で解答を考え出さねばならない。その意味では、様式③は記述式解答形式の色彩も帯びており、認知負荷量が様式②よりも多いと思われる。

最後に、情動的要因である。本文を聞く前から選択肢の中に正解が提示されている様式②と比べた場合、少なくとも本文を聞く前と聞いている間は、自分で解答を考え出さねばならない様式③では、受験者の不安や緊張感は大いだろう。上記の考察から、問題文や選択肢の事前提示による文脈情報の多少に加えて、テスト慣れの程度、認知的負荷量、情動的要因がテストパフォーマンスに影響を与えることがうかがえる。

もちろん、本研究の結果とそれに基づくそうした考察を一般化するには、今後の研究を待たねばならないだろう。しかし、テスト作成者や英語教師が学習者のリスニング力を測定する際、どの項目様式をとるべきかについては様々な要因を考慮しつつ、慎重を期する必要があることを本研究の結果は示している。

6.2 研究課題2

本実験の結果、MCQ リスニングテストの問題文と選択肢の提示時期（項目様式）の相違によるリスニングテストのパフォーマンスの差は、聞き手（受験者）のリスニング力にかかわりなかった。このことは、古典的テスト理論の考え方に基づく2要因分散分析の結果からも、項目応答理論に基づくテスト特性曲線からも裏づけられた。では、なぜ仮説は支持されなかったのだろうか。ここでは方法上の問題点を2つ指摘したい。第1に、各項目様式のテストの信頼性係数が様式①=.73, 様式②=.66, 様式③=.62と十分に高くなかったことが挙げられる。テストの信頼性係数はテストの内的一貫性を示す指標であり、信頼性係数の低さは各項目様式内で測定している要素が必ずしも同様ではなかったことを示している。

第2に、表4から、プリテスト（SLEP テスト）と本テストの相関は必ずしも高いとは言えず、両テストが異なる構成概念（construct）で構成されていた可能性が高いことが挙げられる。その結果、受験者能力（SLEP テスト）と項目様式（本テスト）との間に交互作用が出にくかったのではないかと推察される。リスニングテストのパフォーマンスと受験者能力との間に交互作用がないという結果は、英語教師の日頃の体験的感覚とは異なる。この点は、今後更なる検証が必要であろう。

実験結果より、様式①の正答数は様式③に比べて有意に多く、この傾向は受験者のリスニング能力にかかわりないということが示された。ところが、図2が示しているように、様式①の測定精度は様式②や③に比べておおむね低い。このことから、様式①は受験者の真のリスニング力以上の正答数を出す傾向が強いことをうかがわせる。

しかし、このことだけで様式①の使用が避けられるべきではないだろう。それは、前述したように、本研究では過去に英検の本試験として実際に実施された問題をもとにしてテスト問題を作成したため、英検の項目様式と同一様式をとらない様式①や③では、テストの妥当性や適切性に関して多少の不利な結果が出るかもしれないことは予想されたからである。英検の様式をもとにすることなく、全く新たに項目を起こし様式①や③を作成したならば、異なる結果になった可能性はある。ただ、その場合、相当な手間と時間を要することからくる研究の経済性と実行可能性の問題を解決しなければならない。また、

英検の過去問題を素材にテストを作成することが研究結果の妥当性を著しく損なうものでもない。これらの理由により、本研究の方法をとった。

逆に言えば、英検で採用している項目様式②（本文を聞く前に選択肢のみが提示される様式）が本研究で検証した3つの項目様式の中では測定精度が最も高いという結果が出たことから、その適切性がはからずとも実証される結果となったことは興味深い。実際のコミュニケーションの場面では、しかし、この様式のように発話を聞く前から選択肢が提示されていることはなく、また、用意された選択肢の中から1つを選んで応答することもない。従って、測定の道具としては様式②の精度は高くとも、コミュニケーション能力の一部としてのリスニング力を測定する道具としては、様式②はその妥当性に疑問があると言わざるを得ない。

様式③は3つの様式の中では、唯一、選択肢を事前に提示しない日々のコミュニケーション活動に近い様式であり、その意味では望ましい様式であろう。しかし、信頼性係数が.62（表2）、プリテストとの相関係数が.44（表4）とやや低い結果が出たのは残念である。信頼性係数がYanagawa（2003）の.87と比べて低いのは、被験者のリスニング力の相違によるところが大きいと思われる。Yanagawa（2003）の被験者は高校3年生であるのに対し、本実験の被験者の多くが高校2年生であった。被験者のリスニング力がもう少し全体的に高ければ、テストの信頼性も上がったと思われる。

7 | まとめ

7.1 要約

本研究により、多肢選択式リスニングテストの受験者のパフォーマンスは、問題文と選択肢の提示時期（項目様式）により影響されうることが示された。問題文と選択肢の両方が本文を聞く前に提示される場合、受験者の正答数は多くなる場合があり、この傾向は受験者のリスニング力にかかわりなかった。また、問題文と選択肢の提示時期によりリスニング力測定の精度にも違いが見られた。おおむね精度の高い順に、選択肢のみが事前に提示される様式、問題文のみが事前に提示される様式、問題文と選択肢の両方が事前に提示される様式の順になった。

7.2 教育的示唆

テスト作成者や教師はリスニングテスト受験者のパフォーマンスは必ずしも受験者の絶対的能力を反映しないことを認識しておく必要がある。つまり、テストパフォーマンスはテスト形式や項目様式によってある程度は変わる傾向にあるということをである。このことを念頭に置けば、教師は、実際にテストを行う前に各項目様式の特徴を知り、テストの目的と受験者のリスニング力にあった項目様式を選択し、また、時に複数のテスト形式や項目様式を使用することの大切さをも理解するであろう。教師のこうした配慮により、テストの妥当性と信頼性は向上するはずである。

7.3 今後期待される研究

3つ挙げておきたい。第1に、項目様式による、テスト慣れの程度、受験者にかかる認知的負荷量、情動的負荷量の相違がテストパフォーマンスに与える影響を計量的に検証すること。

第2に、項目様式により正答数と測定精度の相違を生んだ要因を、個々の項目について語彙的魅力度 (Freedle and Kostin, 1996, 1999) と必要な情報 ("necessary information", Buck and Tatsuoka, 1998) の位置の観点から分析・検証すること。

第3に、サンプル数を増やして本研究の結果を再検証すること。

本研究により、多肢選択式リスニングテストにおける日本人高校生英語学習者のパフォーマンスは、

問題文と選択肢の提示時期により影響を受ける場合があること、そして、影響の程度は学習者のリスニング力にかかわりがないことが示唆された。具体的には問題文と選択肢の両方が本文を聞く前に提示される時、受験者の正答数は増える場合があることが示された。また、問題文と選択肢の提示時期によりリスニング力測定精度にも違いが見られた。以上のことから、多肢選択式リスニングテストにおいてどの項目様式をとるのが最も適切であるのか、今後、項目様式とパフォーマンスとの関係について更なる検証が待たれる。

謝 辞

最後に、この研究の機会を与えてくださった(財)日本英語検定協会、選考委員の先生方、とりわけ、メールを通じて研究全般にわたってアドバイスをくださった常磐大学教授の大友賢二先生に感謝いたします。次に、草稿段階で有益なご助言をいただいた早稲田大学教授の松坂ヒロシ先生に深く感謝いたします。更に、データの分析に関してご助言をいただいた常磐大学助教授の中村洋一先生、データ収集にご尽力をいただいた神奈川総合高校の鈴木栄先生、神奈川県立西湘高校の鈴木達也先生、早稲田大学助手の山口高嶺先生、成城高校講師の小林玲浩先生にも深く感謝いたします。最後に、研究の構想段階から有益な批判とコメントを授けてくれたケンブリッジ大学研究助手の Tony (A.B. Green 氏) に、記して感謝します。

参考文献 (*は引用文献)

- * Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- * Bachman, L. and Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Banerjee, J. and Clapham, C. (2003). The TOEFL CBT. *Language Testing*, 20, 1, 11-123.
- Berne, J. (1993). The role of text type, assessment task, and target language experience in L2 listening comprehension assessment. ERIC ED 358737
- Berne, J. (1995). How does varying pre-listening activities affect second language listening comprehension? *Hispania*, 78, 316-329.
- * Brindley, G. (1998). Assessing Listening Abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- * Brown, J. and Yule, G. (1984). *Teaching the Spoken*

Language: An approach based on the analysis of conversational English. Cambridge: Cambridge University Press.

Buck, G. (1990). Testing second language listening comprehension. Unpublished PhD dissertation, University of Lancaster.

* Buck, G. (1991). The testing of listening comprehension: an introspective study. *Language Testing*, 8, 1, 67-91.

Buck, G. (1995). How to become a good listening teacher. In Mendelsohn, D. and Rubin, J. (Eds.), *A Guide for the teaching of second language listening* (113-131).

* Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

* Buck, G. and Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening

- test. *Language Testing*, 15, 2, 119-157.
- *Campbell, D.T. and Fiske, D.W.(1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cohen, A.(1984). On taking language tests: What the students report. *Language Testing*, 1, 70-81.
- Cohen, A.(1998). Strategies and processes in test taking and SLA. In Bachman, L. and Cohen, A. (Eds.), *Interfaces between second language acquisition and language testing research* (90-111).
- Dunkel, P., Henning, G., and Chaudron, C.(1993). The assessment of an L2 listening comprehension construct: A tentative Model for test specification and development. *The Modern Language Journal*, 77, 2, 180-191.
- Farr, R., Pritchard, R., and Smitten, B.(1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27, 3, 209-226.
- Freedle, R. and Fellbaum, C.(1987). An Exploratory Study of the Relative Difficulty of TOEFL's Listening Comprehension Items. In Freedle, R. and Duran, R.(Eds.), *Cognitive and Linguistic Analyses of Test Performance*(162-192). New Jersey: Ablex Publishing corporation.
- Freedle, R. and Kostin, I.(1996). The prediction of TOEFL listening comprehension item difficulty for minitalk passages: Implications for construct validity. *Research Reports*, 56. Princeton, New Jersey: Educational Testing Service.
- Freedle, R. and Kostin, I.(1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16, 1, 2-32.
- Hughes, A.(1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kobayashi, M.(2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing*, 19, 2, 193-220.
- Littlewood, W.(1981). *Communicative Language Teaching: An Introduction*. Cambridge: Cambridge University Press.
- Lund, R.J.(1991). A Comparison of Second Language Listening and Reading Comprehension. *The Modern Language Journal*, 75, 2, 196-204.
- Mendelsohn, D.(1995). Applying learning strategies in the second/foreign language listening comprehension lesson. In Mendelsohn, D. and Rubin, J.(Eds.), *A Guide for the teaching of second language listening* (132-150).
- *Messick, S.(1989). Validity in R.L. Lion.(Eds.), *Educational Measurement* (3-103). Third edition. New York: American Council on Education and Macmillan.
- 中村洋一. (2002). 『テストで言語能力は測れるか～言語テストデータ分析入門～』. 東京: 桐原書店
- 中野幸子. (1994). 「高校生の読みにおける要約問題と多肢選択問題との相関」. *STEP BULLETIN* vol. 6, 30-33. 東京: 日本英語検定協会
- 根岸雅史. (2002). 「英語力はどう測定されるか——建設的な英語教育論のためのテストの常識」. 『英語展望』(ELEC). 109, 12-15.
- Nevo, N.(1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6, 2, 199-215.
- 日本英語検定協会. (1994). 準2級ガイド. 東京: 日本英語検定協会.
- Nissan, S., DeVincenzi, F., and Tang, K.(1996). An Analysis of Factors affecting the difficulty of dialogue items in TOEFL listening comprehension. *Research Reports*, 51. Educational Testing Service.
- O'Malley, J., Chamot, A. and Kupper, L.(1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10, 4, 418-437.
- 大友賢二. (1996). 『項目応答理論入門』. 東京: 大修館書店.
- Oxford, R.(1990). *Language Learning Strategies: What every teacher should know*. Boston: Heinle & Heinle Publishers.
- Peterson, P. W.(1991). A Synthesis of Methods for Interactive Listening. In Celce-Murcia.(Eds.), *Teaching English as a Second or Foreign Language* (106-122). Boston, MA: Heinle & Heinle Publishers.
- Richards, J.C.(1990). *The Language Teaching Matrix*. Cambridge: Cambridge University Press.
- *Sherman, J.(1997). The effect of question preview in listening comprehension tests. *Language Testing*, 14, 2, 185-213.
- *Shohamy, E.(1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 2, 147-170.
- *Shohamy, E.(1996). Competence and Performance in Language Testing. In Brown J. et al.(Eds.), *Performance and Competence in Second Language Acquisition*(138-151). Cambridge University Press.
- Taguchi, N.(2001). L2 Learners' Strategic Mental process during a Listening Test. *JALT Journal*, 23, 2, 176-201.
- 田辺博史. (2002). In 武井編著. 『英語リスニング論』. 東京: 河源社
- Thompson, I.(1995). Assessment of second / foreign language listening comprehension. In Mendelsohn, D. and Rubin, J. (Eds.), *A Guide for the teaching*

of second language listening (31-58). San Diego: Dominie Press.

Tsui, A.B.M. and Fullilove, J. (1998). Bottom-up or Top-Down Processing as a Discriminator of L2 Listening Performance. *Applied Linguistics*, 19, 4, 432-451.

Underwood, M. (1989). *Teaching Listening*. London: Longman.

Ur, P. (1984). *Teaching Listening Comprehension*. Cambridge: Cambridge University Press.

Vandergrift, L. (1999). Facilitating second language listening comprehension: acquiring successful strategies. *ELT Journal*, 53, 3, 168-176.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 3.

Weir, C. J. (1990). *Communicative Language Testing*.

New York: Prentice Hall

Widdowson, H. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Wu, Yu'ian (1998). What do tests of listening comprehension test? — A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15, 1, 21-44.

Yanagawa, K. (2001). Validity of Multiple Choice Questions: The Effects of Answer Options on Listening Comprehension Tests for Japanese High School Learners of EFL. *Proceedings of the 6th Conference of Pan-Pacific Association of Applied Linguistics*, 129-139.

Yanagawa, K. (2003). Relative Difficulty of Three Multiple Choice Listening Comprehension Item Formats for Japanese High School Learners of EFL. *The Bulletin of the Graduate School of Education of Waseda University*, 10, 2, 1-14.

資料

項目様式①

Q Where does the woman's brother work?

- (a) At a restaurant
- (b) At a TV station
- (c) At a cooking school
- (d) At a magazine publisher

項目様式②

表側 (4つの選択肢のみ印刷されている)

Sample

- (a) At a restaurant
- (b) At a TV station
- (c) At a cooking school
- (d) At a magazine publisher





"Go on to the next page." という指示が聞こえたら、次のページを開いて解答する (指示があるまでは進まない)

裏側 (問題文と4つの選択肢が印刷されている)

Sample

Where does the woman's brother work?

- (a) At a restaurant
- (b) At a TV station
- (c) At a cooking school
- (d) At a magazine publisher

項目様式③

表側 (問題文のみ印刷されている)

Sample

Where does the woman's brother work?





"Go on to the next page." という指示が聞こえたら、次のページを開いて解答する (指示があるまでは進まない)

裏側 (問題文と4つの選択肢が印刷されている)

Sample:

Where does the woman's brother work?

- (a) At a restaurant
- (b) At a TV station
- (c) At a cooking school
- (d) At a magazine publisher