

第30回 研究助成

A 研究部門・報告V・英語能力テストに関する研究

パラフレージング・レベルによる錯乱肢の作成

— 誤答分析に基づくテキスト理解度の診断に向けて —

研究者:茨城県／筑波大学大学院 在籍 政所 里佳

《研究助言者:村木 英治》

概要

本研究では、パラフレーズ質問の選択肢における言い換えの程度について、原文の語句を含む割合であるパラフレージング・レベルの観点から、現状の調査および実証研究を行った。調査1では現行の読解テストのパラフレーズ質問の選択肢を対象に、記述と英文中の語が重複している程度について比較した結果、パラフレージング・レベルが低い選択肢の割合はテスト間で異なることがわかった。調査2では、パラフレージング・レベルに基づいて選択肢を作成した読解問題を実施した結果、正答・誤答選択肢の組み合わせによる選択率と魅力度への影響は、テスト得点群によって異なることが示された。テスト得点がより低い受験者は、正答選択肢のパラフレージング・レベルが低い場合に、よりレベルが高い錯乱肢を多く選択し、魅力的だと判断する傾向が見られた。結果より、パラフレーズ質問の選択肢作成には、文章中の記述を含む割合や選択肢間の関係性を併せて検討することが重要であると示唆された。

1

はじめに

多肢選択式問題は、今日までの教育的測定において、最も有効かつ長く使用されている形式の一つである。その理由として、採点の容易さと客観性が確保でき、高い信頼性が期待できるなどの長所があることが挙げられる。このような利点か

ら英語教育分野でも、英検をはじめとする多くの大規模テストにおいて多肢選択式問題が出題されてきた。しかし、多肢選択式問題には多くの問題点も指摘されている。中でも主要な問題の一つに、当て推量による正解の可能性が否定できないことが挙げられる。つまり、4つの選択肢から1つを選ぶこの形式には、受験者が問題で求められる理解を達成できているか否かに関係なく、4分の1の確率で正解してしまう恐れがあるということである。

このような当て推量による正解を防ぐ方策として、これまでの研究では、良い誤答の選択肢、すなわち、能力のある受験者には魅力的には見えないが、能力のない受験者にとっては思わず選んでしまう(根岸, 2017)ような錯乱肢をどのように作成すればよいのかという問い合わせが検証されてきた。しかしながら、このような効果的な錯乱肢の性質について検証した研究からは、そのもっともらしさを構成する要素について、十分に一貫した結果が得られているとは言い難い。その理由として、各研究において対象とした問題の質問タイプやテキストなどの要因が異なっていることが挙げられる。また、複数の選択肢から最終的に1つの選択肢に絞る受験者の解答プロセスを考慮すると、正答選択肢と錯乱肢それぞれの性質のみならず、両者の関係性を含めた検証によって、受験者のパフォーマンスをより詳細に診断できると考えられる。しかし、そのような試みを行っている研究はこれまで多くは見られなかった。

そこで、本研究では、多肢選択式問題読解テス

トにおいて大きな割合を占める質問タイプであるパラフレーズ質問に焦点を当て、記述中にテキスト原文にある語を含む割合であるパラフレージング・レベルに基づく錯乱肢を作成し、その有用性について、正答選択肢との関係や選択肢の魅力度の観点から検証する。テキスト理解が十分に達成されている受験者にとっては、テキスト原文の語を含む割合に関係なく、自信をもって正答選択肢を選ぶことができると考えられる。一方、十分に理解しきれていない受験者の場合は、テキスト原文の語を含む割合の高低が、選択肢の判断を迷わせる可能性が考えられる。

本研究は2つの調査から構成される。調査1では、現行の読解テストのパラフレーズ質問を対象に、パラフレージング・レベルの観点から、正答・誤答の選択肢が含む語と英文中の語はどの程度重複しているのかについて調査を行った。続く調査2においては、パラフレージング・レベルに基づいて正答選択肢・錯乱肢を作成したテストを用いて実験を実施した。結果からその有用性を検討し、弁別力のあるパラフレーズ質問の錯乱肢の作成方法に示唆を与えることを試みた。

2 先行研究

2.1 読解テストにおけるパラフレーズ質問

リーディングテストは、出題形式や質問項目、正答・誤答の選択肢の要因によって、測定される能力が変化する。

例えば、清水(2005)は英検、TOEFL、大学入試センター試験問題の問題を分析し、リーディングテストにおいて用いられる質問タイプは大きく

次の6つに分けられることを示した(表1)。

このように、質問タイプによって解答者に求められるテキストの理解は異なっている。つまり、テスト作成者は各質問タイプが読み手のどのような能力を測っているのかを把握した上で、どのタイプを用いるかを決める必要がある。

一般に、読解において達成される「理解の深さ」については3つのレベルがあるとされる(van Dijk & Kintsch, 1983)。テキストを読んだ読み手の心内に記憶される痕跡は「心的表象」と呼ばれ、そのうち(a)表層的記憶(テキスト内の語や文をそのまま記憶している段階),(b)命題的テキストベース(テキストに含まれる命題の理解),(c)状況モデル(読み手の知識や推論を含めて構築された表象)の順に、テキストの理解は達成される。

したがって、例えば、テキストの中に出現する単語や文がそのまま正答選択肢に含まれる質問を出題した場合、表層的記憶とほぼ同じレベルの理解を測っているといえる。一方、選択肢に含まれる語がテキストに含まれる単語や文と全く同じでない場合、すなわち、単にテキスト内の単語や文の異同を問うものではないパラフレーズ質問や指示質問においては、読み手がテキストの中に含まれる命題をどれだけ理解しているかという命題的テキストベースの理解を測っているといえる(卯城, 2012)。

清水(2005)において明らかになったのは、大規模読解テストにおけるパラフレーズ質問の多さである。特に、英検の長文読解問題においては、およそ9割の設問がこの質問タイプに該当することが示されている。このようにパラフレーズ質問が多くの読解テストにおいて採用されている理由の一つとして、質問作成の容易さ、解答の作成のしやすさから、作成者にとって作りやすいこ

■表1: 読解テストにおける質問タイプ(清水, 2005)

- パラフレーズ質問(文章中の局所的なある一部分を言い換えると、質問とそれに対する正解が得られる質問。明示情報に関する質問)
- 推論質問(文章に基づいて適切に推論されることについて問う質問)
- テーマ質問(パラグラフまたは文章全体の主題について問う質問)
- 提示質問(代名詞または指示表現の先行詞の理解について問う質問)
- 語彙問題(語彙の意味について問う質問)
- 文章構造質問(比較・対照や時間順など文章構造について問う質問やある内容が文章中のどの部分で述べられていたかを問う質問)

とが挙げられる。また、清水は、質問文や選択肢を言い換えることで、文章中のある部分とおおよそ一致するパラフレーズ質問は、受験者自身でパラフレーズする必要はあるが文章中に明示されている情報について解答することを求めるため、下位レベル処理の質問に分類されたとした。

このように、項目作成が容易で困難度がそれほど高くないことから、質問タイプの中でも出題や解答の機会が最も多いと考えられるパラフレーズ質問であるが、未だに明らかにされていない点がいくつもある。

第一に、定義としては同じ「パラフレーズ質問」という分類に該当する項目であっても、選択肢の言い換えの程度は統制されておらず、項目間で異なることが多い。例えば、清水(2005)の行っていた設問分類では、パラフレーズ質問の中に、テキスト内の単語や文をそのまま字義通りに解答すれば正答となる字義的な質問も含まれていた。つまり、パラフレーズ質問における選択肢が、テキスト内の単語や文を言い換えている程度、含んでいる割合について厳密な場合分けを行った試みは未だに十分であるとは言い難い。前述したように、質問項目の選択肢がテキスト内の単語や文をどの程度含んでいるのかによって、読み手のテキストの理解のうち、どの段階を測っているのかが変わること可能性がある。構成概念の検証のためには、選択肢がテキスト内の単語や文を含む程度、言い換えている程度を考慮することが求められると考えられる。

第二に、項目の選択肢どうしの関係性である。前述したように、パラフレーズ質問ではテキストの一部分を言い換えることにより、質問と正答選択肢が作成される。しかし、正答選択肢以外の錯乱肢における、テキストに含まれる単語や文等の表現を含んでいる程度については検討する必要はないのだろうか。多肢選択式テストに解答する受験者の立場に立ってみると、複数の選択肢から1つの正答を選ぶ際、最初から1つの正答に絞って選ぶ場合を除けば、他の選択肢の内容も考慮して吟味すると考えられる。すなわち、正答選択肢と錯乱肢との関係性(e.g., 正答選択肢・錯乱肢においてテキスト内の語句を含んでいる程度)を考慮することは重要であると考えられる。

2.2 多肢選択式問題における錯乱肢の機能と性質

多肢選択式テストにおいて、通常、受験者は複数与えられた選択肢の中から1つの正答を選ぶ。この受験者に与える正答の選択肢や錯乱肢の作り方によって、問題の弁別力が変化する可能性があることから、選択肢の作成方法は問題作成者にとって重要な課題である。その中でも、錯乱肢は、理解が十分な受験者とそうでない受験者を弁別し、項目の有効性を高める上で重要な役割を担っている。

それでは、良い選択肢とはどのようなものなのか。Haladyna and Downing(1989)とHaladyna, Downing, and Rodriguez(2002)は、選択肢全般に関する多肢選択式問題作成のガイドラインを提示した。選択肢どうしの長さを均一に揃えることや、選択肢は類似したものとすることなど複数の選択肢を設ける上での留意事項に加え、錯乱肢については「もっともらしい錯乱肢であること」を提案している。また、多肢選択式テストの難易度を左右する要因を調査したDrum, Calfee, and Cook(1981)において、「錯乱肢の「もっともらしさ」が難易度に有意に影響したことが示されている。加えて、Drum et al. (1981)は正答選択肢についても、低頻度語や新出内容語を含むか否かによって難易度が変化したことを見出している。

このように、もっともらしい錯乱肢の重要性については明らかにされていたが、「もっともらしさ」に影響するのはどのような要因であるかについて、これまでその構成要素を調査した研究は多くなかった。

その中の一つに、Ushiro et al.(2007)がある。彼らは、読解テストにおける錯乱肢の「もっともらしさ」の指標を調査するために、Drum et al. (1981)の指摘したテストの難易度を決定する要因のうち、(a)錯乱肢に含まれる語数、(b)錯乱肢に含まれる内容語の割合、(c)錯乱肢に含まれるテキスト内で用いられていない内容語の割合、(d)錯乱肢に含まれる否定語の総数、(e)錯乱肢に含まれる前置された語の総数、(f)錯乱肢に含まれる照応語の総数、の6つの要因について検証した。日本人大学生は、異なる質問タイプで構成されるTOEFLの多肢選択式問題に解答した。結果とし

て、質問タイプにより、これらの指標がもっともらしさの指標として機能するかどうかは異なっていた。特に、パラフレーズ質問においては、錯乱肢とテキストに含まれる語の重複が、もっともらしさに影響していたことが示されている。全体としては錯乱肢の表面的な性質は錯乱肢のもっともらしさとは関わりが弱いことが示されたものの、質問タイプによって選択肢のもっともらしさに影響する要因が異なる可能性が示唆されている。

もう一つの研究、Ushiro et al. (2008) では、大学生を対象とした実験で読解テストの質問タイプが自由記述式と多肢選択式の難易度に与える影響について検証し、多肢選択式では自由記述式とは異なり、項目難易度が質問タイプによって変化しないこと、そして選択肢の重要性が大きいことを示している。特に、難易度が高い項目ではどのような選択肢があったのか調査を行ったところ、いくつかの項目においてテキスト原文にある語と正答選択肢の語との重複の割合が非常に低い場合に正答率が下がっていたことが明らかになった。一方、錯乱肢の語とテキスト中の語の重複の割合については、項目によって一貫した結果は得られなかった。このことから、多肢選択式テストへの解答時には、受験者が選択肢から得られる情報に大きく頼っており、特にテキストと選択肢の語が重複しているものを選ぶ傾向があることが示唆されている。

また、高木(2011)は、「記憶への保持されやすさ」という指標が錯乱肢の有用性を決定づけるかどうかについて調査を行った。具体的には、テキストの中における他の情報との因果的連続性の強弱を操作することで、記憶に保持されやすい錯乱肢と保持されにくい錯乱肢を作成し、日本人大学生を対象に、その効果を比較検証している。結果として、因果的連続性が強い錯乱肢は、テキストにない情報を記述した錯乱肢よりも、テキスト理解度の低い受験者を多く引き付けることを示した。

上記の研究結果を踏まえると、多肢選択式読解テストにおける選択肢のもっともらしさは、質問タイプや受験者のテキストの理解度により異なることが予想される。特に、前述した通り読解テストで最も多く用いられているパラフレーズ質

問において、選択肢、とりわけ錯乱肢がテキストの情報を含む程度が受験者の解答パフォーマンスにどのような影響を与えるのかについては、検証の余地があると考えられる。したがって、本研究ではこの点について、次節で述べる手法を用いて調査を行うこととした。

2.3 パラフレージング・レベルを用いた多肢選択式問題

本節では、本研究が取り上げる2つの手法について述べる。まず、本研究で主要な選択肢の分類・作成基準として扱うのは、Keck(2006)のParaphrase Typesに基づいて考案した「パラフレージング・レベル」である。

Keck(2006)では、アメリカの大学生を対象に、L1とL2で書かれた1000語程度の文章をそれぞれ1パラグラフ程度におさめるように求める要約タスクを実施し、書き手がテキスト内で使用されている言葉を借用するストラテジーである「パラフレーズ」について調査を行っている。Keckは、書き手の行ったパラフレーズについて、元のテキストに含まれている情報がどの程度要約において複製されているか・言い換えられているかにより、次の4つの種類に分けた:(a) Near Copy, (b) Minimal Revision, (c) Moderate Revision, (d) Substantial Revision。さらに、Keckは、パラフレーズの割合を判断する基準として、「unique links (UL), general links (GL)」という枠組みを用いている。unique linksは、本文において局所的に表れる内容語(e.g., 名詞、形容詞、動詞), general linksは本文において広範囲に複数回表れる内容語であると定義しており、このULの語数の合計数が総語数に占める割合に準じて各パラフレーズが分類される(表2)。具体的には、合計数が50%以上であればNear Copy (NC), 20~49%以上であればMinimal Revision (MIR), 1~19%であればModerate Revision (MOR), 全く含まれない(0%)場合にはSubstantial Revision (SR)に分類されている。

本研究では、このKeck(2006)のParaphrase Typesの算出方法に関し、ULとGLの合計が選択肢の総語数に占める割合を求めるように一部修正したパラフレージング・レベルを基準として、多肢選択式読解テストにおけるパラフレーズ質

問の選択肢の分類を行うこととする。その理由は主に2点挙げられる。1点目に、本研究では表層的な語の性質に絞ってテキスト原文と選択肢の語の重複の程度の変化による影響を調査するため、重複する語の数が総語数に占める割合という、純粹かつ客観的な値を算出するKeckの枠組み

が適していると考えるからである。2点目は、カウントする語の対象を内容語に絞っていることから、本研究でマテリアルとする読解テキストの機能語による影響を抑えられると考えたためである。

■表2: Paraphrase Typesの分類基準 (Keck, 2006)

	ULの語数の合計が総語数に占める割合
Near Copy (NC)	50%以上
Minimal Revision (MIR)	20 ~ 49%
Moderate Revision (MOR)	1 ~ 19%
Substantial Revision (SR)	0%

もう一つの手法として、本研究では、上記のパラフレージング・レベルを用いて作成された選択肢それぞれについて、受験者に対する魅力度を算出する。

多肢選択式の抱える課題は主に2つあるとされる。1つ目は、当て推量による正答の可能性があることである(Kobayashi, 2002)。つまり、内容を理解し、選択肢を吟味した上で正答を選んだ場合も、ランダムに選び4分の1の確率で当たった場合も、同じ「正解」となるため、作成者側からは適切な解答プロセスを経た解答とそうでない解答を区別することができない。2つ目は、解答者が正答を特定することはできなくとも、いくつかの選択肢は不正解であることが分かっている場合に、その部分的知識を結果に反映させることができない点である。

このような多肢選択式テスト受験者の解答プロセスや部分的知識を反映させるために、様々な修正方法が考案されてきた。例えば、Shizuka (1999a, 1999b)では、受験者に解答自信度を3段階で評価させた値を、解答の正解・不正解と合わせて0~5の6段階で組み込む方法を採用すると、テストの信頼性や妥当性が向上したことを示している。この他、自信度に加え、解答までの反応時間や正答率を得点に反映させる方法もテストの信頼性向上を促したことが示されている(e.g., Shizuka, 2003)。

このような解答自信度の枠組みを応用し、

Iimura(2014)では、正答選択肢と錯乱肢の、受験者に対するもっともらしさである「魅力度」を数値化した研究を行っている。Iimuraは熟達度の異なる日本人大学生を対象に、多肢選択式リスニングテストにおける選択肢への解答自信度を調査した結果から、受験者に選ばれる割合の低かった錯乱肢であっても、受験者は魅力的に感じている可能性があることを明らかにした。その上で、選択率などの解答頻度とは別の指標を用いて、各錯乱肢の魅力度を詳細に調べることが重要であるとしている。また、熟達度による違いとして、熟達度の低い学習者は錯乱肢に対してより高い魅力度を感じ、熟達度の高い学習者は正答に対する自信度がより高い傾向があることを明らかにしている。

以上のように、選択肢が受験者に与える影響を検証する上で、選択率といった解答頻度とは別に、受験者が各解答にどの程度自信をもったか、各選択肢をどの程度魅力的だと感じたかを調べるために、解答プロセスを表出化させることは重要であると考えられる。したがって、本研究では、パラフレージング・レベルを用いて作成された正答選択肢と錯乱肢それぞれの選択率に加え、魅力度を算出することにより、各選択肢が受験者にどのように作用するのかを検証する。

3

本研究の目的

本研究では、読解テストで最も多く用いられているパラフレーズ質問において、選択肢がテキストで使用される語を含んでいる程度が、テキスト理解度の異なる受験者の解答パフォーマンスに与える影響を調べることを目的とする。特に、記述の中にテキスト内の語を含んでいる割合であるパラフレージング・レベルを操作して作成された選択肢の有用性を検証するため、本研究では2つの調査を行った。まず調査1では、3つの現行の読解テストを対象に、各項目の選択肢のパラフレージング・レベルにはどのような傾向があるのかを調査した。続く調査2では、調査1から明らかになった傾向に基づいて作成された、異なるパラフレージング・レベルの正答選択肢・錯乱肢が受験者に選ばれた割合や魅力度を分析し、その有効性を検証した。

4

調査

4.1 目的

調査1の目的は、現在運用されている読解テストのパラフレーズ質問を対象とし、各項目の選択肢のパラフレージング・レベルの傾向を明らかにすることである。また、調査結果をテスト間で比較することにより、テストごとの選択肢の特徴およびその理由についても考察する。

4.2 マテリアル

現在運用されている読解テストのうち、英検、TOEIC、TOEFLの読解セクションにおける長文読解問題を使用した。以下に示す通り、本調査では3種類の読解テストの合計124項目の選択肢について分析を行った。

これらのテストを選んだ理由は、次のような共通した特徴を持つためである：(a) 4つの選択肢から成る多肢選択式である、(b) 受験者の人数が多く、大規模に実施されている、(c) 正答選択肢と対応するテキスト記述が明確である。

本調査ではパラフレーズ質問の選択肢の傾向について、テスト間で比較・検討することが目的であるため、調査対象とする問題は、テキストの明示的な内容を問うパラフレーズ質問（清水、2005）に絞った。そのため、(c)を満たしていない、すなわち正答選択肢の情報についてテキスト中に対応する記述が存在していない問題に関しては、以下の分析対象とした項目の中からも除外している。加えて、受験者に対し、テキストで述べられている内容と合致していない選択肢を選ばせる問題や、テキスト全体の主旨を問う問題などは、本研究の焦点とするパラフレーズ質問とは異なるため、分析から除外した。

a. 英検2級（項目数41）（日本英語検定協会、2017；旺文社、2017）

2015年度第1回4B, 4C, 2015年度第2回4B, 4C, 2015年度第3回4B, 4C,
2016年度第1回3B, 3C, 2016年度第2回3B, 3C, 2016年度第3回3B, 3C

b. TOEIC（項目数43）（国際ビジネスコミュニケーション協会TOEIC運営委員会、2016, 2017）

TOEIC Listening & Reading 問題集1, 2よりPart 7 読解問題

c. TOEFL（項目数40）（Educational Testing Service, 2013）

TOEFL iBT リーディング演習問題1～3, 実戦問題1, 2

4.3 手順

2.2節で示したKeck(2006)の分類に基づき、表3のような選択肢の分類基準を作成した。選択肢のパラフレージング・レベルは、「選択肢の記述と英文両方に含まれる語の数／選択肢の記述の総語数」によって算出される割合であり、客観的な数値であったため、全ての項目について調査者1名が評定を行った。

各テストのパラフレーズ質問となる項目の正答・誤答選択肢に、本文中の対応する記述が含ま

れる割合に応じてパラフレージング・レベルを評定した。具体的には、Keck(2006)に基づき、以下のlinkとなる語が英文中に含まれている数をカウントし、それらの総語数に対する割合を算出した(例1)。選択肢と本文中の語との対応について、調査対象とした語は主に内容語であり、原形の変化形も含め、テキスト内の合計数を算出した。関係詞、前置詞、数字以外の冠詞については、英文全体で選択肢の記述内容と関連しない部分でも多く見られることから、本調査の対象からは除外した。

- unique links(UL):本文において、局所的(1～2度)に表れる内容語(e.g., 名詞、形容詞、動詞)。
- general links(GL):本文において、広範囲(3度以上)に多く表れる内容語

■表3: 選択肢のパラフレージング・レベル分類基準

	UL, GLの語数の合計が総語数に占める割合
Near Copy (NC)	50%以上
Minimal Revision (MIR)	20～49%
Moderate Revision (MOR)	1～19%
Substantial Revision (SR)	0%

■例1: 選択肢の評定方法 [英検2級2016年度第3回3C (34)]

- (34) What are chocolate manufacturers concerned about?
1. People prefer **living** a healthy life to eating sweets such as chocolate. (12語)

$$\begin{aligned} \text{UL (太字)} : & \text{people}(1\text{語}) + \text{living}(1\text{語}) = 2\text{語} \rightarrow 2 / 12 = 16.7\% \\ \text{GL (下線)} : & \text{chocolate}(1\text{語}) \times 1 = 1\text{語} \rightarrow 1 / 12 = 8.3\% \\ \text{合計: UL(16.67\%)} & + \text{GL(8.33\%)} = 25.0\% \\ & \Rightarrow \text{Minimal Revision} \text{と評定。} \end{aligned}$$

4.4 結果と考察

各テストのパラフレーズ質問となる項目の正答選択肢および錯乱肢に、本文中の対応する記述が含まれる割合に応じ、パラフレージング・レベルを評定した結果について、それぞれ表4、5に示す。

分類の結果、対象とした読解テストの正答・誤答選択肢の全体的な傾向としてはMIRが最も多く、約半数を占めており、MIRに次いで多かつ

たのはNCであることが明らかになった。つまり、現行の読解テストにおいては、正答・誤答を問わず、ある程度英文中で使用される語を含む選択肢が多く用いられることがわかった。一方で、MOR、SRなどパラフレージング・レベルが低い選択肢の割合に関してはテスト間の違いが見られた。以下より、各テストの選択肢の特徴について、理由とともに考察し、テスト間の比較を行う。

■表4: パラフレーズ質問の正答選択肢の分類結果

	選択肢数	NC	MIR	MOR	SR
英検	41	13 (31.7%)	26 (63.4%)	2 (4.9%)	0 (0%)
TOEIC	43	10 (23.3%)	17 (39.5%)	8 (18.6%)	8 (18.6%)
TOEFL	40	13 (32.5%)	27 (67.5%)	0 (0%)	0 (0%)
Total	124	28 (28.6%)	52 (53.1%)	10 (10.2%)	8 (8.2%)

■表5: パラフレーズ質問の錯乱肢の分類結果

	選択肢数	NC	MIR	MOR	SR
英検	123	29 (23.6%)	86 (69.9%)	8 (6.5%)	0 (0%)
TOEIC	129	27 (20.9%)	34 (26.4%)	14 (10.9%)	54 (41.9%)
TOEFL	120	56 (46.7%)	58 (48.3%)	2 (1.7%)	4 (3.3%)
Total	372	112 (30.1%)	178 (47.8%)	24 (6.5%)	58 (15.6%)

4.4.1 英検

英検では全体的に、節および文単位で提示される選択肢が多くかった。正答選択肢・錯乱肢とともに、MIRが最も多く、次いでNC、MORの順に多いと

いう特徴が共通していた。そして、今回対象とした項目に限って言えば、3つのテストの中で唯一、SRの正答選択肢・錯乱肢は見られなかった。表6は英検の問題からの抜粋である。

■表6: 英検で出題された項目の例(英検2級2016年度第1回3B第1パラグラフより)

Many large snakes eat as rarely as once every three months. When they do eat, however, they eat whole animals, such as deer or wild pigs. In order to digest these large meals, the heart and other organs inside the snake's body suddenly grow much larger. The heart, for example, does this so that it can pump blood around the body much faster. This allows the snake to digest food more quickly. After the snake has digested the meal, the organs go back to their usual size.

(30) Many large snakes change the size of their hearts to in order to

- 1 make more space for the whole animals that they eat. (MIR)
- 2 better digest the blood of the animals that they eat. (MIR)
- 3 be able to catch more food more quickly. (NC)
- ④ help them digest large meals more quickly. (NC)

この設問で正答を得るためには、下線部の3文を理解することが求められる。特に、第3文を理解できれば、記述の内容およびdigest, large, mealsなどの使用されている語彙が本文と一致している正答選択肢4を選ぶことができる。

残りの錯乱肢について詳しく見てみると、いずれの錯乱肢についても、eat, animalsなど、テキスト中に複数回現れる単語や、food, more quickly, など、特定の部分において局所的に出てくる単語が含まれている。このように、本文の記述と正答選択肢・錯乱肢両方の間で語のオーバーラップの程度が高いことが分かる。

英検の項目の多くを占めていたのは、このようにNC, MIRのバラフレージング・レベルに分類される正答選択肢・錯乱肢であり、MOR, SRのレベルに該当する選択肢の数は少なかった。このような結果となった理由は、(a) 選択肢の提示単位、(b) テキストのレベル、に関係していると考えられる。(a)に関しては、この後に述べるTOEICに比べ、英検の選択肢は節もしくは文單

位で提示される場合が多く、記述の語数が多くたために、本文中との語のオーバーラップの割合も増加したという可能性が考えられる。(b)については、本調査では、他のテストとの比較および調査2で参加する協力者の読解熟達度を考慮したため、英検の中でも2級の問題のみをマテリアルとして扱った。例えば、準1級や1級で使用されるテキスト・選択肢を用いて分析した場合、テキストはより長くなるが設問数は減少し、1つの設問で問われる情報量は級が上がれば上がるほど多くなると考えられるため、結果として選択肢では本文で記述された複数の情報を統合した表現が用いられ、MOR, SRの選択肢が増加する可能性は十分考えられる。

4.4.2 TOEIC・TOEFL

以下では、TOEICとTOEFLの概観について併せて見ていく。

まず、TOEICでは全体として、語や句などの単位で提示される選択肢が多かった。今回扱った

■表7: TOEICで出題された項目の例(TOEIC Listening & Reading 問題集1 Part1より)

To: All Employees
From: Su Kyung Cho <skcho@greatcars.co.nz>
Subject: Annual outing
Date: 1 December

I wanted to let everyone know that management has decided to postpone our annual company hiking trip, which is usually held in December. As all of you know, that is the busiest time of year for our car rental agency, and every year it is difficult to fit the trip into our schedule. The new date is 23 January. This should be more convenient for anyone who wants to go.

(161) What is the purpose of the e-mail?

- a To schedule a meeting (MIR)
- b To make hotel reservations (SR)
- c To explain a car rental policy (MIR)
- d To announce a change in plans (MOR)

問題では、正答選択肢はMIR, NC, MORおよびSRの順で多いのに対し、錯乱肢はSR, MIR, NC, MORの順で多かった。また、他のテストに比べ、SRの選択肢が多いことが特徴的であった。表7はTOEICの問題からの抜粋である。

この問題で問われているのは書き手がメールを送った目的であり、第一文および下線部の情報に含まれる、“to let everyone know”を“to announce”へとパラフレーズすることにより、正答選択肢dを解答することができる。一方で、錯乱肢に関しては、aとcはテキスト内にそれぞれschedule, car rentalといったテキスト内で局所的に出現する語が含まれているのに対し、bに関してはいずれの語も本文中で使用されていない。また、いずれの錯乱肢においてもその内容はテキストの記述と合致していない。

このように、TOEICでは、正答選択肢ではテキストの対応する情報についてテキストで使用されている語よりも短い表現で言い換える場合が、錯乱肢においてはテキスト内で局所的に使用さ

れている語を含んでいるか、全く対応する記述が存在しないような場合が多く見られた。

この理由には、他のテストと比較して、TOEICでは選択肢の提示単位が語・句と比較的短く、また、設問ごとのテキストの長さが短いことが挙げられる。テキストの総語数が少なく、選択肢の提示単位のために本文よりも短い表現で言い換える必要性から、テキスト中と重複する語が限定される可能性が考えられる。また、そうして作成される錯乱肢の数には限りがあるため、テキスト内で対応する記述が全く存在しない錯乱肢も含めて提示している可能性もある。

一方、TOEFLにおいては、句単位あるいは節・文単位で提示される選択肢が多く見られており、正答選択肢はMIR, NCのみに集中し、錯乱肢はMIR, NC, SR, MORの順で多かった。表8はTOEFLの問題からの抜粋である。

この設問の正答選択肢cは、テキスト本文中の下線部の情報を理解することで解答でき、受験者には、本文中の“had almost no contact with

■表8: TOEFLで出題された項目の例(TOEFL iBT リーディング演習問題2, The Expression of Emotionsより)

Moreover, people in diverse cultures recognize the emotions manifested by the facial expressions. In classic research Paul Ekman took photographs of people exhibiting the emotions of anger, disgust, fear, happiness, and sadness. He then asked people around the world to indicate what emotions were being depicted in them. Those queried ranged from European college students to members of the Fore, a tribe that dwells in the New Guinea highlands. All groups, including the Fore, who had almost no contact with Western culture, agreed on the portrayed emotions. The Fore also displayed familiar facial expressions when asked how they would respond if they were asked in stories that called for basic emotional responses. Ekman and his colleagues more recently obtained similar results in a study of ten cultures in which participants were permitted to report that multiple emotions were shown by facial expressions. The participants generally agreed on which two emotions were being shown and which emotion was more intense.

(5) According to paragraph 2, which of the following was true of the Fore people of New Guinea?

- a They did not want to be shown photographs. (MIR)
- b They were famous for their story-telling skills. (SR)
- c They knew very little about Western culture. (MIR)
- d They did not encourage the expression of emotions. (MIR)

“Western culture”という記述が選択肢において“knew very little about Western culture”へと言い換えられていることに気づくことが求められている。一方、錯乱肢に関しては、a, dについては局所的(e.g., shown, photographs)および全体的に(e.g., expression, emotions), 本文中でも出現する語が使われているものの、内容が対応する記述は存在しない。また、錯乱肢bに含まれる語および情報も、本文には記述されていない。

上記のようにTOEFLにおいては、正答にはテキストの対応する語を比較的高い割合で含む選択肢が多く、錯乱肢においてはテキスト内で局所的又は全体的に使用されている語を含むか、全く対応する記述が存在しないような場合が見られた。この理由として、TOEFLでは他のテストに比べ、1設問当たりに関連するテキストの分量が多いという特徴が考えられる。このように総語数、およびパラグラフごとの分量も多いテキストでは、設問ごとに問われる範囲が広く、選択肢作成における自由度が比較的高くなり、NC・MIRの選択肢を量的に多く作ることや、多様なレベルでの錯乱肢の作成がしやすかったと考えられる。

4.4.3 調査1のまとめ

3種類の多肢選択式テストのパラフレーズ質問における選択肢を比較した結果、正答・誤答選択肢のパラフレージング・レベルの全体的な傾向としてはMIRが最も多く、約半数を占めており、次いで多かったのはNCであることが明らかになった。つまり、現行の読解テストにおいてパラフレーズ質問で求められている能力は、正答・誤答を問わず、ある程度英文中で使用される語を含む選択肢を選ぶことのできる能力であると考えられる。

一方で、MOR, SRなどパラフレージング・レベルが低い選択肢の割合に関してはテスト間の違いがあることがわかった。今回調査した読解問題において、英検ではSRの選択肢は正答・誤答いずれも見られず、TOEFLではMOR, SRの正答選択肢は見られなかったが、TOEICでは他の2つのテストに比べてSR, MORの正答・誤答選択肢が多くなっていた。

上記の分類結果を踏まえ、調査2では、既存の読解テストにおけるパラフレーズ質問の正答選択肢および錯乱肢のパラフレージング・レベルを

改編した読解テストを作成する。テキスト中の語とのオーバーラップの割合が異なる選択肢によって構成される読解問題を、日本人英語学習者である大学生を対象に実施することにより、各選択肢の受験者を引き分けている程度がどのように異なるのかを検証する。

5 調査2

5.1 目的

調査1より、現行の読解テストにおけるパラフレーズ質問の選択肢のパラフレージング・レベルの傾向はテスト間で異なることがわかった。そこで調査2では、調査1の分類手順に基づいて既存の読解テストの選択肢を改編し、4つの異なるパラフレージング・レベルの選択肢からなる読解問題を作成・提示し、受験者が各選択肢を選んだ割合を明らかにする。加えて、各選択肢が受験者にとってどの程度もっともらしいのかを示すもう一つの指標として、魅力度を明らかにする。また、本調査では、テストの総得点に応じて受験者を上位群、中位群、下位群の3つに分類し、受験者のパフォーマンスによって、各選択肢の選択率や魅力度がどのように異なるのかについても検討する。

5.2 協力者

国立大学に通う43名の日本大学生・大学院生が調査に参加した。ただし、協力者のうち、制限時間内に全ての項目に解答できなかった1名のデータは分析から除外した。協力者の専攻は、医学、工学、生物学、人文学、社会学、教育学など多岐にわたった。

5.3 マテリアル

5.3.1 使用したテスト

調査2では、英検2級過去問題の長文読解セクションより選定した以下の4つの読解問題を、基盤のテストとして用いることとした(表9)。その

上で、基盤としたテストの各項目の正答選択肢および錯乱肢について、4つのパラフレージング・レベルが同じ数ずつ設けられるように改変したテストをマテリアルとして使用した。

英検の読解問題を基盤のテストとした理由は、(a)選択肢の記述の形態が節・文で統一されていること、(b)本文中で使用される語を含んでいる割合が低い選択肢の出現頻度が少ないと、の2点が挙げられる。

(a)に関しては、調査1で明らかになったように、英検の選択肢は節あるいは文の形で記述されている。本調査においては、テキスト原文との語のオーバーラップの割合を調整することで選択肢のパラフレージング・レベルを設定するため、記

述のおおよその語数や提示単位など、各項目の選択肢間の等質性が確保されている点で、英検の読解問題が最適であると考える。

また、(b)については、調査1で明らかになったこととして、現行の読解テストの中でも英検は、本文中で使用される語を含んでいる割合が低い選択肢を用いた項目が少ないことが挙げられる。特に、SRに関しては正答選択肢・錯乱肢のいずれにおいても例が見られなかった。ゆえに、正答選択肢と錯乱肢のパラフレージング・レベルの組み合わせによる受験者への影響を調べる上で、テキスト原文との重複の程度が低い選択肢を用いた検証の余地がより大きいと考えた。

■表9: 調査2で使用した英検過去問より選んだ英文 (旺文社, 2003)

テキスト	総語数	FKGL
2001年度第2回4A Green Cleaners	338	8.4
2001年度第2回4B Drugstore Dilemma	320	10.3
2002年度第1回4B The Placebo Effect	338	11.3
2001年度第3回4B Million-Dollar Math	342	10.7

(注) FKGL = Flesch-Kincaid Grade Level.

5.3.2 錯乱肢の作成

初めに、基盤とした英検の読解問題について、4.3節例1に示す通りの手順で、選択肢中にテキスト原文の語が含まれる割合を算出した。続いて、4択問題における正答・誤答の選択肢のパラフレージング・レベルについて、正答選択肢のレベルが決まると他の錯乱肢は残りの3つのレベルになるような問題パターンを4つ設定した(表10)。

1つの問題パターンごとに4つの選択肢にパラフレージング・レベルを振り分けられるよう、算出して得られたテキスト原文の語の割合と、錯乱

肢のパターン内におけるレベルが異なる場合には調整を行った。具体的には、当該選択肢が設定したいパラフレージング・レベルよりも高い場合には、テキスト原文の記述をパラフレーズした表現(e.g., 上位語、類義語)を使用し、逆に低い場合には、テキスト原文で用いられている表現を使用することにより、修正を行った。調査者が改編した選択肢は、ネイティブスピーカーによって表現や意味に誤りや不自然さがないかを確認され、必要な場合には修正が加えられた。

■表10: 調査2で作成したマテリアルの問題パターン

	P 1	P 2	P 3	P 4
正答選択肢	NC	MIR	MOR	SR
錯乱肢①	MIR	MOR	SR	NC
錯乱肢②	MOR	SR	NC	MIR
錯乱肢③	SR	NC	MIR	MOR

(注) P = パターン (pattern)。

5.3.3 自信度アンケート

協力者の正答選択肢および錯乱肢に対する捉え方を調べるために、読解テストの解答用紙の解答記入欄の隣に、各選択肢への解答の自信度を回答する質問項目を設けた（資料2:110頁参照）。協力者は各選択肢の正誤の解答について、低・中・高の3段階のいずれかで自信度を答えるよう指示され、該当する自信度に丸をつけて回答した。

5.3.4 筆記再生課題

筆記再生課題を使用する目的は、本調査で使用した読解テストの項目は全てパラフレーズ質問であり、表層的記憶および命題的テキストベースの理解を測定していると考えられるが、同様のテキスト理解を測定していると考えられる筆記再生課題について、その成績と読解テストの得点との相関を調べることにより読解テストの妥当性を確認するためであった。筆記再生課題は紙面にて作成・実施し、協力者は、各英文に関して覚えている内容をできるだけ多く書き出すよう指示され、記述する際は単語の羅列ではなく、文章の形で書くことを求められた。本調査で使用したテキストの数は4つと多かったため、協力者は課題に解答する際に、内容を想起する手がかりとして各英文のタイトルを与えられた。

5.4 手順

調査は個別、もしくは4名程度をまとめて一斉に実施された。初めに実験の目的と、個人情報は特定不可能な形で厳重に管理されることを伝え、実験参加への承諾を得た。続いて実験全体の流れと、リーディングテストの問題解答、選択肢の自信度回答の方法について、例を提示しながら説明を行った。

説明の後、協力者は2つの英文についての読解テスト（20分）、筆記再生課題（時間無制限）の順に紙面上で解答を行い、5分間の休憩の後に同様の手順を繰り返し、合計4つの英文について、読解テストと筆記再生に取り組んだ。読解する4つの英文の順番は、カウンターバランスをとった。また、通常のテスト実施時と同様、設問解答の際にはテキストを参照することを可能とした。調査はほとんどの場合、90分程度で終了した。

5.5 採点・分析

5.5.1 リーディングテスト

多肢選択式問題については、正答1項目につき1点を与え、合計得点を算出した。また、各選択肢が選ばれた割合を算出した。

分析について、本調査では有意水準を $\alpha = .05$ として設定し、 $\alpha < .10$ の場合に有意傾向とみなすこととする。球面性が仮定されない要因については Greenhouse Geisser 法による修正を探用し、多重比較については Bonferroni 法による修正を行った。

5.5.2 自信度アンケート

回答データは、受験者の正誤判断と自身の解答に対する自信度に応じ、1~6の6段階でコーディングされた（表11）。これ以降では、正誤判断と解答の自信度を組み合わせることで得られた評定値を「魅力度」として扱う。

表11: 正答選択肢・錯乱肢の魅力度の算出方法

回答自信度	正誤判断	魅力度
高	正	6
中	正	5
低	正	4
低	誤	3
中	誤	2
高	誤	1

5.5.3 筆記再生課題

Ikeno (1996) に基づき、英語教育学を専攻する大学院生2名と調査者により4つの英文をアイディアユニット(IU)に分割した（一致率：91.90%、93.46%）。採点では各IUにつき、約3分の2の情報が再生された場合に1点を与えた。初めに解答データ30%について、大学院生2名と調査者がペアで採点し（一致率：96.57%、90.18%），不一致点を協議で解決した後、さらに明確化した基準に基づき、残りのデータは調査者1名が採点した。

5.6 結果

5.6.1 リーディングテストの信頼性

本調査で実施した全20項目のテストについて信頼性係数を算出したところ、 $\alpha = .69$ であった。そのため、信頼性が低い4項目を除外したところ、

16項目で $\alpha = .72$ の値が得られた。本調査の項目数およびサンプルサイズは小さいため、ある程度の信頼性が確保されたといえる。表12は、項目を修正した後のリーディングテストの記述統計および内的一貫性を示している。

■表12: リーディングテストの記述統計と信頼性係数 (K=16)

N	M	SD	Max	Min	α
42	12.60	2.71	16	5	.72

(注) 最高点は15点。

5.6.2 リーディングテストの妥当性

本調査のマテリアルは、英検で使用されたテスト項目を用いたため、基本的にはテストとしての妥当性は満たされていると考えられる。ただし、留意すべき点として、元の選択肢において含まれていた語を改編して作成した項目であったことが挙げられる。

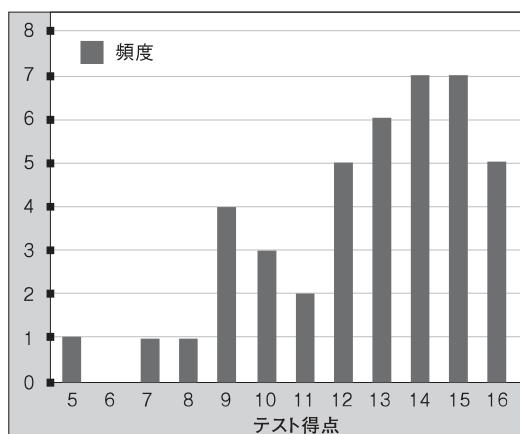
Messick (1995) はテストの構成概念妥当性に關し、その中において、(a) 内容的側面、(b) 本質的側面、(c) 構造的側面、(d) 一般化可能性の側面、(e) 外的側面、(f) 結果的側面の6つの側面があると規定している。本調査では、この中の (b) 本質的側面に関して、使用したテスト項目の妥当性検証を行った。具体的には、各協力者の多肢選択式読解テストの得点と筆記再生課題で得られた再生率との相関を分析し、テスト解答時に達成されているテキスト理解から、読解後の記憶表象に保持されている内容の程度は予測できるかどうかを検証した。

筆記再生率のデータが正規分布を満たさないことから、ノンパラメトリック検定であるスピアマンの順位相関係数 (ρ) を用いて分析したところ、 $\rho = .783$ ($p = .003$) という強い相関が示された。これより、本調査で使用した読解テストは、英文の表層的記憶・命題的テキストベースの理解を測定しているというある程度の妥当性が確認された。

5.6.3 テスト得点

分析対象とした16項目について、協力者の平均得点、標準偏差を表13に示す。全協力者42名の平

均値は12.60点であった。今回のテスト得点別の人數分布の傾向(図1)、および、上位群27%、下位群27%、中位群46%に分けることが一般的な言語テスト分野の慣習に基づき(Hughes, 2003)、協力者のうち11点以下の者を下位群($n = 12$)、12～14点の者を中位群($n = 18$)、15、16点の者を上位群($n = 12$)に分類した。一元配置分散分析により、上中下群間の総得点には有意な差があることが確認された、 $F(2, 39) = 106.67$, $p < .001$, $\eta^2 = .85$ 。



■図1: テスト得点別の人數分布

次節からは、協力者全体、および上記のようにして得られたテスト得点上位・中位・下位の3つの群の傾向について詳細に調査する。具体的には、各パラフレージング・レベルの選択肢の選択率と魅力度それぞれについて、全体的傾向を概観した後、テストの得点群ごとの傾向について調査する。

■表13: テスト得点群別の協力者の得点

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
上位	12	15.42	0.51	15.00	16.00
中位	18	13.11	0.83	12.00	14.00
下位	12	9.00	1.71	5.00	11.00
合計	42	12.60	2.71	5.00	16.00

5.6.4 選択肢の選択率

表14に各選択肢が選ばれた割合を示す。まず、これらの4つの選択肢の選択率について、一元配置分散分析によって比較したところ、NCの選択率がSRよりも有意に高くなっていた($p = .007$)。一方、MIR、MORの選択率については、他

の種類との差は確認されなかった($ps > .05$)。これらのことから、全体的な傾向として、選択肢の記述内に本文で使用される語が50%以上含まれる場合、本文で使用される語をほとんど含んでいない選択肢よりも受験者に選ばれる可能性が高いことがわかった。

■表14: 各選択肢が選ばれた割合

NC		MIR		MOR		SR	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
0.27	0.07	0.26	0.08	0.26	0.08	0.22	0.06

続いて、このような各選択肢の選択率の傾向が、協力者のテスト得点および正答の選択肢の種類によってどの程度異なるのかを検証する。

4つの問題パターンにおける、テスト得点群別の各選択肢の選択率を表15および図2に示す。協力者ごとの各選択肢の選択率を従属変数とし、3(総得点:上位、中位、下位) × 4(問題パターン:P1, P2, P3, P4) × 4(選択肢の種類:NC, MIR, MOR, SR)の三元配置分散分析を行った。その結果、選択肢の種類の主効果、 $F(3, 117) = 3.46$, $p = .019$, $\eta^2 < .01$ 、および、問題パターン×選択肢の種類の交互作用、 $F(6.68, 260.55) = 321.50$, $p < .001$, $\eta^2 = .79$ 、問題パターン×選択肢の種類×総得点の交互作用、 $F(13.36, 260.55) = 15.36$, $p < .001$, $\eta^2 = .08$ がそれぞれ有意であった。

以下では、有意であった、問題パターン×選択肢の種類×総得点の交互作用の解釈を行う。この交互作用についてさらに検証するために下位検定を行った結果、総得点の単純・単純主効果が、3つの組み合わせ(P1: SR, P2: NC, P4: MOR)を除いた13個全ての問題パターン×選択肢の種類の組み合わせにおいて有意であった($ps <$

.05)。つまり、正答選択肢のパラフレージング・レベルと受験者の得点により、各選択肢の選ばれる割合が異なっていたことが示された。

各問題パターン(P)において選択肢が選ばれた割合について、得点群間で多重比較を行った結果、下位群はP1, P2, P4においてそれぞれ正答のレベル(NC, MIR, SR)を選んでいる割合が上位群、中位群よりも低く($ps < .001$), P3において正答のMORを選んでいる割合が上位群よりも低かった($p < .001$)。錯乱肢に関しては、P1においてMIR, MOR($ps < .05$), P2においてMOR($ps < .001$), P3においてSR($ps < .001$), P4においてMIR($ps < .01$)を選んでいる割合が、それぞれ上位群・中位群よりも高いことが示された。さらに、下位群は上位群よりも、P3においてNC($p = .004$), P4においてNC($p = .002$)を選んでいる割合が有意に高かった。

これらの結果をまとめると、正答選択肢がいずれのパラフレージング・レベルであっても、下位群は上位群・中位群よりも正答を選ぶことが困難であり、他のレベルで作成された錯乱肢をより多く選択していたことがわかった。また、上位

表15: 各問題パターンにおける各選択肢の選択率

問題	熟達度	NC		MIR		MOR		SR	
		M	SD	M	SD	M	SD	M	SD
P1 (NC)	上位	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	中位	0.89	0.18	0.04	0.10	0.01	0.06	0.06	0.11
	下位	0.65	0.17	0.21	0.18	0.10	0.13	0.04	0.10
	合計	0.85	0.20	0.08	0.14	0.04	0.09	0.04	0.09
P2 (MIR)	上位	0.04	0.10	0.94	0.11	0.02	0.07	0.00	0.00
	中位	0.11	0.15	0.82	0.21	0.03	0.08	0.04	0.10
	下位	0.15	0.13	0.52	0.20	0.23	0.20	0.10	0.13
	合計	0.10	0.14	0.77	0.24	0.08	0.15	0.05	0.10
P3 (MOR)	上位	0.00	0.00	0.02	0.07	0.98	0.07	0.00	0.00
	中位	0.07	0.12	0.14	0.13	0.76	0.18	0.03	0.08
	下位	0.15	0.13	0.06	0.11	0.65	0.25	0.15	0.13
	合計	0.07	0.11	0.08	0.12	0.79	0.22	0.05	0.10
P4 (SR)	上位	0.00	0.00	0.04	0.10	0.02	0.07	0.94	0.11
	中位	0.06	0.11	0.06	0.11	0.08	0.15	0.81	0.16
	下位	0.15	0.13	0.23	0.20	0.19	0.24	0.44	0.19
	合計	0.07	0.11	0.10	0.16	0.10	0.17	0.74	0.25

(注) P = パターン (pattern)。()内は正答選択肢の種類を示す。

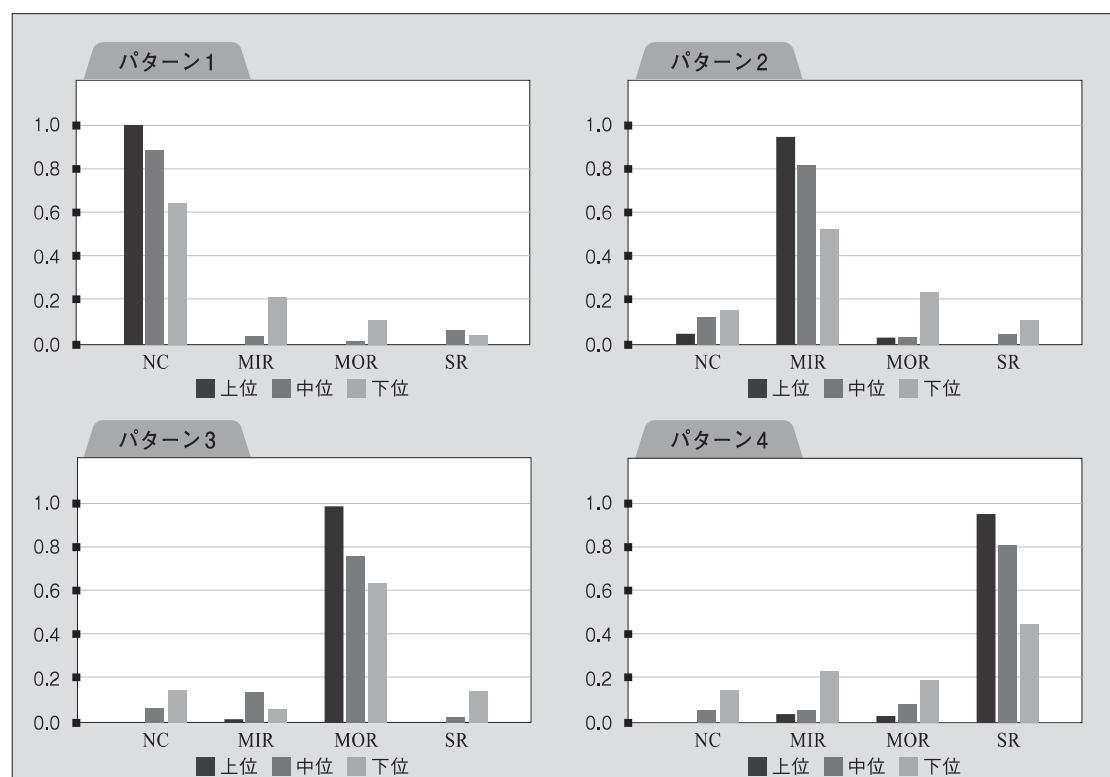


図2: 各問題パターンにおける各選択肢の選択率

群に比べ、MOR, SRといった正答選択肢のバラフレージング・レベルが低い場合に、テキストと重複する語の割合が最も高いNCの錯乱肢をより多く選択する傾向があったことが示された。

一方、中位群と上位群の選択率の間に差が見られた項目は、P3であった。P3において、上位群は中位群よりも正答のMORを選んでいる割合が高かったのに対し($p = .009$)、中位群は上位群に比べ、MIRをより多く選ぶ傾向があったことがわかった($p = .020$)。すなわち、正答選択肢のバラフレージング・レベルがMORであったとき、上位群に比べ、中位群の受験者はそれを選ばず、テキスト原文との重複している語がより多く含まれるMIRを選ぶ傾向にあったということが示唆された。

5.6.5 選択肢の魅力度

表16は、各選択肢の魅力度の評定値を示している。まず、全体的な4つの選択肢の魅力度について、一元配置分散分析で比較した結果、NCの評定値がMOR($p = .002$), SR($p = .006$)よりも高いことがわかった。MIRについては他の種類

との差は見られなかった($p > .05$)。これらのことから、全体的な傾向として、選択肢の記述内に本文で使用される語が50%以上含まれる場合、本文で使用される語を含む割合が20%程度以下の選択肢よりも受験者を引き付けやすく、正答と判断される確率が高いことがわかった。

次に、各バラフレージング・レベルで作成された正答選択肢および錯乱肢の魅力度について、協力者のテスト得点群ごとに評定値の傾向は異なるのかどうかを検証するため、各テスト得点群の4つの種類の正答選択肢・錯乱肢に対する魅力度評定値を従属変数とし、3(総得点:上位、中位、下位) × 4(選択肢の種類:NC, MIR, MOR, SR)の二元配置分散部分析を行った。

まず、正答選択肢に関しては、選択肢の種類の主効果が有意であり、 $F(3, 117) = 6.58, p < .001, \eta^2 = .07$ 、NCがMORよりも高く($p < .001$)、SRよりも高い傾向にあることが明らかになった($p = .026$)。一方で、総得点の主効果、 $F(2, 39) = 0.18, p = .084, \eta^2 < .01$ 、および、選択肢の種類 × 総得点の交互作用 $F(6, 117) = 0.85, p = .534, \eta^2 = .02$ 、は有意でなかった。つまり、テス

■表16: 各選択肢の魅力度

問題	熟達度	全体		正答		誤答	
		M	SD	M	SD	M	SD
NC	上位	2.44	0.18	2.40	0.16	0.92	2.15
	中位	2.65	0.30	2.49	0.25	3.42	2.56
	下位	2.64	0.45	2.34	0.45	4.74	1.58
	合計	2.58	0.33	2.42	0.30	3.08	2.62
MIR	上位	2.39	0.32	2.34	0.31	1.25	2.30
	中位	2.52	0.39	2.36	0.33	3.69	2.40
	下位	2.69	0.52	2.37	0.43	4.60	1.55
	合計	2.53	0.42	2.36	0.35	3.26	2.50
MOR	上位	2.24	0.24	2.21	0.23	0.83	1.99
	中位	2.23	0.30	2.15	0.26	1.89	2.47
	下位	2.51	0.41	2.20	0.39	4.42	1.49
	合計	2.31	0.34	2.18	0.29	2.31	2.49
SR	上位	2.34	0.22	2.34	0.22	0.00	0.00
	中位	2.33	0.36	2.26	0.34	1.86	2.41
	下位	2.38	0.52	2.17	0.44	4.25	2.09
	合計	2.35	0.37	2.26	0.34	2.01	2.49

トの得点群に関わらず、正答選択肢においてテキスト原文と同じ語が使用されている割合が高い場合、割合が低い場合よりも、受験者に自信をもって選ばれやすかったことがわかった。

一方、錯乱肢に関しては、総得点の主効果が有意であり、 $F(2, 39) = 51.62, p < .001, \eta^2 = .31$ 、魅力度は下位群において上位群($p < .001$)、中位群($p < .001$)よりも高く評定され、さらに、中位群において上位群よりも高く評定されていた($p < .001$)。他の、選択肢の種類の主効果、 $F(3, 117) = 2.44, p = .068, \eta^2 = .03$ 、および、選択肢の種類×総得点の交互作用 $F(6, 117) = 0.63, p = .705, \eta^2 = .02$ 、は有意でなかった。のことから、錯乱肢の記述においてテキスト原文と同じ語を含む割合の高低に関わらず、総得点の高い受験者に比べ、低い受験者ほど、錯乱肢に引き付けられやすかったことが示唆された。

続いて、このような選択肢の魅力度について、選択率と同様に、協力者のテスト得点および正答の選択肢の種類によってどの程度異なるのかを検証する。4つの問題パターンにおける、テスト得点群別の各選択肢の魅力度を表17および図3に示す。協力者ごとの各選択肢の魅力度を従属変数とし、3(総得点:上位、中位、下位) × 4(問題パターン:P1, P2, P3, P4) × 4(選択肢の種類:NC, MIR, MOR, SR) の三元配置分散分析を行った。その結果、問題パターンの主効果、 $F(3, 117) = 7.10, p < .001, \eta^2 < .01$ 、および、問題パターン×選択肢の種類の交互作用、 $F(6, 11, 238.25) = 307.14, p < .001, \eta^2 = .76$ 、問題パターン×選択肢の種類×総得点の交互作用、 $F(12, 22, 238.25) = 12.49, p < .001, \eta^2 = .06$ がそれぞれ有意であった。

ここからは、有意であった、問題パターン×選択肢の種類×総得点の交互作用の解釈を行う。この交互作用についてさらに検証するために下位検定を行った結果、総得点の単純・単純主効果が、5つの組み合わせ(P1: MOR, SR, P2: NC, SR, P3: MIR)を除き、11個全ての問題パターン×選択肢の種類の組み合わせにおいて有意であった($ps < .05$)。つまり、正答選択肢のパラフレージング・レベルと受験者の得点により、各選択肢の魅力度が異なっていたことがわかった。

各問題パターンにおける選択肢の魅力度につ

いてテスト得点群間で多重比較を行った結果、P1, P2, P4の正答の選択肢(NC, MIR, SR)の下位群に対する魅力度は上位群、中位群よりも低く($ps < .01$)、また、P3の正答MORの下位群に対する魅力度は上位群よりも低かった($p = .005$)。一方、錯乱肢に関しては、P1のMIR($ps < .05$)、P2においてMIR, MOR($ps < .01$)、P4のMIR($ps < .05$)の下位群に対する魅力度が、それぞれ上位群・中位群よりも高いことが示された。さらに、下位群は上位群に比べ、P3においてNC($p = .004$)、P4においてNC, MOR($ps < .01$)の魅力度評定値が有意に高かった。

これらの結果をまとめると、どのパラフレージング・レベルで作成された正答選択肢であっても、下位群は上位群・中位群よりも強い自信をもって正答と判断することが困難であり、他のレベルで作成された錯乱肢により引き付けられていたことがわかった。また、上位群に比べ、MOR, SRといった正答選択肢のパラフレージング・レベルが低い場合に、テキストと重複する語の割合がより高いNC, MORの錯乱肢をより魅力度を感じる傾向があったことが示された。

一方、中位群と上位群の差に関しては、P3におけるMOR, MIRで差が見られた選択率の結果とは異なり、いずれの項目においても中位群と上位群の魅力度評定値の間に統計的有意差は確認されなかった。

5.7 結果のまとめと考察

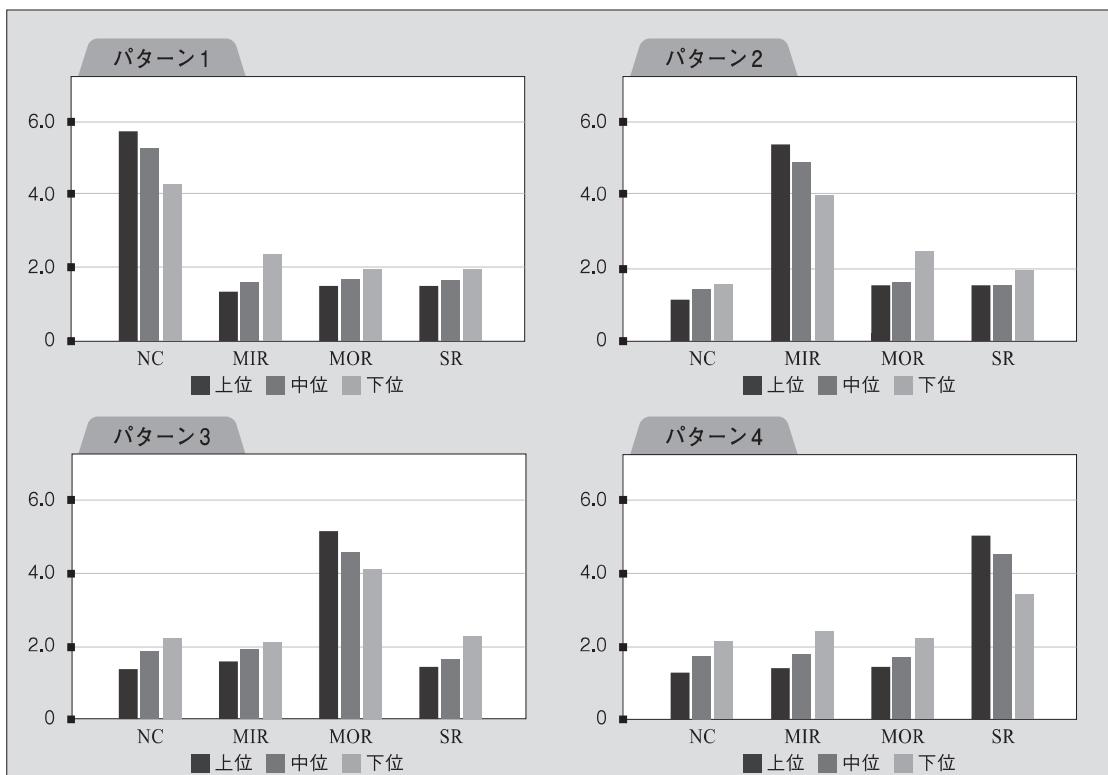
本調査において得られた主な結果は、以下の4点にまとめられる。

- 全体的に、テキスト原文と同じ語が50%以上含まれる選択肢は、本文で使用される語を0~20%程度しか含まない選択肢よりも受験者に選ばれやすく、自信をもって正答だと判断されやすい。
- 錯乱肢の記述においてテキスト原文と同じ語を含む割合の高低に関わらず、高得点の受験者よりも低得点の受験者ほど、錯乱肢を魅力度的であると判断する。

■表17: 各問題パターンにおける選択肢の魅力度

問題	熟達度	NC		MIR		MOR		SR	
		M	SD	M	SD	M	SD	M	SD
P1 (NC)	上位	5.67	0.34	1.31	0.36	1.42	0.51	1.44	0.40
	中位	5.19	0.73	1.57	0.53	1.61	0.65	1.63	0.48
	下位	4.21	0.73	2.31	0.94	1.90	0.59	1.90	0.67
	合計	5.05	0.85	1.71	0.74	1.64	0.61	1.65	0.54
P2 (MIR)	上位	1.15	0.35	5.23	0.52	1.54	0.53	1.50	0.44
	中位	1.43	0.44	4.76	0.65	1.64	0.40	1.54	0.43
	下位	1.60	0.57	3.92	0.86	2.46	0.66	1.94	0.83
	合計	1.40	0.48	4.65	0.84	1.85	0.64	1.64	0.59
P3 (MOR)	上位	1.38	0.38	1.58	0.43	5.13	0.56	1.42	0.40
	中位	1.86	0.56	1.93	0.67	4.49	0.74	1.65	0.66
	下位	2.21	0.76	2.13	0.57	4.08	0.92	2.27	0.62
	合計	1.82	0.65	1.89	0.61	4.55	0.83	1.76	0.67
P4 (SR)	上位	1.27	0.27	1.42	0.39	1.44	0.40	5.02	0.54
	中位	1.74	0.60	1.81	0.53	1.72	0.54	4.50	0.67
	下位	2.15	0.55	2.42	0.65	2.25	0.88	3.40	0.82
	合計	1.72	0.60	1.87	0.65	1.79	0.69	4.33	0.92

(注) P = パターン (pattern)。()内は正答選択肢の種類を示す。



■図3: 各問題パターンにおける各選択肢の魅力度

c 正答選択肢がテキスト原文と同じ語を含んでいる割合の高低に関わらず、低得点の受験者は高・中程度の得点の受験者よりも正答を選ぶことが困難であり、他の錯乱肢をより多く選択した。特に、高得点の受験者に比べ、正答選択肢がテキスト原文と同じ語を含んでいる割合が低い場合に、テキストと重複する語の割合がより高い錯乱肢に引き付けられやすい。

d 正答選択肢のテキスト原文と同じ語を含む割合が1~19%の場合、高得点の受験者は正答を中程度の得点の受験者よりも多く選ぶ一方、中程度の得点の受験者はテキストと重複する語の割合がより高い(20~49%)錯乱肢を選ぶ傾向がある。

まず、aに関しては、先行研究から予測された点と一致する。Ushiro et al.(2008)では、多肢選択式テストへの解答時には、受験者が選択肢から得られる情報に依存し、特にテキストと選択肢の語が重複しているものを選ぶ傾向があることが示唆されていた。本調査においても、選択肢が選ばれた割合、そして選択肢の魅力度の観点からも、受験者が選択肢とテキストの語が重複している程度が高いほど魅力的であると判断し、引き付けられる傾向にあったと考えられる。

次に、bに関しては、先行研究と同様の結果が得られた。多肢選択式リスニングテストの選択肢の魅力度を調査したIimura(2014)において、熟達度の低い受験者は高い受験者よりも錯乱肢を魅力的だと判断する傾向があることが示されていた。本調査においても同様に、読解テストにおけるパラフレーズ質問においてパラフレージング・レベルを操作した選択肢を用いた場合、低得点の受験者は高・中程度の得点の受験者よりも、中程度の得点の受験者は高得点の受験者よりもテキストの理解が十分でなかったために、錯乱肢をもっともらしく感じる度合いが高かった可能性が考えられる。

続いてcであるが、高・中程度の得点の受験者は低得点の受験者よりも正答選択肢を多く選んでいることが明らかであるため、これは当然の結果であるといえる。しかし、低得点の受験者は高得点の受験者よりもテキスト理解が十分でな

かったために、テキスト原文と同じ語を含んでいる割合が低い正答選択肢を選ぶことができず、テキストと重複する語の割合がより高い錯乱肢に引き付けられる傾向にあったと考えられる。

dに関しては、本調査において唯一、高得点・中程度の得点の受験者を区別する結果となった。すなわち、中程度の得点の受験者は高得点の受験者よりもテキスト理解が十分でなかったために、テキスト原文と同じ語を含んでいる割合が低い正答選択肢を選ぶことができず、テキストと重複する語の割合がより高い錯乱肢に引き付けられたと考えられる。このことから、パラフレーズ質問の選択肢においてテキストの理解が中程度の受験者と十分に達成されている受験者を分ける閾値は、このMIR(20~49%)、MOR(1~19%)というレベルである可能性が考えられる。一方で、魅力度の侧面から考えると、正答をMORとするP3において、MIRに対する魅力度評定値に、中・高得点の受験者間で有意差は見られなかった。つまり、中程度の得点の受験者は自信をもって錯乱肢であったMIRを正答と判断したわけではなく、判断を迷った結果として、消極的に選択したという可能性も考えられる。

6

結論と今後の課題

本研究では2つの調査を通して、多肢選択式読解テストにおいて、パラフレージング・レベルをパラフレーズ質問の選択肢作成に利用する可能性を検証した。以下に、本研究から得られた主要な結果をまとめると。

調査1では、英検、TOEIC、TOEFLの読解問題においてパラフレーズ質問を対象に、パラフレージング・レベルの観点から、正答・誤答の選択肢が含む語と英文中の語はどの程度重複しているのかについて調査を行った。結果から、テストの種類によって一部のパラフレージング・レベルの傾向が異なることが示された。全体的な傾向としてはMIRが最も多く、約半数を占めており、次いで多かったのはNCであった一方、MOR、SRなどパラフレージング・レベルが低い選択肢の割合に関してはテスト間に違いがあることが明ら

かになった。英検やTOEFLではMOR, SRの正答選択肢の数が少ないので対し、TOEICでは他の2つのテストに比べてMOR, SRの正答・誤答選択肢が多い傾向があることがわかった。

続く調査2では、パラフレージング・レベルにより作成された、テキスト原文で使用される語を含む割合が異なる正答選択肢・錯乱肢、およびその組み合わせが効果的に機能するかどうかについて、選択率および選択肢の魅力度の観点から検証した。結果、正答選択肢・錯乱肢の組み合わせによる選択率と魅力度への影響は、テスト得点群によって異なることが示された。選択率に関して、テスト得点が低い又は中程度の受験者は、高得点の受験者よりも、正答選択肢のパラフレージング・レベルが低い場合に、よりレベルが高い錯乱肢を多く選択することが示された。魅力度については、テスト得点が低い受験者は高得点の受験者よりも、正答選択肢のパラフレージング・レベルが低い場合に、それより高いレベルの錯乱肢をより魅力的であると感じることが示された。

総括すると、パラフレージング・レベルの異なる錯乱肢の機能は、正答選択肢との組み合わせ、および、受験者の得点によって異なることが明らかになった。すなわち、高得点の受験者にとっては、正答選択肢および錯乱肢のパラフレージング・レベルに関わらず、正答を選んでいた割合は同程度であった。一方、得点が低い又は中程度の受験者の場合は、高得点の受験者に比べ、選択肢のテキスト原文の語を含む割合の高低に影響を受け、特に正答選択肢のパラフレージング・レベルが低い場合に、よりレベルが高い錯乱肢を多く選ぶ割合が高かった。特に、中程度の得点の受験者と高得点の受験者を区別する境目となつたのは、正答が原文に含まれる語を1~19%程度含む場合に、それよりも多い割合(20~49%)で含む錯乱肢に引き付けられるか否かであった可能性が示された。一方で、魅力度の観点から考えると、中程度の得点の受験者は錯乱肢であったMIRについて、判断を迷った結果、消極的に選択したという可能性も否定できない。このような受験者の誤った選択を導く原因を明らかにするには、今後の研究において、オンラインの手法を併用しながら解答プロセスをより詳細に調べる

必要があるだろう。これらの結果より、弁別力のあるパラフレーズ質問の項目を作成する際には、選択肢のパラフレージング・レベルすなわち文章中の記述を含む割合について、そして正答選択肢と錯乱肢の組み合わせ、受験者の判断する選択肢の魅力度を合わせて考慮することが有用であることが示唆された。

これらの調査によって、本研究はパラフレーズ質問の項目作成における、選択肢の表層的な性質について検討する意義を見直す機会となつたといえる。表層的な性質はもっともらしさとの関わりが希薄である可能性を指摘する研究もあるが、パラフレーズ質問など、質問タイプによっては難易度を左右しうる要因として考慮すべきではないかと考えられる。さらに、本調査では、選択肢に対する回答自信度と正誤判断を組み合わせた魅力度を導入したこと、各選択肢に対する学習者の認知面についても詳細に調べることが可能となった。今後も多角的な手法により、選択肢要因と学習者の認知面が統合的に調査されることが望まれる。

最後に、本研究における限界点として、調査2において作成した読解テストの信頼性、妥当性、項目数、一般化可能性が十分ではなかったことが挙げられる。特に、今回基盤のテストとして用いたのは英検2級の読解問題のパラフレーズ質問であった。ゆえに、本研究での成果について、英検の他の級においても同様の結果が得られるのかどうか、さらに、英検の形式以外の読解テストに応用することができるかどうかについては、今後さらなる調査が必要となるだろう。本調査ではパラフレージング・レベルに焦点を当てたが、その基準の再設定も含め、より幅広い観点から多肢選択式問題における効果的な錯乱肢の性質について検証することが望まれる。

謝 辞

本研究を実施する機会を与えてくださいました公益財団法人 日本英語検定協会の皆さま、選考委員の先生方に厚く御礼申し上げます。特に、私の研究を担当してくださいました村木英治先生には、貴重なご助言を賜りましたこと深く感謝申し上げます。

そして、筑波大学大学院の卯城祐司先生には、本研究の計画から実施、報告書の作成に至るまで親身にご指導をいただきましたこと、心より御礼申し上げます。最後に、本調査を実施するにあたり、採点作業や調査にご協力をいただいた皆様に厚く御礼申し上げます。

参考文献(*は引用文献)

- * Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, 16, 486-514.
- * Educational Testing Service. (2013). The Official Guide to the TOEFL Test (4th Edition).
- * Haladyna, T.M., & Downing, S.M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in education*, 2, 37-50.
- * 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, 309-333.
- * Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- * Iimura, H. (2014). Attractiveness of Distractors in Multiple-Choice Listening Tests. *JLTA Journal*, 17, 19-39.
- * Ikeno, O. (1996). The effects of text-structure-guiding questions on comprehension of texts with varying linguistic difficulties. *JACET Bulletin*, 27, 51-68.
- * Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15, 261-278.
- * Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19, 193-220.
- * 国際ビジネスコミュニケーション協会TOEIC運営委員会(編). (2016).「公式TOEIC Listening & Reading 問題集 1」. 東京:凸版印刷.
- * 国際ビジネスコミュニケーション協会TOEIC運営委員会(編). (2017).「公式TOEIC Listening & Reading 問題集 2」. 東京:凸版印刷.
- * Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- * 根岸雅史. (2017).『テストが導く英語教育改革 無責任なテストへの処方箋』. 東京:三省堂.
- * 日本英語検定協会. (2017).『英検2級一次試験過去問』.
http://www.eiken.or.jp/eiken/exam/grade_2/solutions.html (2017年7月1日閲覧)
- * 旺文社(編). (2003).『英検2級過去6回全問題集』. 東京:大日本印刷.
- * 旺文社(編). (2017).『英検2級過去6回全問題集』. 東京:大日本印刷.
- * 清水真紀. (2005).「リーディングテストにおける質問タイプ—パラフレーズ・推論・テーマ質問と処理レベルの観点から」. *EIKEN BULLETIN*, 17, 48-62.
- * Shizuka, T. (1999a). Combining response correctness and confidence level rating to produce polychotomous data from MC dichotomous items. *Japan Colleges of Technology Education Journal*, 22, 243-252.
- * Shizuka, T. (1999b). Combining response correctness and confidence level rating to produce polychotomous data from MC dichotomous items (II). *The Kanto-Koshinetsu Association of Teachers of English (KATE) Bulletin*, 13, 73-86.
- * Shizuka, T. (2003). The effect of clustered objective probability scoring with truncation (T-COPS) on reliability. In H. Yanai, A. Okada, K. Shigematsu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 273-280). Tokyo: Springer.
- * 高木修一. (2011).「多肢選択式問題は公平か—記憶保持の強さから」. *EIKEN BULLETIN*, 23, 13-29.
- * 卯城祐司(編). (2012).『英語リーディングテストの考え方と作り方』. 東京:研究社.
- * Ushiro, Y., Morimoto, Y., Hijikata, Y., Nakagawa, C., Watanabe, F., Kai, A., Shimizu, H., Koga, T., Ohno, M., Umehara, C., Hamano, Y., Narumi, T., Tao, N., Shimada, S., Kiyama, M., Suzuki, R., Kurisu, Y., & Gomi, N. (2007). What makes distractors plausible in multiple-choice reading tests? *JLTA Journal*, 10, 56-67.
- * Ushiro, Y., Nakagawa, C., Morimoto, Y., Hijikata, Y., Watanabe, F., & Kai, A. (2008). Effects of question types on item difficulty in two reading test formats: Open-ended and multiple-choice. *Annual Review of English Language Education in Japan (ARELE)*, 19, 201-210.
- * Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

資料1 調査2で用いたテキスト例(英検2級2001年度第2回4Aより).....

Green Cleaners

One of the biggest problems facing the world today is how to clean up areas that have been polluted by industry or war. Often these areas contain dangerous metals or chemicals that can poison people or wildlife. Until now, the main way to clean up such areas has been to dig up the polluted earth and take it away. One problem with this method is that it is very expensive. Moreover, this simply moves the pollution to a different place rather than truly solving the problem.

Recently, scientists have been studying plants and trees to see if they can provide a better solution to the pollution problem. People have known for a long time that some plants and trees are good at removing dangerous metals and chemicals from the earth. Scientists are now studying exactly how they do this in order to develop new kinds of plants that may be used to clean up polluted land. Scientists have actually discovered that there are two main ways in which plants help to clean up the earth. One way is by directly absorbing metals and chemicals through their roots and storing them in their branches and leaves. The other way is by attracting insects and other small creatures that can consume dangerous materials in the earth.

Scientists have made important progress with these methods. In America, for example, a kind of grass has been used to clean up areas polluted by oil. But such methods also have some problems. One problem is that they take a long time to work. Even though scientists have developed plants that work more quickly, it still takes many years to clean up an area by using them. Another problem is that when animals eat such plants, they can be poisoned or spread toxic material to other animals or people. Plants cannot be used in every case of pollution. But it seems likely that in the future they will play a major role in helping to repair the damage human beings have caused to nature.

資料2 調査2で用いた多肢選択式問題の例

(注) () 内は選択肢のパラフレージング・レベルを表す。(実施時には無記載)

※ NC = Near Copy, MIR = Minimal Revision, MOR = Moderate Revision, SR = Substantial Revision.

- (1) The traditional way to clean up polluted areas

解答

- 1 is an inexpensive means to deal with the problem. (MOR)
- 2 can poison human beings and wildlife nearby. (NC)
- 3 does not function in places with dangerous metals. (MIR)
- 4 does not really have a chance at resolving the issue. (正: SR)

自信度

低	中	高
低	中	高
低	中	高
低	中	高

- (2) Scientists

解答

- 1 have realized that bugs carry harmful substances. (SR)
- 2 are researching the relationship between harmful metals and substances. (MOR)
- 3 are trying to use plants to solve the problem of earth pollution. (正: NC)
- 4 have found that plants can release poisonous metals. (MIR)

自信度

低	中	高
低	中	高
低	中	高
低	中	高

- (3) What is one function of plants and trees?

解答

- 1 They utilize metals and substances for growing rapidly. (MOR)
- 2 They absorb metals and chemicals and store them inside. (正: MIR)
- 3 They keep small creatures away from polluted areas. (NC)
- 4 They hold waste underground so that it is less harmful. (SR)

自信度

低	中	高
低	中	高
低	中	高
低	中	高

- (4) What is one problem of using plants to clean up polluted earth?

解答

- 1 Scientists have not been able to make plants work more quickly. (NC)
- 2 It takes a long time to plant a region with plants and trees. (MIR)
- 3 There is a possibility that animals will feed the harmful plants. (正: MOR)
- 4 Forests are seriously harming the environment globally. (SR)

自信度

低	中	高
低	中	高
低	中	高
低	中	高

- (5) Which of the following statements is true?

解答

- 1 Plants are acting as a rapid solution to the pollution problem. (MIR)
- 2 Military conflicts are one of the sources of soil contamination around society. (正: SR)
- 3 Researchers have created plants that do not harm the ground. (MOR)
- 4 Animals are more effective at removing pollution than plants. (NC)

自信度

低	中	高
低	中	高
低	中	高
低	中	高