

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 10 日現在

機関番号：27401

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23520692

研究課題名(和文)多肢選択式リスニングテストにおける選択肢の数の検証

研究課題名(英文)The Effects of the Number of Options in Multiple-Choice Listening Tests

研究代表者

飯村 英樹(IIMURA, Hideki)

熊本県立大学・文学部・准教授

研究者番号：30382831

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：多肢選択式リスニングテストにおける選択肢の数の影響を、問題形式ごとに検証した。3択と4択を比較した結果、テストの信頼性や弁別力には大きな差はなかったが、項目困難度に関しては違いが見られた。描写問題のような比較的単純な形式(写真と一致する英文を選ぶ)の場合は、3択と4択で困難度は変わらないが、TOEICのPart 4のような認知的負荷の高い形式(本文がモノローグで、複数の質問文に答える)では3択の方が、困難度が低くなるという結果になった。研究全体の調査結果から、リスニングテストにおける最適な選択肢数を検証するためには、問題形式および錯乱肢について詳細な分析が必要であることが示唆された。

研究成果の概要(英文)：This study investigated the effects of the numbers of options (four-option vs. three-option) on different listening tasks. The data were analyzed from the perspective of item difficulty, discrimination, reliability, and distractor performance. The results indicate that the number of options had no notable impact on item difficulty in picture and dialogue tasks, but had a significant effect in the monologue task. The results also suggest that the number of options had no great impact on item discrimination and reliability in all tasks, and that the number of items containing all discriminating distractors was different between the three tasks. The overall findings provide evidence that task difference and distractor analysis are important elements in the study of optimal number of options and the development of multiple-choice listening tests.

研究分野：英語教育

キーワード：リスニング 多肢選択式 問題形式 錯乱肢 妥当性 項目困難度 弁別力 信頼性

1. 研究開始当初の背景

(1) リスニングテストにおける選択肢数

リスニングテストにおいて、多肢選択式は採点の容易さや信頼性の高さから幅広く使われている。実際に、大学入試センター試験や英検、TOEIC など国内外の主要なテストはすべて多肢選択式である。

多肢選択式の特徴は、その名の通り複数の選択肢の中から正解を選ぶところにあり、そのため、当て推量による正解の可能性を否定できない。当て推量の影響を避けるために、多くの場合は、正解肢1つに錯乱肢と呼ばれる誤答を3つ加えて、4択にする手段を講じる(表1)。なぜこれほどまでに4択が採用されているのだろうか。

表1 リスニングテストにおける問題形式別の選択肢数

問題形式	TOEIC	センタ ー	英検準 2級	英検2 級	英検準 1級
写真	4択	4択			
応答	3択	4択	3択		
対話	4択	4択	4択	4択	4択
説明文	4択	4択	4択	4択	4択

実際のところ、4択が他の選択肢数に比べて特に優れているという明確な理由は見当たらない。日本におけるテストのガイドラインを定めた『テスト・スタンダード』においても、最適な選択肢数は、実用面からみて4択かあるいは5択が用いられているとの記述にとどまっている(日本テスト学会, 2007)。

(2) 3択の優位性

4択が圧倒的に採用されている現状に対して、言語テスト理論の分野では、4択と3択を比較した場合、3択の方が優れているとの指摘がされている。その理由としては、3つの錯乱肢があっても、実際に弁別力を持ち機能するのは1つないし2つにとどまり、3つ目の錯乱肢を置く意味がないこと(Haladyna, 2004)、選択肢の数を減らして、項目数を増やす方がテストの信頼性が増し、その結果として内容的妥当性が高まること(Rodriguez, 2005)、3択においても当て推量によってのみ全問正解する確率はほぼゼロに等しいこと(Downing, 2006)、などが挙げられる。また大学入試の読解問題を4択と3択で比較し、難易度・信頼性・弁別力・実質選択肢指数の点で、両者にほぼ違いがないことを実証的に検証した研究もある(Shizuka, Takeuchi, Yashima, & Yoshizawa, 2006)。

(3) 問題形式の影響

表1に示した通り、多肢選択式リスニングテストの問題形式には、問題冊子に印刷されている写真や絵を適切に表している英文を選ぶ「描写問題」、会話において、相手の質問に対して適切な応答を選ぶ「応答問題」、対話の内容に関する質問に対して適

切な英文を選ぶ「会話問題」(ダイアログ)、会話の代わりに説明文など1人の話し手による「説明文問題」(モノログ)などがある。

このような問題形式の違いは、得点に大きな影響を与えると指摘されている(Bachman & Palmer, 1996)。実際にTOEICや英検の準2級の応答問題では、4択ではなく3択が採用されていることも考慮すると、問題形式によって適切な選択肢の数が異なる可能性が考えられる。したがって、本研究ではこの点を踏まえ、問題形式別に選択肢数を比較し、それぞれの形式に応じた適切な選択肢数を明らかにする。

2. 研究の目的

本研究の目的は、多肢選択式リスニングテストにおける適切な選択肢数を明らかにすることである。上述のように、国内外の主要なリスニングテストのほとんどは、4択(正解肢1つ・錯乱肢3つ)を採用しているが、言語テスト理論においては、3択(正解肢1つ・錯乱肢2つ)の方が好ましいとされている。本研究では、問題形式別に選択肢数(4択と3択)の異なるリスニングテストを作成・実施して、4択の代替としての3択の可能性を探る。

3. 研究の方法

以下では、本研究で実施した3つの調査の目的とリサーチ・クエスチョン(RQ)を提示し、続いて調査協力者や実験マテリアル、データ分析法等について記す。

(1) 調査1

様々な問題形式の多肢選択式リスニングテストをどのようなプロセスで解答しているのかを、テストテイキング・ストラテジーの観点からプロトコル(有声思考法)を用いて調査する。

RQ 1-1 熟達度によって、ストラテジー使用に違いはあるか。

RQ 1-2 問題形式によって、ストラテジー使用に違いはあるか。

RQ 1-3 正答・不正答によって、ストラテジー使用に違いはあるか。

調査協力者は日本人大学生10名。協力者は英語熟達度(TOEIC IPテストのリスニングセクション)によって、上位と下位それぞれ5名に分けた。

調査マテリアルは、TOEIC公式リスニング問題集(ETS, 2011)のうち、描写6問、応答6問、会話(6パッセージ, 18問)、説明文(6パッセージ, 18問)を用いた。

調査協力者には、1問ずつ解答が終わった

直後に、どのように解いたのかを振り返って声に出して答えてもらった(回顧的有声思考法)。

調査協力者の解答(プロトコル)は, Badger and Yan (2012)と Goh (2002)を参考に作成したストラテジー分類表(認知ストラテジー: 8 カテゴリー, 26 ストラテジー; メタ認知ストラテジー: 6 カテゴリー, 34 ストラテジー)により, 専門家 2 人によってコーディングした。

(2) 調査 2

選択肢数の違い(4 択 vs. 3 択)が, 異なる問題形式において, 難易度・信頼性・弁別力・錯乱肢の機能に与える影響を調査した。

RQ 2-1 選択肢の数は項目困難度に影響を与えるか。またそれは, 問題形式によって違いがあるか。

RQ 2-2 選択肢の数は弁別力に影響を与えるか。またそれは, 問題形式によって違いがあるか。

RQ 2-3 選択肢の数は信頼性に影響を与えるか。またそれは, 問題形式によって違いがあるか。

RQ 2-4 選択肢の数は錯乱肢の機能に影響を与えるか。またそれは, 問題形式によって違いがあるか。

調査協力者は, 日本人大学生 118 名。このうち 62 名を 4 択グループ, 残り 56 名を 3 択グループに分けた。

調査材料は, TOEIC リスニング問題集を使用し, 3 つの形式(描写, 会話, 説明文)を対象とした(ETS, 2011)。描写は 15 問, 会話と説明文は 7 パッセージ, 21 問をそれぞれ使用した。

3 択問題は, 本実験とは異なる 90 名の日本人大学生にオリジナルの 4 択を受験してもらい, 項目分析の結果, 最も選択されなかった錯乱肢を 1 つ削除することで作成した。

本調査では, 2 つのグループの熟達度を調整していないため, それぞれの問題形式で共通項目(描写: 10 問, 会話と説明文: 6 パッセージ 18 問)を設定し, そのテスト結果をもとにラッシュモデルによる等化を行った。

(3) 調査 3

一般に錯乱肢が機能しているかどうかは, 何人の受験者に選ばれているかという頻度によって決定される。しかし, 受験者に選ばれなかったとしても, 錯乱肢としての機能(正解に見える)を果たすものがあるかもしれない。調査 3 では, 応答問題(正答肢 1 つと錯乱肢 2 つの合計 3 つの選択肢から構成)における錯乱肢の魅力度を, 独自に作成したアンケートを使って調査する。

RQ 3-1 正答肢と錯乱肢で受験者の自信度は異なるか。

RQ 3-2 錯乱肢の魅力度は, 受験者の自信度と頻度の点で違いがあるか。

調査には TOEIC リスニング問題集(ETS, 2011)の応答問題 15 問を使用した。

調査協力者は日本人大学生 75 名で, テストの正答率により, 3 つのレベルに分けた(上位群: 14 名, 中位群: 44 名; 下位群: 17 名)。

錯乱肢の魅力度を調査するために, 正解と不正解への自信度を尋ねるアンケート用紙を作成した(図 1)。

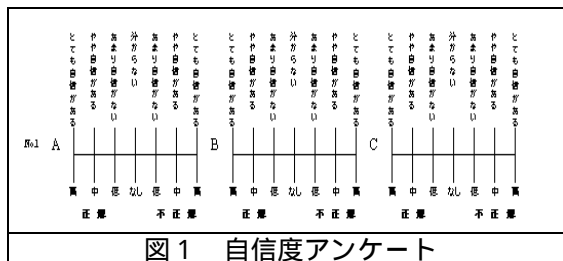


図 1 自信度アンケート

このアンケートでは, 3 つの選択肢全てについて, 正解と思うか不正解と思うか, さらにその判断にどの程度の自信があるかを尋ねる。正解肢については, 正解への自信度になるが, 不正解の選択肢については, 錯乱肢としての魅力度の判定基準となる。つまり, 受験者が自信を持って正解であると判断した錯乱肢は, 魅力度の高い錯乱肢と考えることができる。一方, 受験者が自信を持って不正解であると判断した錯乱肢は, 魅力度の低い錯乱肢ということになる。この調査紙を使用して, 錯乱肢の魅力度を 7 ポイント(7: 自信を持って正解; 1: 自信を持って不正解)で換算し, 錯乱肢の魅力度を計算した。

項目分析を行い, 各選択肢を選んだ人数を割り出し, 前述の錯乱肢の魅力度との関連を検証した。また熟達度によって, 錯乱肢の魅力度に違いがあるかを調査するために, 二元配置の分散分析を行った。さらに点双列相関係数により, 各項目の弁別力を求めた。

4. 研究成果

本章では, 3 つの調査で設定した RQ に対する答えをまとめ, 最後に研究全体を総括する。

(1) 調査 1 のまとめ

RQ 1-1 熟達度によって, ストラテジー使用に違いはあるか。

・上位群は, メタ認知ストラテジーを使用して, 質問に答えるためにテキスト全体のどの部分が重要なかを認識していた。

・下位群は, 本文と選択肢で重複している語を探す認知ストラテジー(fixation)を多用していた。

RQ 1-2 問題形式によって、ストラテジー使用に違いはあるか。

・熟達度に関わらず、最も難易度の高い形式（説明文）では、他の形式よりもストラテジー数が減少した。

・上位群では、難易度の低い形式では、メタ認知ストラテジーを使用するが、難易度が高くなると、認知ストラテジーの fixation に頼る傾向が見られた。

・下位群では、全体的に lexical matching に頼る傾向があるが、描写問題では上位群と同様にメタ認知ストラテジーを使用していた。

RQ 1-3 正答・不正答によって、ストラテジー使用に違いはあるか。

・上位群では、難易度の低い形式では、正解した時の方が使用するストラテジーは多いが、難易度が高くなると、不正解の時の方がストラテジーの数が増える傾向が見られた。

・下位群では、形式に関わらず、正解した時の方がストラテジーの数が多い。また正解・不正解に関わらず、使用するストラテジーの種類（認知ストラテジー）は、ほぼ同じであった。

（ 2 ） 調査 2 のまとめ

RQ 2-1 選択肢の数は項目困難度に影響を与えるか。またそれは、問題形式によって違いがあるか。

・問題形式によって、選択肢数の影響が異なる。すなわち、難易度の低い形式（描写や会話）では、項目困難度に変化はないが、難易度の高い形式（説明文）では、3 択の方が困難度が低くなる。

RQ 2-2 選択肢の数は弁別力に影響を与えるか。またそれは、問題形式によって違いがあるか。

・選択肢数は弁別力に影響を与えず、問題形式による違いも見られなかった。

RQ 2-3 選択肢の数は信頼性に影響を与えるか。またそれは、問題形式によって違いがあるか。

・選択肢数は信頼性に大きな影響を与えず、問題形式による違いも見られなかった。

RQ 2-4 選択肢の数は錯乱肢の機能に影響を与えるか。またそれは、問題形式によって違いがあるか。

・頻度に関しては、問題形式の難易度が高くなるにつれて（描写<会話<説明文）、機能する錯乱肢が増える傾向が見られた。

・弁別力に関しては、高い弁別力を持つ錯乱肢の数が最も多かったのは、3 択でも 4 択でも会話問題であり、続いて説明文、一番少なかったのは描写問題であった。

（ 3 ） 調査 3 のまとめ

RQ 3-1 正答肢と錯乱肢で受験者の自信度は異なるか。

・熟達度の高い受験者は、熟達度の低い受験者よりも自信を持って正答肢を選ぶ傾向が見られた。

・熟達度の低い受験者は、熟達度の高い受験者よりも錯乱肢に魅力を感じる（錯乱肢を正答肢と勘違いする）傾向が見られた。

RQ 3-2 錯乱肢の魅力度は、受験者の自信度と頻度の点で違いがあるか。

・錯乱肢の頻度と魅力度は、必ずしも一致しないことが明らかになった。すなわち、受験者に選ばれなかったとしても、魅力度の高い錯乱肢の存在が確認された。

（ 4 ） 総括

本研究の目的は、リスニングテストにおける適切な選択肢数を、問題形式の観点から明らかにすることであった。

調査 1 は、様々な問題形式におけるテスト受験者の解答プロセスを明らかにすべく行われた。ストラテジー使用という点からプロトコルデータを分析した結果、問題形式の難易度が受験者の用いるストラテジーの数と種類に影響を与えていることが示された。言い換えると、多肢選択式リスニングテストにおいて、問題形式の違いは解答プロセスに影響を与えることが示唆された。

調査 2 では、3 択と 4 択の違いを、項目難易度・弁別力・信頼性・錯乱肢の機能という 4 つの観点から量的な分析を試みた。実験の結果、項目困難度のみが選択肢数の違いの影響を受けることが明らかになった。そしてそれは、問題形式によって影響の度合いが異なることも示唆された。すなわち描写問題では、3 択でも 4 択でも項目困難度には違いがないが、説明文では違いが見られ、3 択の方が困難度が低かった。これは問題形式の難易度が影響しているものと考えられる。すなわちモノログ形式では、質問文に加え選択肢も文字提示されているため、4 択と 3 択では読む分量が異なる。このことが項目困難度の違いに影響をもたらした可能性が考えられる。他の観点に関しては、選択肢数の違いは大きな影響を与えていなかった。これは選択肢の数が異なっても、テスト項目の本質的な部分、すなわち本文や質問文、正答肢が同一であるためだと考えられる。

調査 3 では、受験者の解答に対する自信度という点でアンケートを行い、錯乱肢の魅力度を検証した。実験の結果、最終的に受験者に選ばれなかったとしても、実際には機能していた（受験者を惑わす）錯乱肢の存在が確認された。このことは、従来から行われている項目分析による受験者数ごとの錯乱肢の評価では不十分であることを示唆している。

したがって、今後、多肢選択式リスニングテストにおける適切な選択肢数を考える際には、開発の段階で、従来型の項目分析に加えて、本調査で使用したような錯乱肢の魅力度調査も併せて行う必要性があると言える。

<引用文献>

- Bachman, K. F., & Palmer, A. (1996). *Language Testing in Practice*. OUP.
- Badger, R., & Yan, X. (2012). The use of tactics and strategies by Chinese students in the listening component of IELTS. In L. Taylor & C. J. Weir (Eds.), *IETS Collected Papers 2* (pp. 454-486).
- Downing, S. M. (2006). Selected-response item formats in test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 287-301). Mahwah, NJ: Lawrence Erlbaum.
- ETS (2011). 『TOEIC テスト公式プラクティスリスニング編』国際ビジネスコミュニケーション協会
- Goh, C. (2002). Exploring listening comprehension tactics and their interaction patterns. *System*, 30, 185-206.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- 日本テスト学会 (編) (2007) 『テスト・スタンダード：日本のテストの将来に向けて』金子書房
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Shizuka, T., Takeuchi, O., Yashima, T., & Yoshizawa, K. (2006). A comparison of three- and four-option English tests for university entrance selection purposes in Japan. *Language Testing*, 23, 35-57. doi: 10.1191/0265532206lt319oa

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計2件)

Limura, Hideki. (2014). Attractiveness of distractors in multiple-choice listening tests. 日本言語テスト学会. *JLTA Journal*, 17, 19-39. 査読有.

Limura, Hideki. (2015). The effects of the number of options on different tasks in multiple-choice listening tests. 全国英語教育学会. *Annual Review of English Language Education in Japan (ARELE)*, 26, 237-252. 査

読有.

〔学会発表〕(計3件)

飯村英樹. 「選択肢の数と問題形式がリスニングテストの難易度に与える影響」第40回全国英語教育学会徳島研究大会. 2014年8月9日. 徳島大学(徳島県徳島市).

Limura, Hideki. Relationship between task difference and test-taking strategies in multiple-choice listening tests: A qualitative Study. 日本言語テスト学会第18回全国研究大会. 2014年9月20日. 立命館大学(滋賀県草津市).

飯村英樹. 「多肢選択式リスニングテストにおけるストラテジー使用について：熟達度、問題形式、正答・不正答の観点から」第43回九州英語教育学会大分研究大会. 2014年12月6日. 大分大学(大分県大分市).

6. 研究組織

(1)研究代表者

飯村 英樹 (IIMURA, Hideki)
熊本県立大学・文学部・准教授
研究者番号：30382831