

令和元年6月8日現在

機関番号：22302

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K02790

研究課題名(和文) 多肢選択式リスニングテストにおける錯乱肢の魅力度について

研究課題名(英文) Attractiveness of distractors in multiple-choice listening tests

研究代表者

飯村 英樹 (Iimura, Hideki)

群馬県立女子大学・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) 3,100,000円

研究成果の概要(和文)：本研究は多肢選択式リスニングにおける錯乱肢の魅力度(もっともらしさ)を構成する要因を検証した。日本人大学生を対象としたTOEIC(R)のテストデータを分析した結果、応答問題では「応答キーワード」(質問文の疑問詞に対する答えとなるキーワードを含む)と「語彙的連想」(質問文の語句と関連する語句を含む)、長文問題では「同語」(本文の語句を含む)が、錯乱肢をもっとらしくする要因であることが示唆された。

研究成果の学術的意義や社会的意義

本研究は国内外の多くの主要な英語テストで採用されている多肢選択式テストにおける錯乱肢の魅力度を決定する要因を明らかにする。錯乱肢は正答肢に似ていなければならないが、明らかに誤答でなければならないという厳しい制約があるために作成が非常に難しい。現状では、有効な錯乱肢作成のガイドラインがなく、テスト作成者は複数の錯乱肢を作り出すことに苦慮しており、作成した錯乱肢は実際には機能しないことも多い。本研究の成果は、作成すべき、または避けるべき錯乱肢がどのようなものかを明らかにし、より妥当性の高いテスト作成に寄与する。

研究成果の概要(英文)：This study aimed to investigate the plausibility of distractors in a multiple-choice listening test. Using data from the TOEIC(R) test administered to Japanese university students, this study examined the characteristics of distractors that could attract test takers. The results showed that in question-response tasks “keywords” (a distractor including words or phrases corresponding to the interrogatives of the questions) and “word association” (a distractor including words or phrases associated with the questions) were found to contribute to distractor plausibility. The results also indicated that in dialogue and monologue tasks, “overlap” (a distractor including the same words or phrases used in the text) would be the most influential factor to make distractors attractive.

研究分野：英語教育

キーワード：リスニングテスト 多肢選択式 問題形式 錯乱肢 妥当性 テストテイキングストラテジー 重回帰分析

様式 C - 19, F - 19 - 1, Z - 19, CK - 19 (共通)

1. 研究開始当初の背景

リスニングテストでは、多肢選択式や真偽判定、情報転移、自由記述、ディクテーションなど多様な形式が用いられている。とりわけ多肢選択式は、採点が容易であり高い信頼性が確保しやすいことに加え、記述式に比べて多くの項目数を出題することができ、その結果として内容的妥当性が高められるため (Downing, 2006), TOEIC® や TOEFL®, 実用英語技能検定 (英検) など国内外の主要な英語テストにおいて幅広く用いられている。

多肢選択式テストの特徴は、文字通り複数の選択肢の中から正解を選ぶところにある。そのため、当て推量 (guessing) による正解の可能性を否定できないという指摘もあるが (Hughes, 2003), 正答肢 1 つに錯乱肢と呼ばれる誤答を複数加えることによって、その影響を避ける工夫がなされている。上述の主要な英語テストの多くは 4 択 (正答肢 1 つ + 錯乱肢 3 つ) で構成されている。

しかし、言語テスト理論においては、4 択問題の場合、3 つの錯乱肢すべてが受験者を引きつけることはほとんどなく、1 つあるいは多くても 2 つにとどまることが指摘されている (Rodriguez, 2005)。国内における実証研究においても、錯乱肢すべてが機能するわけではないことが報告されている (Shizuka, Takeuchi, Yashima, & Yoshizawa, 2006)。

錯乱肢には、受験者の能力を弁別するという重要な働きがあるが (Brown, 2005), テスト作成のガイドラインでは、錯乱肢の作成手順が詳細に記述されているとは言えず、「できるだけ「もっともらしく」(plausible) 作成するよう努めること、との表現にとどまっている (Haladyna, Downing, & Rodriguez, 2002)。

つまり、これまでの研究では、錯乱肢を「もっともらしく」する要因、換言すれば、どのような錯乱肢が受験者にとって魅力的なのかは明らかになっていないと言える。

上記の背景を踏まえ、本研究では多肢選択式リスニングテストにおける錯乱肢の魅力度を決定する要因を明らかにする (図 1 参照)。

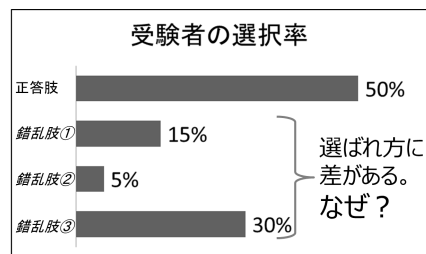


図 1 本研究の概要図

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錯乱肢の魅力度を決定する要因を明らかにすることである。多肢選択式テストは、1 つの正答肢と複数の錯乱肢 (誤答) から構成されているため、錯乱肢は重要な役割を担っている。錯乱肢が適切に機能しているかどうかは、どのレベルのテスト受験者をどれだけ引きつけられたか、によって判断される。本研究は、受験者の英語熟達度に応じた錯乱肢の魅力度を実証的に明らかにすることによって、より妥当性の高いリスニングテスト作成への提案を行う。

3. 研究の方法

本研究では 2 つの調査を実施した。調査 1 では応答問題、調査 2 では長文問題 (100 語 ~ 150 語程度の会話文と説明文) を扱った。

(1) 調査 1 (応答問題)

テストデータ

日本人大学生 199 名が受験した TOEIC®リスニングセクションの Part 2 の 30 問 (Educational Testing Service, 2005) を使用した。応答問題は 1 つの質問文を聞いた後に、3 つの応答文 (3 択) を聞き、その中から最も適切に回答している選択肢を 1 つ選ぶ形式である。質問文も選択肢も問題用紙には記載されておらず、音声のみによる提示であり、放送回数は 1 回であった。

コーディング

寺尾・安永・石井・野口 (2015) を参考に、錯乱肢の魅力度を検討するために 5 つの要因を設定した (表 1)。なお 5 つの要因に該当する錯乱肢の抽出にあたっては、英語教育を専門とする大学教員に協力を仰ぎ、筆者と 2 人で行った。意見が分かれた箇所は話し合っ

表 1. 調査 1 における錯乱肢の魅力度の要因と例

要因	例 (定義)
1. 同語	(質問文中の語句が含まれている) 質問文: When can you submit the <u>report</u> to Mr. Chang? 錯乱肢: Today's weather <u>report</u> .
2. 派生語	(質問文と派生の関係にある語が含まれている) 質問文: When will you be <u>returning</u> to work? 錯乱肢: Please <u>return</u> it by two o'clock.
3. 音声的類似	(質問文中の語に音声的に似ている語が含まれている) 質問文: Do you <u>prefer</u> to work the morning shift, or are you available later? 錯乱肢: I'll be happy to <u>refer</u> you.
4. 語彙的連想	(質問文の語と関連する語が含まれている) 質問文: Where did you get your <u>hair</u> cut? 錯乱肢: It was getting too <u>long</u> .
5. 応答キーワード	(質問文の疑問詞に対する答えとなるキーワードが含まれている) 質問文: <u>Where</u> should I send the revised version of the contract? 錯乱肢: He's probably from <u>the United States</u> .

分析方法

Bachman (2004) にしたがって、テストの総得点をもとに参加者を上位および下位の約 27% それぞれを上位群と下位群、その中間を中位群として、3 つのグループに分けた。この結果にもとづき、項目分析を行い、各錯乱肢を選んだ受験者の割合 (頻度) を算出した。また各錯乱肢の弁別力を検証するために、点双列相関係数を用いた。錯乱肢の有効性は、頻度に関しては 5% 以上の受験者から選ばれているか (Haladyna & Downing, 1993)、弁別力に関しては点双列相関係数の値がマイナスになっているかどうかで判断した (Green, 2013)。以上の分析を通して有効な錯乱肢を抽出した。

次に錯乱肢の魅力度をどの程度予測できるかを検証するために重回帰分析を行った。説明変数は表 1 で示した 5 つの要因とし、従属変数は上述の分析によって得られた有効な錯乱肢の選択率 (頻度) を用いた。重回帰分析の前提として、多重共線性 (許容度と VIF) と外れ値 (クックの距離)、残差の独立性 (ダービン・ワトソン検定) に問題が生じていないか (平井 & 高木, 2017) を確認した。

(2) 調査 2 (長文問題)

テストデータ

日本人大学生 199 名が受験した TOEIC®リスニングセクションの Part 3 (10 パッセージ・30 問) および Part 4 (10 パッセージ・30 問) を使用した (Educational Testing Service, 2005)。長文問題 (Part 3 と Part 4) は、100 語程度の英文 (会話文あるいは説明文) を聞き、内容に関する 4 択の 3 つの問題に答える形式である。質問文と選択肢は問題用紙に記載されており、放送回数は 1 回であった。

コーディング

調査 1 と同様に寺尾・安永・石井・野口 (2015) を参考に錯乱肢の魅力度を検討するために 6 つの要因を設定し (表 2)、筆者ともう一人の大学教員の 2 名で抽出した。

表 2. 調査 2 における錯乱肢の魅力度の要因と例

要因	例 (定義)
1. 同語	(本文中の語句が含まれている) 質問文: <i>Where is Anna's apartment?</i> 錯乱肢: <i>It is next to the <u>big department store</u>.</i> *本文中で "big department store" が使われている。
2. 類義語	(本文中の語句の類義語が含まれている) 質問文: <i>Where is Anna's apartment?</i> 錯乱肢: <i>Ashley's <u>flat</u>.</i> *本文中で "apartment" が使われている。
3. 派生語	(本文中の語句の派生語が含まれている) 質問文: <i>When will Ashley come to Anna's apartment?</i> 錯乱肢: <i>In <u>nearly</u> 30 minutes.</i> *本文中で "near" が使われている。

4. 否定語	(“no” や “not” “never” が含まれている) 質問文: <i>Where is the big apartment?</i> 錯乱肢: <i>It is <u>not</u> near the metro.</i> * 否定を表す “not” が使われている。
5. 音声的類似	(本文中の語句と音声的に似ている語句が含まれている) 質問文: <i>Where is Ashley now?</i> 錯乱肢: <i>She is near the <u>light</u>.</i> * 本文中に “right” が使われている。
6. 限定詞	(“all” “only” “very” など限定詞が含まれている) 質問文: <i>What did Ashely find?</i> 錯乱肢: <i>Anna <u>only</u> found the coffee shop.</i> * 限定詞の “only” が使われている。

分析方法

調査 1 と同じ分析を行った。

4. 研究成果

2 つの調査の結果のまとめを提示し、最後に研究全体を総括する。

(1) 調査 1 のまとめ

上位群の錯乱肢の選択率を予測するのに有意なモデルはなかった。予測への寄与が最も大きい変数は「語彙的連想」であったが、錯乱肢の選択率を予測するのに有意ではなかった ($\beta = .25$, $t = 1.36$, $p = .18$)。中位群においても、上位群と同様に錯乱肢の選択率を予測するのに有意なモデルは見出せなかった。予測への寄与が最も大きい変数は「応答キーワード」であるが、錯乱肢の選択率を予測するのに有意ではなかった ($\beta = .17$, $t = 1.17$, $p = .25$)。上位群・中位群と異なり、下位群においては、「応答キーワード」と「語彙的連想」からなるモデルが予測に有意であり、「語彙的連想」は ($\beta = -.24$, $t = -1.75$, $p = .08$)、「応答キーワード」は ($\beta = .30$, $t = 2.18$, $p = .03$) であり、説明率は 11% であった。すなわち、下位群にとって魅力のある錯乱肢は、語彙的連想を含まず、「応答キーワード」を含むものということになる。「応答キーワード」は、中位群でも魅力度の予測に寄与する要因として最後まで残っていたこともあり、錯乱肢の魅力度に寄与すると考えられる。

応答問題を解く際に、多くの受験者は答えのポイントとなる疑問詞を聞き逃さないよう注意していると予想されるが、質問文全体の理解に失敗した場合は、疑問詞に対応するキーワードだけを頼りに選択肢を選ぶことになる。そのため、「応答キーワード」を含んだ錯乱肢が受験者にとって魅力のあるものになったと思われる。例えば、アメリカの「天気」を尋ねているのに、日本の「天気」を答えるような応答を指す。質問文の理解が不十分であった場合、このタイプの錯乱肢が受験者を引きつけることは十分に考えられる。

しかしもう 1 つの変数「語彙的連想」がなぜ錯乱肢の魅力度のマイナスの要因となるのかについては解釈が困難である。1 つの可能性として、錯乱肢に語彙的に関連する語句が含まれていると、受験者が「引っ掛け問題」と捉えて警戒して選ばない、ということがあるかもしれない。

(2) 調査 2 のまとめ

上位群では、「同語」($\beta = 5.43$, $t = 4.02$, $p < .001$) と「派生語」($\beta = 4.99$, $t = 2.40$, $p = .017$) からなるモデルが錯乱肢の魅力度を説明するのに有意であり、説明率は 13% ($F(1, 136) = 9.89$, $p < .001$) であった。中位群では、「同語」のみが有意であり ($\beta = 4.77$, $t = 3.16$, $p = .002$)、説明率は 7% ($F(1, 136) = 10.00$, $p = .002$) であった。下位群においては、同語 ($\beta = 4.16$, $t = 3.15$, $p = .002$)、限定詞 ($\beta = -8.29$, $t = -2.33$, $p = .021$) および派生語 ($\beta = 4.35$, $t = 2.14$, $p = .034$) によるモデルが有意であり、説明率は 11% ($F(1, 136) = 5.68$, $p = .001$) であった。

全体として、「同語」がどのレベルの受験者においても錯乱肢のもっともらしくする最も影響力のある要因であることが明らかになった。これは聞き手がリスニングテストを受験する際に本文と選択肢が一致するものを正答として選ぶいわゆるマッチングストラテジー (Iimura, 2011) を使うためだからだと考えられる。熟達度が高い受験者でも難易度の高い問題についてはこのストラテジーを使う傾向があることがわかる。加えて、上位群と下位群においては、「派生語」が錯乱肢の魅力度を予測する要因となっていた。これは、熟達度の高い受験者は難易度の高い質問にも答えられるようにできるだけ多くの情報をとらえようとするからであると考えられる。一方、下位群においては、マッチングストラテジーと同様にパッセージ全体の内容理解というよりも選択肢にある語句と似たもの、すなわち派生語があればそれを選ぶ傾向があるためと考えられる。

「限定詞」に関しては、下位群においてのみ錯乱肢の魅力度を予測する要因となっていた。これはおそらく「知識に基づく当て推量」(informed guessing) の 1 つと考えられる。すなわ

ち限定詞を含んでいる選択肢は正当肢ではないと考え、自動的に排除している可能性が考えられる。

(3) 総括

日本人大学生を対象とした TOEIC®リスニングセクションの受験データを用いて、錯乱肢の魅力度を検証した。錯乱肢の魅力度は受験者の英語熟達度によって異なる可能性があることを踏まえ、受験者を上位・中位・下位の3つのレベルに分けた。分析では重回帰分析を用いた。応答問題では錯乱肢の魅力度を説明すると考えられる5つの要因を、長文問題では6つの要因を設定し、説明変数とした。従属変数は各錯乱肢を選んだ受験者数とした。

分析の結果、応答問題においても長文問題においても受験者の熟達度によって、錯乱肢の魅力度は異なることが分かった。しかしながら、応答問題では上位群と中位群では、錯乱肢の魅力度を説明できる有意なモデルを見出すことができなかった。また有意なモデルが発見できた応答問題の下位群および長文問題の全レベルにおいても、本研究で設定した説明変数による説明率は7~13%にとどまった。すなわち錯乱肢の魅力度の90%前後は別な要因による可能性が高いことを意味する。このような低い説明率になった原因としては、錯乱肢のコーディングに問題があった可能性が考えられる。本研究では、英語教育を専門とする大学教員2名で抽出作業を行ったため、テストの作り手の視点から抽出作業を行ったと言える。そのため、例えば「語彙的連想」の場合、受験者は言語的知識の不足により語彙的な関連性が見出せなかった可能性がある。さらには受験者が持っている非言語的知識によって、錯乱肢に誤った語彙的関連性を見出した可能性も考えられる。「音声的類似」についても、作り手が考える類似とは異なる観点で英語の語句に似ている要素を受験者が見出している可能性も否定できない。そのため、今後の研究では受験者にどのように錯乱肢を選んだのか、あるいは選ばなかったのかを詳細に尋ねる有声思考法(発話プロトコル)などによる検証が必要である。

<引用文献>

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge University Press.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment* (New ed.). New York: McGraw-Hill College.
- 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 Associates.
- Educational Testing Service (2005). 『TOEIC 新公式問題集 (Vol. 1)』東京: 国際ビジネスコミュニケーション協会.
- Green, R. (2013). *Statistical analyses for language testers*. London, England: Palgrave Macmillan.
- Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice item? *Educational and Psychological Measurement*, 53, 999-1010.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press. (静哲人(訳)(2003). 『英語のテストはこう作る』東京: 研究社.)
- Iimura, H. (2011). The effects of format difference on strategy use in multiple-choice listening tests: A verbal protocol analysis. *JACET Journal*, 53, 19-34.
- 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.
- 寺尾尚大, 安永和央, 石井秀宗, & 野口裕之 (2015). 「能力別にみた錯乱肢の効果に関する検討: 多肢選択式の英語文章読解テストを用いて」『日本テスト学会誌』11, 1-20.
- 平井明代 & 高木修一 (2017). 「回帰分析: 変数間の因果関係を予測する」平井明代(編)『教育・心理系研究のためのデータ分析入門: 理論と実践から学ぶ SPSS 活用法』(pp. 165-190) 東京: 東京図書

5. 主な発表論文等

[雑誌論文](計3件)

飯村英樹(2018)「受験者を引きつける錯乱肢の特徴: リスニングテストの場合」九州英語教育学会『九州英語教育学会紀要』第46号, 51-59.

Hideki IIMURA. (2018). Distractor Plausibility in a Multiple-Choice Listening Test.

日本語テスト学会. *JLTA Journal*, 21, 65-81.

https://doi.org/10.20622/jltajournal.21.0_65

Hideki IIMURA. (2016). Are the least frequently chosen distractors the least attractive?: The case of a four-option picture-description listening test. 日本語テスト学会. *JLTA Journal*, 19, 46-65.

https://doi.org/10.20622/jltajournal.19.0_46

〔学会発表〕(計3件)

飯村英樹 「受験者を引きつける錯乱肢の特徴：リスニングテストの場合」 第43回全国英語教育学会島根研究大会. 2017年8月19日. 島根大学.

飯村英樹 「リスニングテストにおける錯乱肢の分析：TOEIC テストの Part 3 と Part 4」 第45回九州英語教育学会福岡研究大会. 2016年12月10日. 福岡工業大学.

飯村英樹 「錯乱肢の分析：リスニングテストの場合」 第42回全国英語教育学会埼玉研究大会. 2016年8月21日. 獨協大学.

〔その他〕

ホームページ等

<https://iimurahideki.wixsite.com/iimurahideki>

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。