

第34回 研究助成

**A 研究部門・報告 I • 英語能力テストに関する研究**

# マクロルールの階層性に基づく英文読解問題の提案: 詳細情報理解から概要把握まで

研究者: 茨城県／筑波大学大学院 在籍 佐藤 連理

《研究助言者: 寺内 一》

**概要**

要約課題に取り組む際に学習者が辿る規則として、マクロルール（削除、一般化、構成）が存在する。本研究では、この規則に基づき詳細情報から概要把握までを問うことが可能な多肢選択式設問を作成し、既存の設問及び要約課題との比較検証を行った。調査1ではリーディングテストに含まれる設問を分類した結果、英検では準1級と2級を除き、難易度の高い級の設問の方がより難しいルールを含むなど、その構成内容に有意差が見られた。一方で、高校卒業程度とされている2級も削除することで解答できる設問の割合が高く、一般化や構成を用いて概要を問う設問は依然として少なかった。調査2では、①新規に作成した一般化、構成を同程度含んだテスト、②従来の削除設問を多く含んだテスト、③要約テストの3つを大学生・大学院生を対象に実施した結果、多肢選択式テストでは削除と比較して構成設問の方が有意に正答率が低く、また要約では見られなかった一般化・構成についても多肢選択式で問われれば正答できることが示された。得られた結果について、マクロルールに基づく多肢選択式テストの妥当性という観点から考察を行った。

**1****はじめに**

平成29年から30年にかけて行われた学習指導要領の改定により、外国語科の目標は育成したい

資質・能力が「知識及び技能」「思考力、判断力、表現力等」「学びに向かう力、人間性」という3つに整理され、特に高等学校学習指導要領では以下のような目標が立てられた。それは、「外国語によるコミュニケーションにおける見方・考え方を働きかせ、外国語による聞くこと、読むこと、話すこと、書くことの言語活動及びこれらを結びつけた統合的な言語活動を通して、情報や考えなどを的確に理解したり適切に表現したり伝え合ったりするコミュニケーションを図る資質・能力を次の通り育成すること(MEXT, 2018)」である。特に「思考力、判断力、表現力等」では「日常的な話題や社会的な話題について、外国語で情報や考えなどの概要や要点、詳細、話し手や書き手の意図などを的確に理解」することができる力の育成を目標としており、中学校、高等学校のどちらにおいても読解において1文ごとの意味理解(下位処理)だけでなく書き手の大まかな意図を捉えること(上位処理)が求められるように変化している。

これまで、字義レベルの理解は短文解答や多肢選択式課題で測られることが多い一方で、概要の把握には要約課題が広く使用されていた。これは、読み手はテキストの内容を理解した上で重要情報を選択・統合することで文全体の概要を理解することから、要約課題を課すことで読み手がテキストの表層的な部分のみを理解しているのか、あるいはテキストで使用されている語や語順から離れて変形できるレベルなのかを測定することができるからである(Sherrard, 1989; 邑本, 1992)。この要約課題の1つの特徴として、元あっ

たテキストよりも少ない語数でその内容をまとめることが必要となることが挙げられる。その際、学習者が内容を凝縮する方法には法則性があることが示されている(Kintsch & van Dijk, 1978)。学習者がテキストを読み、その後に内容を書き起こすというのは技能統合型課題に分類され、その認知的な負荷は高いものの、新学習指導要領でも求められている活動の一つである。しかしながら、このような活動では学習者の解答が多岐に渡るため、採点に時間がかかる。また、複数人での採点においてはテストの評価者間信頼性確保の観点から、特に大規模なテストでは実施が困難であるという課題も存在する。以上を踏まえると、詳細情報を多肢選択式テストで問うように、要約課題で測れる内容を多肢選択式に応用できるようなテストの作成が求められていると言えるだろう。そこで、本研究では、学習者が要約作成時に辿る処理を3つに分類したマクロルールを基に、その階層性を利用して詳細理解から大意把握の間の段階的なテキスト理解度を測定する設問を検討する。より具体的には、以下の3つの点を検証し、効果的な読解テスト作成に示唆を与える。

- (1) 多肢選択式を用いた大規模テストにおける長文セクションでは、どのようなマクロルールの使用が求められるのか。また、難易度が異なるテストでは、マクロルールの使用に違いが見られるのか。
- (2) マクロルールを応用した新型の多肢選択式課題と、調査1に基づく従来の項目を用いた旧型の多肢選択式課題を比較すると、その信頼性・妥当性に違いが見られるのか。
- (3) 新型テスト、旧型テスト、要約テストの3つにおけるマクロルールの使用や正答率に違いはあるのか。

## 2

## 先行研究

### 2.1 要約課題と認知的負荷

近年、要約課題や議論型のエッセイ課題など、

技能統合型のタスクであるReading-to-Writeタスクの使用頻度が増えている。その理由として、新学習指導要領に記載されているというのももちろんだが、タスクをより現実世界での生活に近づけるため(Weigle et al., 2013)ということもあるのだろう。しかしながら、要約課題には指導者にとって採点に時間がかかり、信頼性の確保が難しい。さらに、学習者にとっても負荷の高い課題であるとされている。

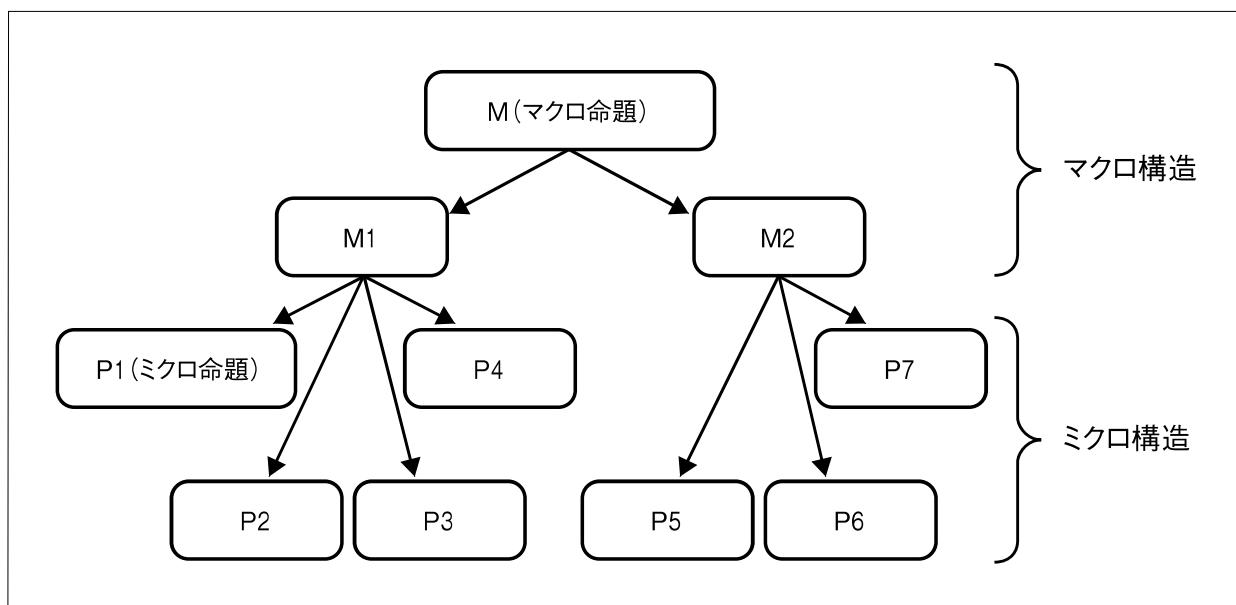
要約課題とは読んだテキスト内容を限られた字数内で説明する課題であり、学習者がテキストに書かれていた内容全てを出力するだけでは達成できないものである。凝縮を必要とする課題であることから、読み手の心的表象を測る手立てとして広く利用されている。心的表象とは、文章を読んだ際に脳内に残るテキストの内容やイメージのことである。これは表層的記憶、命題的テキストベース、状況モデルと呼ばれる3つの段階を持ち(van Dijk & Kintsch, 1983)、状況モデルとは、文章からの情報と読み手がそれ以前に持っていた知識が統合された、テキストによって記述される状況全体の理解のことである。読み手は英文を読む際、情報を削除したり統合したりして、一貫した心的表象を構築する。要約課題では、この読み手の心的表象を反映するため、読み手がどの程度文章内容を理解することができたかを測定することができるのだが、その性質上読み手が情報の取捨選択や統合を行う必要がある。

しかしながら、この統合処理の存在が要約など技能統合型課題を難しくさせる要因の一つでもある。人間の認知資源は限られており、特にL2の学習者はその資源の多くを翻訳に割くことから(Horiba, 2000)、文脈間のつながりを推測したり、テキスト全体を俯瞰したりするなど対局的な処理が難しくなるのである。Segev-Miller(2007)の研究においても学生を対象に主にプロセスログを用いて実験を行った結果、大学生レベルの協力者であってもソーステキストの統合に困難を抱えるなどその認知的負荷が高いことを示している。これについてはL1研究(Mateos et al., 2008; Segev-Miller, 2007; Spivey & King, 1989; Mateos et al., 2008)だけでなくESL/EFL研究(Yang & Plakans, 2012; Yang, 2014)においても同様に示されている。

また、日本人 EFL 学習者を対象とした Ushiro et al. (2008) の研究では要約課題において局所的な統合は可能であるものの、パラグラフや文章全体を結びつけるような統合は依然として困難であることが示された。

## 2.2 テキストの階層性とマクロルール

前記で述べた情報の取捨選択及び統合が要約作成に必須のプロセスなのだが、このような段階的な理解はテキストの階層性とも関連している。図1が示すように、文章は大きくマクロ構造とミクロ構造によって構成されているのである (Kintsch, 1988)。



■図1: テキストの階層性

1つの述部と1つ以上の項からなる意味単位を命題と呼び、ミクロ構造はその命題を単位とする局所的な下位概念のことである。その一方で、マクロ構造とは、複数のミクロ命題を取り扱い、統合したもので、上位概念を表すことができる。つまり、ミクロ命題とはテキストに含まれる局所的な情報のことであり、テキスト読解時に学習者は語を解析し、意味のまとまりを理解するなど下位処理を行った後、命題や文、パラグラフレベルの情報を統合したり、既知の知識と組み合わせたりするなどの上位処理を行い、テキストを大局的に理解していく。なお、ミクロ命題がテキストに明示的に書かれている情報であるのに対し、マクロ命題はトピックセンテンスとしてテキストに明示される場合もあれば、明示されておらず学習者が自ら生み出すことが必要となる場合もある。

特に明示されていないマクロ命題を構築するため、学習者が情報を選び、統合するこの規則を要約規則、あるいはマクロルール (Kintsch &

van Dijk, 1978; van Dijk & Kintsch, 1983) と呼ぶ。マクロルールは先行研究 (Brown & Day, 1983; Keck, 2006) によって分類方法に多少の異なりが見られるが、本調査では Kintsch and van Dijk (1978) 及び van Dijk and Kintsch (1983) が提唱した「削除」「一般化」「構成」という3つのマクロルールに基づいて実験を実施する。

第一のルールである削除とは、その名の通りマクロ命題を構成するにあたり、不要なミクロ命題を削除するというものである。例えば、“Mary played with a ball. The ball was blue.” という2つの文を1つにしようとする場合、「青い」という情報が不要だと判断されれば “Mary played with a ball.” とその情報が削除されたものがマクロ命題として読み手に認識される。

第二のルールである一般化とは、複数の命題（語）をより一般的で上位の命題に置き換えることでマクロ命題を構成するというものである。例として、“Mary played with a doll. Mary

played with blocks.”という2文を“Mary played with toys.”という1文に要約することができる。この時、具体的に示されていたa doll及びblocksという名詞がより抽象化された、上位概念であるtoysに変換されていることがわかるだろう。

最後に、第三のルールとして構成が存在する。これは最も難易度の高いルールであり、読み手がテキストを読み、本文に明示的に示されていないマクロ命題を作り出すというものである。テキストで述べられた内容をそのまま用いるのではなく、一連の動作をより包括的な内容を表す上位表現に作り變えるなど、学習者に推論することが求められるのである。具体的には、“I went to the station, bought a ticket, and took a train.”という1文から “I traveled by train.”というマクロ命題を得ることができるだろう。この構成の難しさについてはUshiro et al. (2008)だけではなく、Ushiro et al. (2009)でも述べられている。テキストの語数に対して要約の字数が短ければ、より内容を凝縮する必要があるため、より複雑な上位処理が必要になると考えられる。

また、マクロルールとは異なる切り口から畠本(1992)もL1学習者の要約産出方法を4つに分類している(表1)。こちらにおいても、テキスト依存度というテキスト本文と作成された要約文のオーバーラップに基づいて分類をしており、要約課題においてテキストの存在が学習者の産出

過程に大きく影響を与えること、またそれゆえにコピー & ペーストのような方略をとってしまう学習者が一定数存在することも示されている。依存度が低ければ推論や抽象化(Kintsch & van Dijk(1978)のマクロルールの「構成」に当たる)が起こりやすく、またテキスト中には明示されていない文の要約文における占有率が高くなると述べている。一方で、依存度の高い要約はテキスト情報を取捨選択的に抜き出して要約を作成するやり方であるとしている。それゆえ、同一表現や類似表現、言い換えが増え、依存度の高い要約と比べて推論や抽象化が少ないと述べた。

なお、実際に協力者が作成した要約を分析した実験では、L1と比べてL2ではテキストのコピーが増えることから、L2はソーステキストに依存する傾向にあると報告した研究(Weijen et al., 2019)もあれば、全体としてテキストの使用率が9%と低く、EFLとESL環境にある学習者の習熟度による差異は見られなかった(Weigle & Parker, 2012)という研究結果も存在する。また、Sherrard(1989)は、L2での筆記による要約と口頭による要約を要約作成の方略、読み手の要約能力の向上、要約の評価という3つの観点から分析した。方略に関する分析結果より、熟達度の高い学習者はテキストの表面的な語彙や命題の順序から離れ高度なマクロルールを用いた要約を作成する一方、熟達度の低い学習者はマクロルールにおける削除のみを使用することが明らかとなった。

■表1: 畠本(1992)による4つの要約産出方略とその特徴

方略名	特徴
凝縮型	テキスト依存度が低く、要約文章は比較的短い傾向にある。テキストの表現ではなく、読み手の理解に基づき作成される。推論を含む。
具現型	テキスト依存度が低く、要約文章は長めの傾向にある。 凝縮型と同様に個人の理解に基づくが、文章内容をある程度具体的な表現を用いて記述する。 テキストに明示されている文を用いる傾向にあり、推論が少なくなる。
複写型	テキスト依存度の高い方略で、要約文章は長めの傾向にある。 テキスト内容を要約に含めるか含めないかの決定をし、含める場合にはテキストで使用された表現をそのまま使用する。
換言型	テキスト依存度の高い方略で、テキスト中の表現をそのまま複写するのではなく、自分なりに解釈して記述する。

### 2.3 多肢選択式テストへの応用に向けて

多肢選択式テストは、特に大規模なテストで学習者の読解力を測るのによく用いられる手法である。採点に時間もかからず、評価者間で採点にずれが生まれないということがその普及の大きな理由の1つであると考えられる。一方で、多肢選択式テストとソースベースの意見文を書く課題を比較した Scouller(1998) は L1 学習者のストラテジーを調査したところ、ライティング課題が課されている群では統合などより深いレベルの処理が行われているのに対し、多肢選択式テストに答えるために取り組んでいた群は暗記など表層的なレベルの処理が行われるということを示唆しており、多肢選択式では技能統合型のように大局的で深い理解を問うことの困難さを示唆している。同様に、選択肢が提示されるがゆえに、その語彙の重なりなどから学習者がテキストを十分に理解していなくても正解を推測することができてしまう、といった懸念点も多肢選択式テストには存在する (Kobayashi, 2002)。

上記に示したような限界点を克服できる可能性として、多肢選択式のテストで 2.2 節に示したマクロルールを利用することができる。テキスト内で使用された語彙や表現を消却したり、単語レベルでの言い換えをしたりした選択肢(削除レベル)、テキスト内で複数の語を用いて具体的に示されていた内容を上位概念に置換するなど、抽象化した選択肢(一般化レベル)、そして文と文、パラグラフとパラグラフを統合することで得られるテキストに明示されない情報を問う選択肢(構成レベル)を用いることで、従来の表層的な設問では測りきれなかった学習者の理解段階を問うことができる可能性がある。なお、多肢選択式テストにおいてテキスト本文と選択肢の重なり(オーバーラップ)によって設問の難易度が変わることは先行研究によって示されており (Ushiro et al., 2008)，要約において一般化や構成といったもとのテキストから離れて変形することが難しいように、多肢選択式でもこのオーバーラップが少ない一般化や構成を用いた設問の難易度が高くなると考えられる。

### 2.4 本研究の目的と構成

本研究の目的は、要約時の情報の取捨選択・統合時に行われるマクロルールの階層性を利用し、多肢選択式テストで効果的に学習者の詳細情報理解だけではなく、概要の理解も測定する方法を検討することである。なお、本研究ではこの「概要の理解」、「概要を捉える力」をテキストを読んだのちに一般化及び構成(2.2節参照)することができる力と定義する。テキストに書かれてある情報から、テキストに書かれていなかいう推論を必要とする設問まで含めることで、読み手の理解状況を段階的に測定することが可能になり、多肢選択式テストでは表層的な部分の理解しか測れない、といった既存の問題点の解決にもつながると考える。加えて、パフォーマンステストのように実施及び採点に時間のかかる課題を用いずとも、ペーパーテストで学習者の思考力・判断力・表現力等を問う方法に寄与することができるだろう。

本研究では、現在行われている英語読解テストの設問分析を行い、傾向を掴んだ後、要約課題に代わり、学習者の詳細情報から概要把握までの理解能力を測定しうる設問の作成を試み、従来の要約課題や多肢選択式設問と比較することで信頼性について検証する。この目的を達成するため、本研究では2つの調査を実施する。調査1では、英検に使用される長文問題において、新学習指導要領に示される概要を捉える能力をどのように測定しているかマクロルールの観点から分析し、級ごとの特徴を明らかにする。続く調査2では、階層性を利用して学習者の理解段階を測定する設問を作成し、その設問の妥当性・信頼性を従来の多肢選択式設問及び要約課題と比較し、調査する。本研究における検証課題(Research Questions; RQs)は以下の通りである。

RQ1 英検3級、準2級、2級、準1級において、解答に必要なマクロルール(削除、一般化、構成)の傾向は異なるか。(調査1)

RQ2-1 多肢選択式で実施した新型テストと旧型テストの項目識別力及び信頼性係数に違いはあるのか。(調査2)

RQ2-2

新型多肢選択式テスト、旧型多肢選択式テスト、及び要約テストにおけるマクロルールの使用と正解確率に違いはあるのか。(調査2)

### 3

## 調査1:マクロルールに基づく設問分析

### 3.1 目的

調査1では、RQ1(英検3級、準2級、2級、準1級において、解答に必要なマクロルールの傾向は異なるか)を検証する。

要約を評価する際に用いられるマクロルールの観点に基づき、設問を分類することにより、中学卒業程度から大学中級程度までを対象目安とした各級でどのような読解力が求められているのかを明らかにする。

### 3.2 方法

#### 3.2.1 マテリアル

調査1では、英検における設問を分類するため

表2: 設問概要

級	級の目安	抽出設問	語数	設問数	設問あたりの語数
3級	中学卒業程度	3C	250語程度	5	50語
準2級	高校中級程度	4B	300語程度	4	75語
2級	高校卒業程度	3C	350語程度	5	70語
準1級	大学中級程度	3C	500語程度	4	125語

#### 3.2.2 手順及び分析

調査1では、以下の手順で各級に含まれる設問を使用されているマクロルールの観点から分類を行った。2.2節で示したKintsch and van Dijk (1978) 及び van Dijk and Kintsch(1983)の3つのマクロルールに基づき、表3のような設問分類基準を作成した。

設問分類は調査者と英語教育学を専門とする大学院生がペアとなり、以下の手順で行った。まず、設問とその正答選択肢から、正答するのに必

に、調査対象級に共通して読解の最終設問として設置される科学的・社会的な内容について述べた長文及びそれに対する理解内容確認問題に焦点を当てた。具体的には、2016年第3回から2021年第2回までの間に実施された5年分、各級15回分の設問を分析した。抽出した設問及び級の目安は表2の通りである。なお英検1級にも同様の長文問題が存在するが、語数が800語程度と大幅に増え、その差が最も大きいところで3倍を超えることから今回は分析対象から除外した。

4つの級を比較すると、最も簡単な3級は250字のテキストに対して設問数が5問存在するのに対し、準1級では500字に対して設問数が4つと、一つの設問に対して読むべき語数がかなり増加していることがわかる。この観点では準2級と2級の読むべき語数が難易度の順序と合わないが、概ね3級から準1級にかけて不要情報を削除する設問よりも、情報を一般化あるいは構成することで正答できる設問数が増えていくと予測される。

要な情報が本文中のどの情報であるかを特定した。その上で、名詞など主具体的な情報のみで答えとなる場合は(a)具体・詳細情報に分類し、それ以外を(b)削除、(c)一般化、(d)構成のいずれかに分類した。なお、当てはまるマクロルールが複数あると考えられる場合は、協議によって最も近いカテゴリーに分類し、各級ごとに解答に必要なマクロルールの傾向が異なるかどうかを検証するため、カイニ乗検定を実施した。

■表3: 分類基準

分類	内容
(a) 具体・詳細情報	正答選択肢に主語と動詞が含まれておらず、テキスト本文中に示された時間や場所などの名詞や数値情報のみで解答することができる項目
(b) 削除	テキストの内容をそのままコピー、または単語レベルで言い換えを行い、必要な部分のみを取り出したり不要情報を削除したりすることで解答できる項目 例) Mary played with a ball. The ball was blue. → “Mary played with a ball.”
(c) 一般化	複数の命題を上位命題に置き換え、テキスト内容を抽象化・一般化することで解答できる項目 例) Mary played with a doll. Mary played with blocks. → “Mary played with toys.”
(d) 構成	一連の動作を包括的な上位表現に置き換えるなど、テキスト本文に明示的に示されていないアイディアを推論し、作り出すことで解答できる項目 例) I went to the station, bought a ticket, and took a train. → “I traveled by train.”

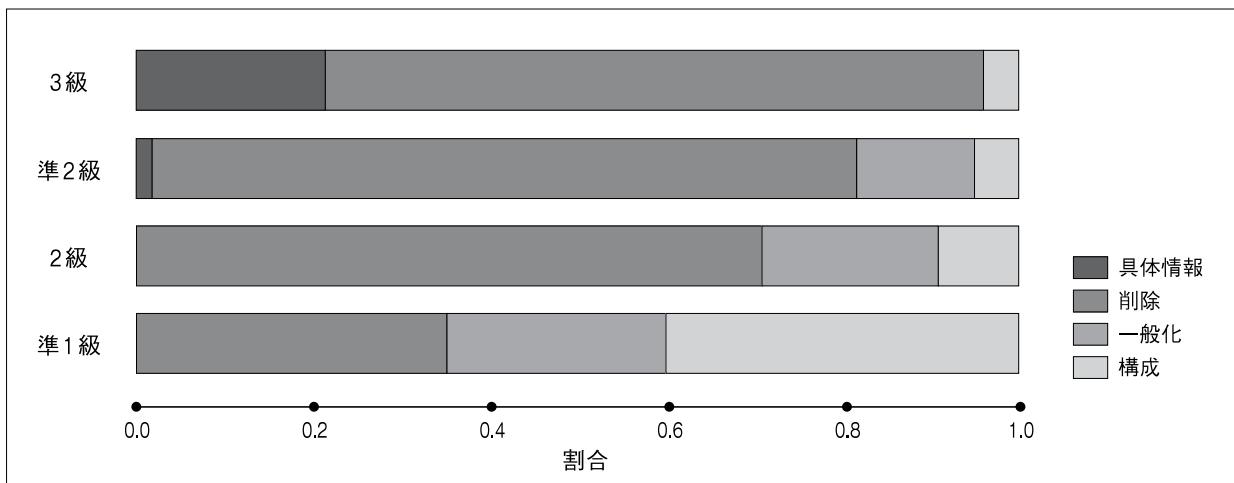
### 3.3 結果と考察

各級の30%分を分類した結果、一致率は

84.15%であった。不一致を協議によって解決したのち、残りの70%を調査者が分類した。各級ごとの設問分類結果を表4及び図2に示す。

■表4: 設問分類結果

	項目数	(a) 具体・詳細情報	(b) 削除	(c) 一般化	(d) 構成
3級	75	16(21.33%)	56(74.67%)	0(0.00%)	3(4.00%)
準2級	60	1(1.67%)	48(80.00%)	8(13.33%)	3(5.00%)
2級	75	0(0.00%)	53(70.67%)	15(20.00%)	7(9.33%)
準1級	60	0(0.00%)	21(35.00%)	15(25.00%)	24(40.00%)
合計	270	17(6.30%)	178(65.93%)	38(14.07%)	37(13.70%)



■図2: 設問分類結果

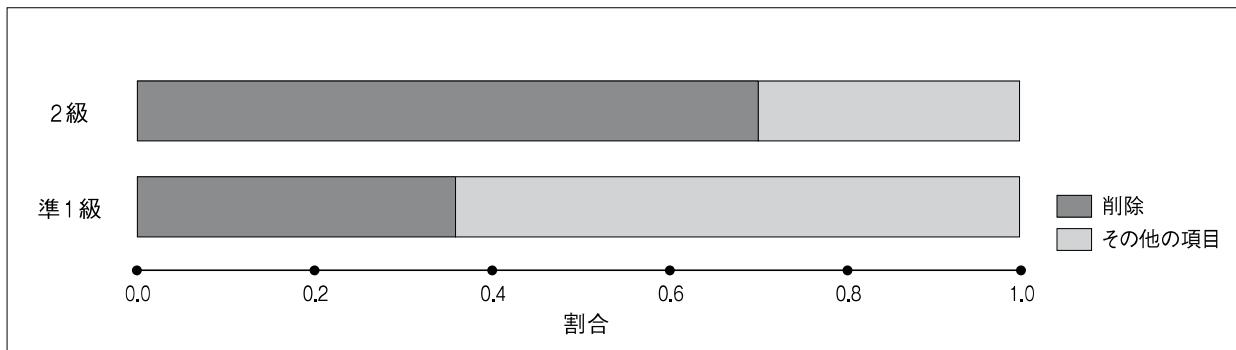
独立性の検定のため、カイ二乗分析を行ったところ、有意差が見られた( $\chi^2(9) = 106.44$ ,  $p < 0.01$ ,  $V = 0.36$ )。ポンフェローニ法を用いて多重分析を実施すると、2級と準1級間( $\chi^2(3) = \text{NaN}$ ,  $p = \text{NaN}$ ,  $V = \text{NaN}$ )を除き、3級と準2級( $\chi^2(3) = 20.44$ ,  $p < 0.01$ ,  $V = 0.16$ ), 3級と2級( $\chi^2(3) = 32.68$ ,  $p < 0.01$ ,  $V = 0.20$ ), 3級と準1級( $\chi^2(3) = 62.35$ ,  $p < 0.01$ ,  $V = 0.28$ ), 準2級と2級( $\chi^2(3) = 3.35$ ,  $p < 0.01$ ,  $V = 0.06$ ), 準2級と準1級( $\chi^2(3) = 30.03$ ,  $p < 0.01$ ,  $V = 0.19$ )の

間でそれぞれ有意差が見られた。

検証できなかった2級と準1級について、(b)削除の項目のみに焦点を当てて再分析を行った。表5及び図3で示すように(b)削除の項目数と、それ以外の(a)具体・詳細情報、(c)一般化、(d)構成の項目に分けてカイ二乗分析を実施した結果、有意差が見られ( $\chi^2(1) = 17.12$ ,  $p < 0.01$ ,  $V = 0.36$ )、2級は準1級と比べて(b)削除の項目を有意に多く含むことが示された。

■表5: 削除に基づく2級と準1級の比較

	項目数	(b) 削除	その他の項目(a, c, d)
2級	75	53(70.67%)	22(29.34%)
準1級	60	21(35.00%)	39(65.0%)
合計	135	74(54.81%)	61(45.19%)



■図3: 削除に基づく2級と準1級の比較

### 3.3.1 英検3級

英検3級では、具体・詳細情報及び削除で解答できる項目が大半を占めており、概要的な理解を問う項目はそれほど見られなかった。例として、

具体・詳細情報を問う設問を表6に示す。選択肢と本文のオーバーラップ度合いが極めて高く、itが何を指し示しているのかなど文章の結束性さえ掴めれば本文にそのまま答えが示されている

■表6: 具体・詳細情報設問例(英検3級 2021年第2回 3Cより抜粋)

Later, in 1808, Audubon moved to Kentucky. He owned a store there, but he still painted pictures of birds in his free time. Audubon's store didn't make enough money, so he closed it in 1819. He decided to travel down the Mississippi River. On the way, Audubon painted pictures of the birds that he saw. His biggest dream was to paint pictures of all the different kinds of American birds.

(28) When did Audubon close his store?

1. In 1808.
- ② In 1819.
3. In 1826.
4. In 1827.

ため、正答を導き出すことができる設問である。選択肢も語や句レベルで構成されており、英文の大意を素早く掴むスキミング能力よりも、必要な情報(ここではAudubonが店を開めた年)のみを素早く読み取るスキヤニング能力を測ることを目的としていると考えられる。

その一方で、設問数は少ないながらも英検3級の特色として、構成を含んだ設問が存在したことでも挙げられる(表7)。5年分の調査の結果、構成を含んだ設問はいずれも、英文の最終問題として設置される“What is this story about?”というものだった。

■表7: 構成設問例(英検3級 2021年 第1回 3C より抜粋)

#### A Special Meal

Thailand is a country in Southeast Asia. It is a very popular place to visit for people from other countries. About 38 million people went there in 2018. Many people like to visit Thailand because it has a lot of beautiful beaches and temples. It has some interesting festivals, too. One is called the Monkey Buffet Festival in the city of Lopburi.

The Monkey Buffet Festival is not a traditional festival. Because of this, it is different from other festivals in Thailand. Lopburi started the festival because the city wanted more visitors. During this festival, a lot of fruit and vegetables are put on tables near an old temple. However, the food is not for visitors. It is for monkeys that live near the temple.

In Thailand, monkeys are a symbol of good luck. There is a traditional story in Thailand about a prince named Rama.

(30) What is this story about?

1. The oldest city in Southeast Asia.
2. The history of Thailand.
- ③ An interesting event in Thailand.
4. A temple built by monkeys.

表7が示すように、テキスト全体を通じてタイで行われている祭りの一種について述べられているのは事実だが、トピックセンテンスなどにおいて明示的に示されてはいない。第3パラグラフではシンボルとしてのmonkeyに言及されるなど、

祭りからやや離れる内容も含まれる中で、一貫して伝えられる内容を問う設問となっており、表層的なテキストベースの理解だけでは導き出せない概要を問うた設問であるといえるだろう。

今回調査した15回分に限り、英検3級の大問3

■表8: 構成設問例(英検3級 2020年 第3回 3Cより抜粋)

#### Tom Longboat

Tom Longboat was a member of the First Nations\* and a famous long-distance\* runner. He was born in Ontario, Canada, on June 4, 1887. He grew up in a poor family on a farm, and his father died when Tom was five. When he was a child, he didn't like school. At that time, Canadian schools didn't want the First Nations people to speak their own languages.

(30) What is this story about?

- ① A famous Canadian runner.
2. A popular long-distance race in Canada.
3. A group of Canadian people in the war.
4. A special school for the First Nations people.

Cにおいて必ず最終設問(30)には上記と全く同じ問い合わせが存在した。それにもかかわらず、15題中12題は削除に分類されており、その理由は表8で示すようにトピックセンテンスなどとして本文の中で明示的に示されていたためである。

このことは、同じ問い合わせであってもテキストに明示的に提示されているか否かによって解答に必要なマクロルールが異なること、また設問の難易度も変わることを示している。

### 3.3.2 英検準2級

英検準2級では3級同様に削除で解答できる設問が最も多かった(80.00%)一方、具体・詳細情報を問う設問は極端に減っていることが大きな違いとして挙げられる。削除とは、語レベルでの言い換えや不要な部分を省くというマクロルールであり、テキスト内容と正答選択肢のオーバーラップがかなり高いことが特徴である。例えば、

表9のような場合がこれに当たる。

この設問の正答選択肢を選ぶためには、重要度の高いトピックセンテンスを含む以下の下線部2文を理解する必要がある。これらの文から重要度の高い情報を抜き出し、不要な情報を省くことで、Hing is a spice which is used in Indian food to make the flavor stronger. という答えに辿り着くことができる。この設問では Hing, spice, used, Indian, food(= cooking), stronger, flavor などが正答選択肢と本文のどちらにも含まれており、内容及びテキストベースでのオーバーラップが極めて高く、マクロルールの中では難易度の低い削除というルールが用いられていることがわかる。正答選択肢の文構造を見ても、テキスト本文と同じ Hing を主語としており、テキストの表層的な情報との違いが少なく、変形する必要もないため比較的難易度が低いと考えられる。

■表9: 削除設問例(英検準2級 2021年第2回 4Bより抜粋)

Hing is a spice which is widely used in Indian cooking. Many traditional Indian dishes are made from vegetables like potatoes and beans, and hing is added to the dishes to give them a stronger flavor. However, hing has a very bad smell until it is added to food and heated. The smell is so bad that hing must be kept inside a closed box, or everything nearby will start to smell like it.

- (34) Hing is a spice which is
1. made from potatoes and beans.
  2. kept closed to open window.
  - ③ used in Indian food to make the flavor stronger.
  4. popular in Indian because of its wonderful smell.

### 3.3.3 英検2級

英検2級は、先に述べた英検3級及び準2級と同様に削除が設問全体の半分以上を占めているという点が共通している。その一方で、対象学年が上がっていることにも由来しているのか、一般化で解答できる設問が2割、構成で解答できる設問が約1割を占めるなど、テキストベースの表層的な情報から離れ、読み手自身が概要を捉えているかどうかを問う設問が多いことが特徴的であった。一般化とは、複数の命題を上位命題に置き換えるというマクロルールであり、テキスト内で示された内容が抽象化され、上位概念に置き換えられているような場合がこれに該当する。

以下に提示した表10の例は、SmartEnCity プロジェクトというエストニアで実際に行われた事例を取り上げたテキストについての設問である。正答選択肢を導くためには、下線部の内容を理解した上で、一般化のルールを適用することが必要であると分類された。本文ではそのプロジェクトの目的がアパートによるエネルギー消費量を減らすことだけではなく、アパートそのものをより現代的で、見た目の良いものにすることにあると記されている。前者については同じくテキスト内波線部で示された内容の言い換えであり、選択肢とのオーバーラップが比較的高いのだが、後者については modern と good-looking という

命題で説明されていた部分を“improve”という動詞を用いることで短く端的に内容を抽象化し

ているといえるだろう。

■表10: 一般化設問例(英検2級 2021年 第1回 3Cより抜粋)

Estonia, however, is turning its apartment buildings into environmentally friendly homes that use energy efficiently. The project began in 2016 with 17 apartment buildings in Tartu, a city in the south of the country, and is known as SmartEnCity. Estonia is now a member of the European Union (EU), and it received financial help from the EU for the project. The aim is not only to reduce the amount of energy consumed by the apartments, but also to turn them into **modern, good-looking** homes.

(34) What is one of the goals of the SmartEnCity project?

1. To teach people in Taru how to make their homes environmentally friendly.
- ② To improve old apartment buildings so that they are more energy-efficient.
3. To provide financial help for poor people in a city in the south of Estonia.
4. To reduce the amount of energy used to build houses in the European Union.

この他にも、本文内で summer, winter という名詞で表されていたものが“different seasons”という上位概念に変わったり、テキスト内で dry, sunny places が選択肢では“places where there is not much rain”と説明されたりするなど、二つ程度で具体的に説明されていた部分を取り上げ、上位概念に置き換える設問が見られた。

### 3.3.4 英検準1級

英検準1級は、本調査の中でテキストベースから離れ、変形を必要とする概要把握問題(65.0%)が、削除などで解答できる表層的な理解を問う問題(35.0%)と比較して多かった唯一の級である。設問を4つの観点で分類した場合には2級との差が見られなかったものの、削除の項目だけに絞って再分析をすれば英検準1級の方が有意にその項目数が少ないことが示されるなど、英検準1級では表層的な理解だけで解ける設問が少ないとが明らかになった。加えて、英検準1級は他の3つの級と比べて、構成によって解答できる設問の占める割合が最も多いのも特徴的である。先に示した英検3級にも特定の問い合わせで読み手に構成のルールを用いることを求める設問が存在したが、英検準1級ではその設問文は種々様々であり、問い合わせよりも選択肢におけるテキスト本文とのオーバーラップが極めて低く、結果的に構成が必要となる傾向が見られた。

構成とは、名詞などの複数の命題を上位概念に

置き換えた一般化からさらに難易度が上がった、異なる文やパラグラフに書かれた情報を統合することが求められるルールである。これはつまりテキスト内で明示的に示された一連の動作を読み手が包括的な事実として新しく作り出すことで正答できる設問のことであり、表11は実際に出題された構成の設問例である。

このパラグラフではマフィアとレモンの歴史的な関係について述べており、時系列に沿って述べられている。あくまで事実が列挙されているだけであり、パラグラフ全体で述べられている内容が凝縮されたトピックセンテンスなどは存在せず、設問に答えるためには読み手が自らマフィアとレモン農家及びレモン市場の関係を推論しなければならない。ここでは、force farmers to accept its service という要素及びその後行われた began acting as middlemen between sellers and exporters という要素から双方を力で従わせた (control both the people who grew lemons and those who sold them overseas) という暗示的なマクロ命題を得ることができる。

英検準1級では、上記のような構成のマクロルールを用いなければ解答できない問題が最も多く、これらの設問に解答するためにただ表層的な情報を取捨選択するだけではなく、内容を理解した上で明示されない情報を推論するというより高次の能力が必要とされていると言えるだろう。

■表11：構成設問例(英検準1級 2021年 第1回 3Cより抜粋)

While the Mafia may have originally provided a legitimate service that kept lemon crops safe from thieves, that did not stay true for long. It began forcing farmers to accept its services against their will, using violence and intimidation when it encountered resistance. Mafia members then began acting as middlemen between sellers and exporters, manipulating the market to ensure substantial profits. They soon forced their way into other areas of the industry, such as transportation and wholesaling, and eventually their power grew to cover all aspects of lemon production. Some politicians attempted to address these activities, but widespread government corruption allowed the Mafia to extend its influence to many areas of politics and law enforcement.

(34) The Mafia used its involvement with lemon farmers to

- ① control both the people who grew lemons and those who sold them overseas as part of its way to make more money and increase its power.
2. successfully generate greater profits from the lemon industry by demanding that farmers grow larger amounts of lemons.
3. persuade a few politicians to ignore its criminal activities, despite overwhelming opposition from the government.
4. gain control over the entire lemon industry, despite being unable to obtain the cooperation of key members of the police.

### 3.4 調査1のまとめ

多肢選択式で行われる英検の4つの級を対象に、マクロルールに基づき設問を分類した結果、全体的な傾向として削除の割合が高いものの、各級には構成内容に有意差があり、級が上がるにつれてその割合が下がることが明らかになった。つまり、特に対象年齢が低いほどテキスト内で使用される語や語順から離れて変形することで解答できるような概要把握問題よりも、字義的なレベルでの詳細情報の理解を問う設問が多く、学習が進むにつれて難易度が上がっていく構造となつておらず、級によって正答選択肢に用いられるマクロルールの種類が異なっていることが明らかになった(RQ1)。これは予測通り、級が上がり、設問あたりに必要な語数が増えるにつれて、より高度な処理が求められるという結果である。具体的には、3級は具体・詳細情報及び削除を用いた項目が極めて多く、準2級も割合は多少下がるものとの傾向が強かった。2級では削除は約70%に留まり、一般化と構成による概要を問う設問が増加しており、準1級では概要把握を求める設問が全体の65%を占めていたというのがその難易度の高さを示している。その一方で、高難易度である構成が中学卒業程度を目安としている英検

3級で見られたことを踏まえると“What is this story about?”など、そのテキストのテーマを問うような設問を使用することで比較的容易に概要把握を問う設問を作成することができるだろう。その際、Ushiro et al.(2008)が示したように、多肢選択式テストであっても選択肢に本文内で明示的に示されていない情報が含まれている場合に設問の難易度が上がることから、削除程度の理解ではなく概要を捉える力を測りたい場合にはテキスト内から該当情報を削除する、あるいは明示されない情報を問うなど注意が必要である。

調査1の結果を踏まえ、続く調査2では、上記の分類に基づき日本人英語学習者を対象に、(a)調査者が作成した新型多肢選択式テスト(削除1問、一般化2問、構成2問)、(b)旧型多肢選択式課題(削除4問、一般化1問)、(c)要約課題の3つのテストを実際に協力者に課し、(a)及び(b)が(c)とどう関連しているのか、その信頼性及び妥当性について検証した。マクロルールを利用した設問を用いることで学習者の理解段階をより具体的に測ることができ、なおかつ知識・技能だけではなく思考力・判断力・表現力等の測定も可能となる設問作成に寄与することが期待される。

## 4 調査2

### 4.1 目的

調査2では、RQ2-1(多肢選択式で実施した新型テストと旧型テストの項目識別力及び信頼性係数に違いはあるのか)及びRQ2-2(新型多肢選択式テスト、旧型多肢選択式テスト、及び要約テストにおけるマクロルールの使用と正解確立に違いはあるのか)を検証する。調査1の結果から、英検では級が上がるにつれてより高難易度となる一般化・構成を含む設問が増えるものの、高校卒業程度とされている英検2級レベルであっても削除の割合が高く、高校の学習指導要領で求められるような概要把握を問う設問が少ないことが明らかになった。

そこで、調査2では、マクロルールの持つ階層性を利用し、これまで要約課題で測定されていた

概要把握能力を多肢選択式で測定する問題を新たに作成し、要約課題との関連性について検証する。(a)削除、一般化、構成を全て含む新型多肢選択式テスト、(b)削除で解答できる項目の多い旧型多肢選択式テスト、(c)要約課題の3つを実際に協力者に課することで、それぞれのテストの特徴及び新型テストの有用性について明らかにする。

### 4.2 方法

#### 4.2.1 参加者

日本大学生・大学院生35名を対象に調査を行った。協力者の専攻は教育学、人文学、理工学、心理学、医学と多岐にわたっていた。

#### 4.2.2 マテリアル

調査2では実際に英検2級で使用されたことのある3つのテキストをベースに(a)新型テスト、(b)旧型テスト、(c)要約課題を実施した。表12に各マテリアルの概要を示す。

表12: マテリアル概要

テキスト	本文総語数	FRE	FKGL	CEFRレベル
Ocean Treasure (2021年 第2回 3C)	360	61.8	8.3	C1
Building a Better Future (2021年 第1回 3C)	360	56.9	9.3	C1
Adult Brains and Reading (2019年 第3回 3C)	352	63.1	9.6	C1

注)FRE = Flesch Reading Ease; FKGL = Flesch-Kincaid Grade Level

#### 4.2.3 手順

調査は伝染病の流行状況を鑑み、全てオンラインで実施された。協力者は調査者とメールにてやり取りを行い、辞書を使用しないこと、翻訳アプリ等を使用しないこと、各課題20分以内に解き切ることなどの注意事項を確認した上で課題に取り組んでおり、不明点はいつでも確認できる状況にあった。協力者はPDFにて3つの英文が提示され、それに付随して多肢選択式のテストが2つ、要約課題が1つあることを事前に予告されていた。なお、テキスト順序及び課題の順序についてはカウンターバランスが取られており、協力

者は問題を上から順番に解くように指示された。協力者は長文及び設問を印刷して取り組むこと、読解中にメモを取ったりマーカーを引いたりすることが許可されていた。また、課題遂行中はいつでもテキストに読み戻ることが可能だった。要約課題については日本語で250字以内でまとめるように指示されたが、これは第二言語を用いた場合学習者のライティング能力によっては産出に困難を抱えることが想定され、母語を用いた方が学習者の理解が反映されやすいためである(Yu, 2008)。

#### 4.2.4 採点と分析

協力者によって産出された要約プロトコルは1文ずつ採点された。調査1の設問分類基準と同様に文の中で使用されているマクロルールの観

点に基づいており、その基準は表13のとおりである。なお、1文の中に複数のマクロルールが混在する場合は、すべて個別にカウントされた。

■表13: 要約評価基準

(b) 削除	テキストの内容をそのままコピー、または単語レベルで言い換えを行い、必要な部分のみを取り出したり不要情報を削除したりする
(c) 一般化	複数の命題を上位命題に置き換え、テキスト内容を抽象化・一般化する
(d) 構成	一連の動作を包括的な上位表現に置き換えるなど、テキスト本文に明示的に示されていないアイディアを推論し、作り出す

採点については英語教育を専攻する大学院生と調査者がペアになり、それぞれ全プロトコルのうちの30%を分類し、一致率は89.60%であった。不一致箇所は協議によって解決され、残りのデータは調査者1名で採点を行った。

その上で、RQ2-1に回答するため、古典的テスト理論に基づいて新型テストおよび旧型テストの項目識別力と信頼性係数を算出した。また、それぞれのテストの構成概念妥当性を検証するために探索的因子分析を実施した。

また、RQ2-2に回答するため、要約テストに対してはポアソン回帰に変量効果を加味した一般化線形混合効果モデリングを実施した。要約テストと新型および旧型テストの比較においては目的変数を2値データに変換し、ロジスティック回帰に変量効果を加味した一般化線形混合効果モデリングを実施した。

#### 4.3 RQ1の結果と考察

表14は新型テストおよび旧型テストの正解数と不正解数のクロス集計表である。なお、旧型テストにおいては構成が存在していなかったため空欄となっている。全体的な傾向として、削除に関する問題は新旧のテストで約9割の正解率を示した。しかし一般化の問題の正解率に関しては、新型テストで88%、旧型テストで77%となっており、約10%の違いが見られた。新型テストの構成に関しては、正解率が80%と他の問題と比べて最も低かった。

■表14: 2つのテストにおける正解と不正解の割合

		新型テスト		旧型テスト	
		n	%	n	%
削除	正解	33	94.3	128	91.0
	不正解	2	5.7	12	9.0
一般化	正解	62	88.6	27	77.0
	不正解	8	11.4	8	23.0
構成	正解	56	80.0	-	-
	不正解	14	20.0	-	-

表15は要約テストにおける、3つのマクロルールの発生頻度とその割合を表す。要約プロトコルから得られた総プロトコル数は200であった。この表からはdeleteが協力者全ての要約課題で

生じていた一方、他のマクロルールの使用率を見ると一般化で約3割、構成で約5割と使用率が下がっていることがわかる。

■表15: 要約課題におけるマクロルールの発生頻度

	要約テスト	
	n	%
削除	164	100
一般化	13	31.4
構成	23	48.6

上記のマクロルールの使用率及び正答率を見ると、削除については先行研究と同様、要約課題でも使用率が他二つのルールと比べて有意に高く、学習者の間で最もよく用いられる産出方略であることが示された(邑本, 1992; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983; Segev-Miller, 2007; Ushiro et al; 2008)。特にBrown and Day(1983)は削除のようなコピー & ペーストによって要約を産出するのは幼い子どもによく見られる傾向であるとしていたが、本研究では大学生程度であってもその傾向が見られた。加えて、EFL環境下ではテキストから離れることが難しいことから、ソーステキストへの依存度合いが高いことが示されており、先行研究と同様の結果となった(Weijen et al., 2019)。

続いて、テストの妥当性検証のため、新型テストにおける項目識別力を点双列相関分析(テストの総合点と項目に正解したかどうかの2値データの相関)を実施した。その結果は表16の通りであ

る。項目識別力とは、測定しようとしている特性が、その能力値の高いテスト受験者とそうでない受験者をどの程度区別できるかを示す。具体的には、識別が大きい項目ではテスト得点が低い学習者はその項目に誤答し、テスト得点が高い学習者はその項目に正答するという傾向のことである。つまり、この値が高いほど受験者の能力を妥当に測定が可能となり、表16の結果、構成に関する項目1は識別力が.226と低く、受験者の能力を識別できていなかった。一方、一般化に関する項目はいずれも.65を超えており、今回の受験者の弁別に寄与していた。

表17は旧型テストにおける項目識別力である。新型テストと同様に点双列相関分析を実施した結果は、新型テストと異なり、一般化に関する項目の識別力が低かった。このことから、新型テストの一般化問題は旧型テストよりも受験者の弁別に寄与していたと言える。

■表16: 新型テストの項目識別力

	構成(1問目)	構成(2問目)	一般化(1問目)	一般化(2問目)	削除
テスト得点	.226	.503	.653	.674	.503

■表17: 旧型テストの項目識別力

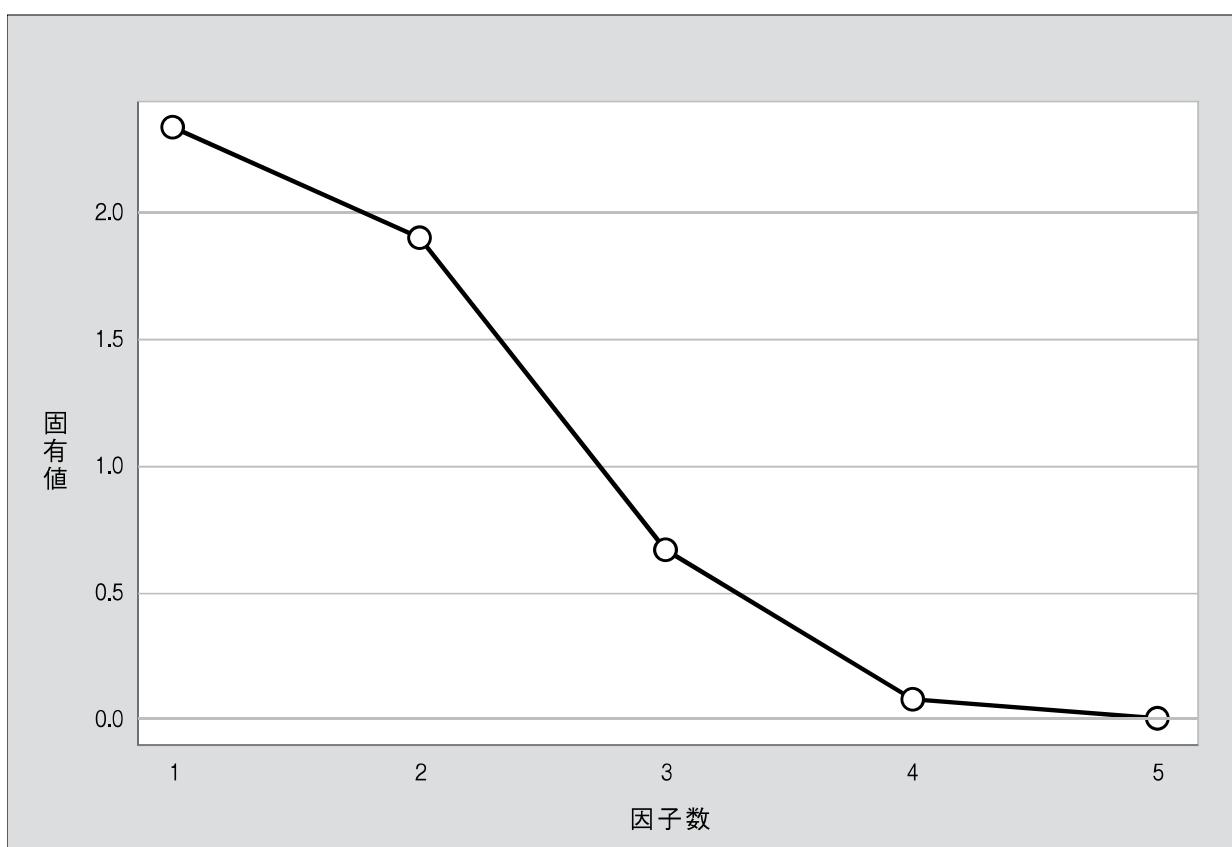
	一般化	削除(1問目)	削除(2問目)	削除(3問目)	削除(4問目)
テスト得点	.339	.615	.487	.339	.511

なお、一般化が新型テストで項目識別力が安定しているのに対し、構成の2項目が識別力の点でばらつきが見られるのは、項目の依存性及び受験者の依存性によるものであると考えられる。これについては旧型テストで削除の数値にばらつきが見られるのと同様である。この数値が低い場合、熟達度の低い学習者だけではなく、熟達度が高く本来設問を正答できる学習者であっても正答できないなど、弁別力が低いことを意味する。古典的テスト理論では、この識別力はテストを受ける受験者そのものが持つ能力や発達段階、年齢などの特性(受験者の依存性)、そしてテストの設問そのものの特性である語彙や文法などの難しさによる項目の依存性が存在するとされている。これらの依存性を解消するため、協力者集団の熟達度やその他の条件が正規分布するよう母集団を増やし、さらに項目数を増やして実験を行うことが必要だろう。

また、テストの信頼性についてはクロンバッックのアルファを用いて算出した。テストの信頼性はテスト得点の安定性と直結する指標であり、こ

の値が大きいほど同じ能力を測定していると捉えることができる。今回、新型テストにおける信頼性係数は  $\alpha = .34$ 、旧型テストは  $\alpha = -.039$  といずれのテストも項目数が非常に少なかったため小さな値を取っているが、新型テストの方が旧型テストよりも高い信頼性を示した。

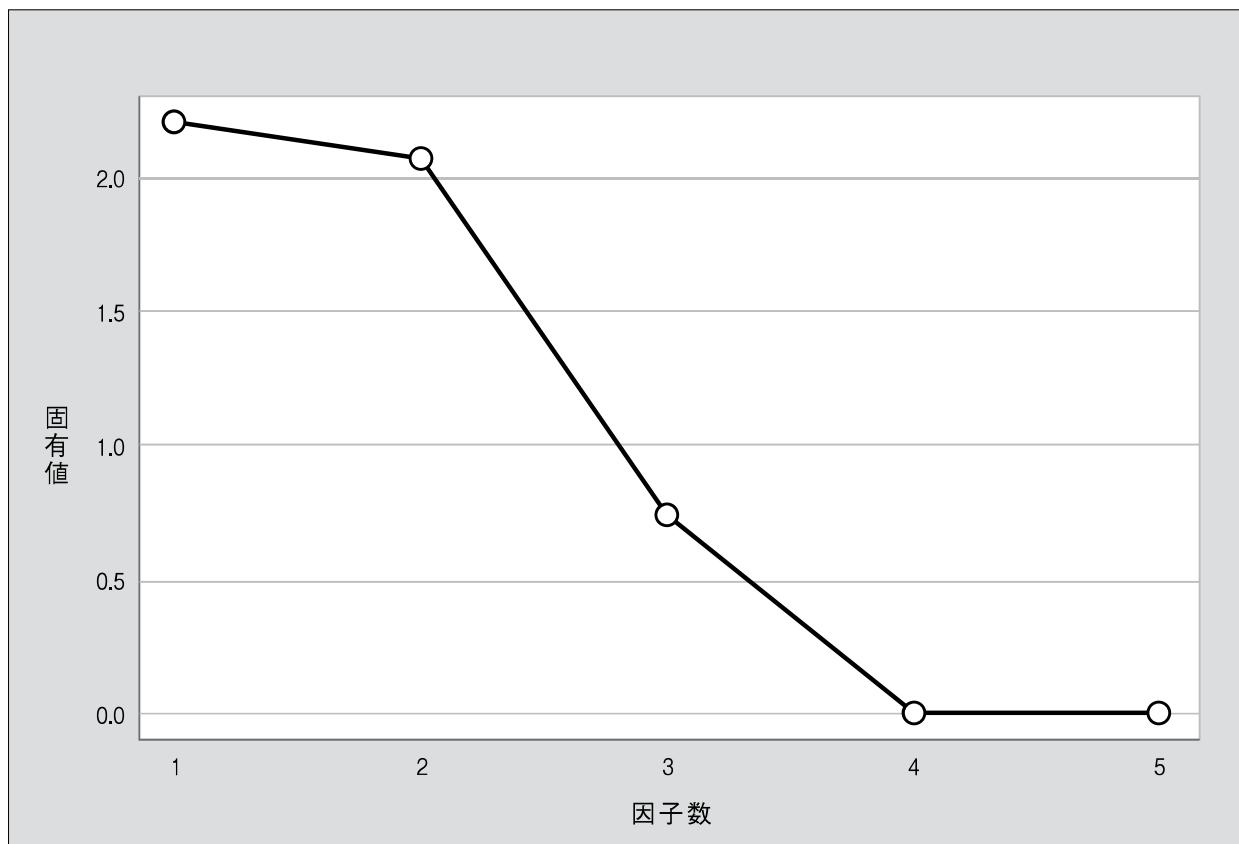
さらに、新型テストで測定している能力はいくつあるかを確認するため、探索的因子分析を行った。新型テストでは削除、一般化、構成の3つの能力を仮定していることから、3因子が検出されると予想された。しかし、協力者数および項目数が少なかったことから、検出されたのは2因子であった(図4)。このことから、今回の設問に回答する際に必要であったマクロルールは3つではなく2つだった可能性がある。このことについては、マクロルールの持つ類似性という観点からRQ2でも引き続き取り扱っている。



■図4：新型テストにおけるスクリープロット

図5は旧型テストに対して行った因子分析の結果である。想定通り、一般化と削除の2つの因子

が検出されたが、こちらもサンプル数を増やした調査が必要となる。



■図5：旧型テストにおけるスクリープロット

#### 4.4 RQ2の結果と考察

RQ2では、RQ1の結果に基づき、信頼性が高いと判断された新型テストが要約課題と代替可能かどうかを検証するため、2つのテストの比較検

証を行った。表18は要約テストの分析結果であり、目的変数を各マクロルールの生起頻度、固定効果をマクロルールの種類、テキストと協力者を変量効果とした一般化線形混合効果モデリングの結果である。

■表18: 要約テストに関する一般化線形混合効果モデリングの結果

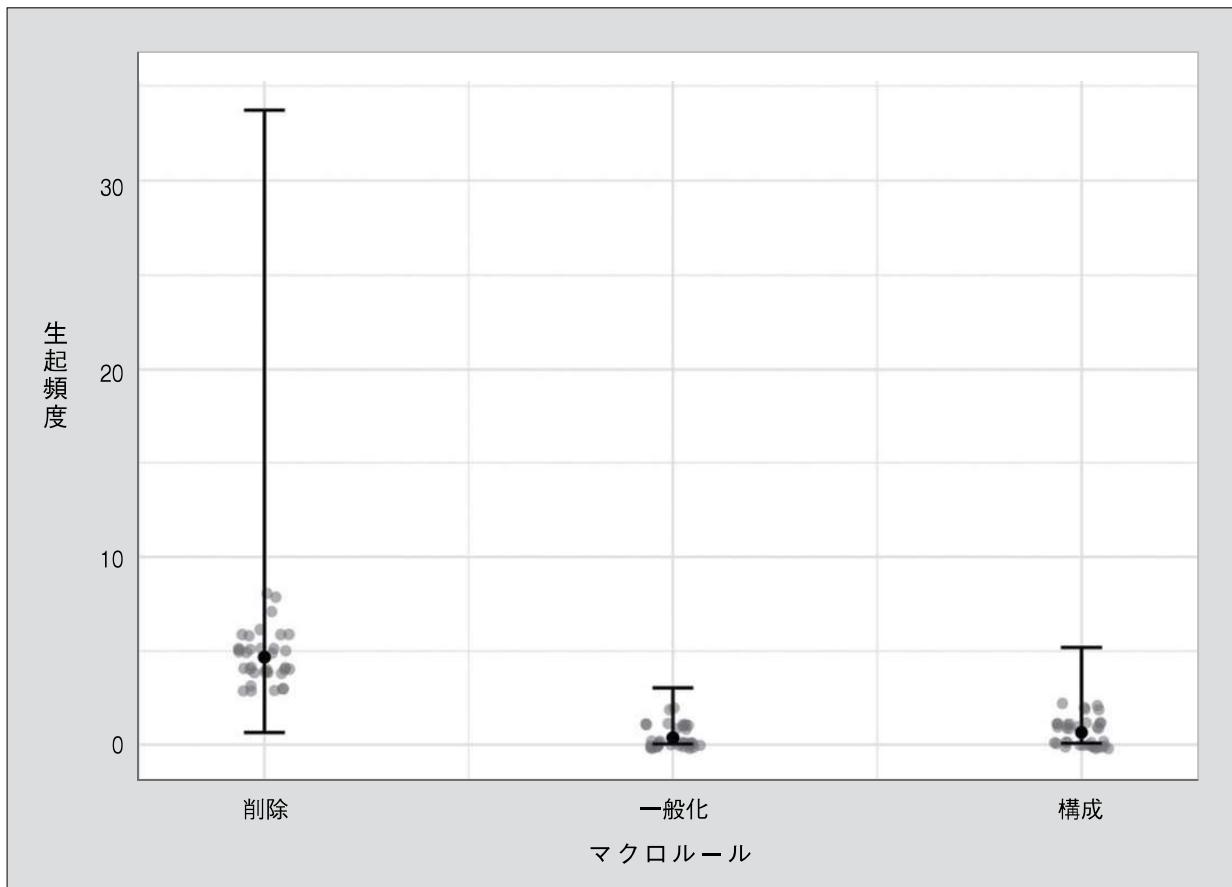
パラメタ	固定効果				変量効果	
	$\beta$	$SE$	$z$	$p$	協力者	テキスト
					$SD$	$SD$
切片	0.044	0.119	0.378	.706	0	0
削除-一般化	-2.535	0.288	-8.798	<.001***	-	-
一般化-構成	0.571	0.347	1.644	.100	-	-

分析の結果、削除は一般化よりも有意に頻度が高かった。このことから、協力者は要約の際に一般化よりも削除を有意に多く使用する傾向があったと言える。図6は分析結果に基づく予測であり、

削除がマクロルールの中で最も高頻度であった。図6に示されている通り、一般化と構成は生起頻度が少なく、統計的な有意差は見られなかった( $p = .100$ )。

なお、要約課題で上位の処理が少なかった理由として、要約産出方略に関する知識が不足していたために協力者が規則を使うことができなかつた可能性も存在するが、要約課題における語数を少なくすることで解決される可能性も存在する(Ushiro et al., 2009)。この研究では要約課題の

字数をより制限することで、一般化や構成の使用が見られることが報告されている。このことから、本研究のように多肢選択式と要約課題を比較する場合、要約課題で指定する語数について、さらに慎重に検討することで異なる傾向が見られる可能性があるだろう。



■図6：予測プロット

要約課題においてより高次の規則の発生頻度が低いことが本調査の結果明らかになったが、要約課題においてマクロルールが使用されていなければ、学習者は一般化や構成に関する能力がないと評価するのは誤りである。新型の多肢選択式テストでは一般化、構成のどちらにおいても80%を超えていていることを踏まえ、新型テストで測定された3つのマクロルールと実際の要約でマクロルールを協力者が用いたかどうかの関係性について、以下のような一般化線形混合効果モデリングを実施した(表19)。目的変数は、新型テストにおけるそれぞれのマクロルールの設問に正解したかどうか2値データと要約テストにおけるそれぞれのマクロルールが発生したかどうかの2

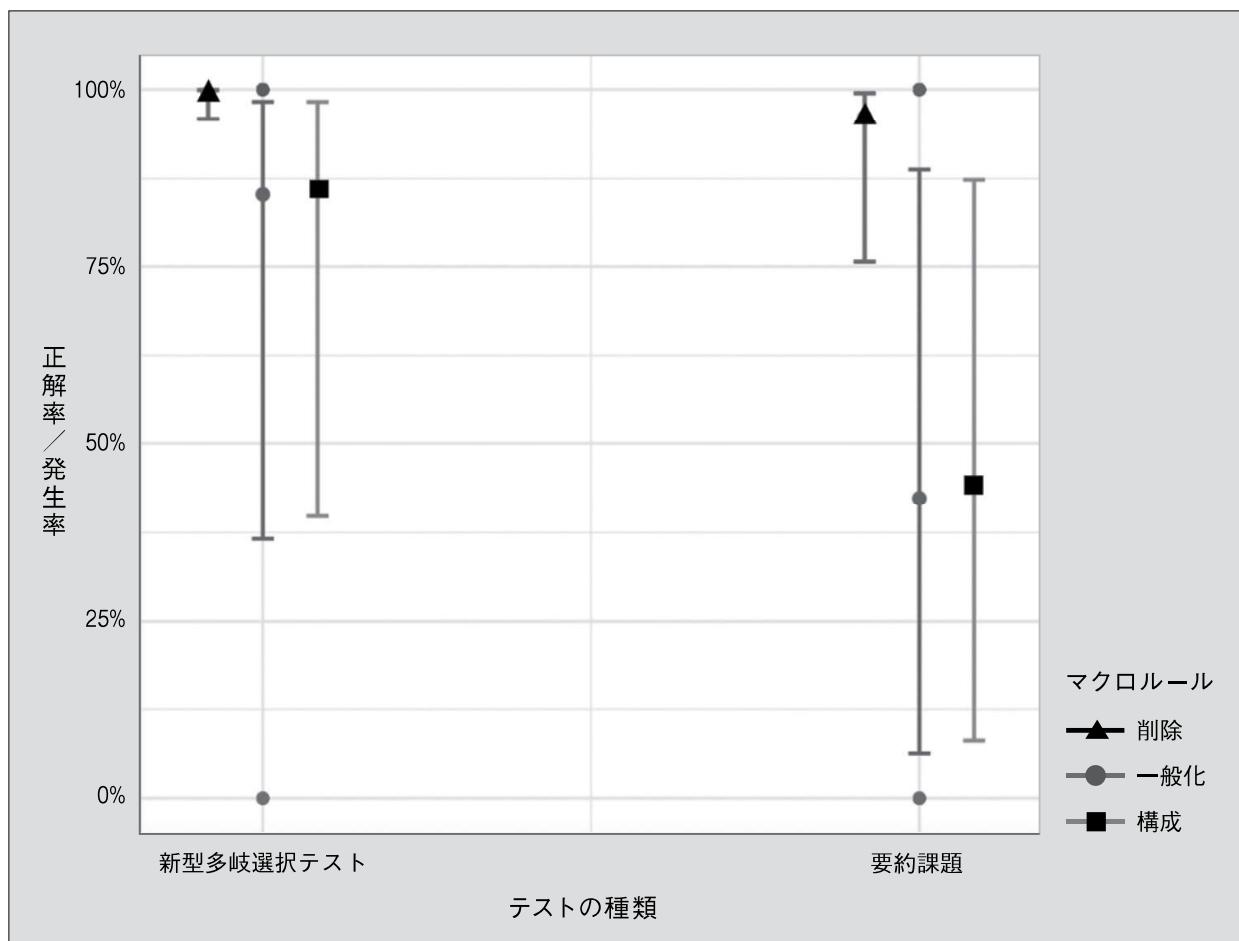
値データである。固定効果はテストの種類と3種類のマクロルールであり、協力者とテキストを変量効果とした。その結果、新型テストよりも要約課題において正解確率が有意に低かった。このことから、新型テストは要約課題よりも一般化と構成を正確に測定していると言える。また、図7に示される予測プロットの通り、構成や一般化の正解率および発生確率は削除に比べて有意に低く( $p < .001$ )、構成と一般化の間には有意な差は見られなかった( $p = .853$ )。スクリープロットにおいても3つのマクロルールを応用した新型テストで2つの因子のみが検出されたことから、構成と一般化は類似した構成概念であるように思われる。マクロルールにおいて削除、一般化、構

成の順で難しくなる (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983) というのは元々のテキストと要約で産出される文章とのオーバーラップの観点からも支持されているが、本調査では削

除よりも一般化、構成の難易度が高いことは示されたものの、一般化と構成の関係についてはさらなる検証が必要であると言えるだろう。

■表19: 要約テストと新型テストの関係に関する一般化線形混合効果モデリング

パラメタ	固定効果				変量効果	
	$\beta$	$SE$	$z$	$p$	協力者	テキスト
					$SD$	$SD$
切片	2.940	0.407	7.232	<.001***	.778	0
テスト	-2.057	0.373	-5.515	<.001***	-	-
削除-一般化	-3.510	0.807	-4.352	<.001***	-	-
一般化-構成	0.064	0.347	0.186	.853	-	-



■図7: 予測プロット

#### 4.5 調査2のまとめ

調査2では、実際に協力者にマクロルールの考え方を応用した(a)新型多肢選択式テスト、削除

レベルの設問が多い(b)旧型多肢選択式テスト、(c)要約課題の3つを課すことで新旧の多肢選択式テストの比較及び要約課題との関連性を検証した。その結果、得られた結果は次の2点である。

## (1) 要約課題と多肢選択式課題に見られるマクロルール使用能力の差

RQ2-1において、実際に設問に回答した協力者のデータをもとに新型テスト及び旧型テストの項目識別力及び信頼性の比較を行った結果、母数が少ないため数値は低いものの、新型テストの方が旧型テストよりもより高い信頼性を示した。特に新型テストの一般化については識別力が高く、多肢選択式テストで学習者の理解を測定する手法として、マクロルール利用の有用性を押し出す結果となったと言える。また、要約課題だけではなく、多肢選択式を用いた新旧どちらのテストにおいても削除の項目は正答率が90%を超えていたことから、学習者にとって比較的実施しやすく、易しいものであったことが推察される。

その一方で、要約課題では一般化と構成の使用率が50%を下回るなど、一見学習者はテキストの表層的な部分の理解しかできていないかに思われたが、新型テストであれば正答率は70%から80%と高い傾向にあったことを踏まえると、要約課題のみから学習者の理解度合いを測定することの危険性が示された。

## (2) マクロルールにおける一般化と構成の類似性

本調査において、「概要を捉える力」をマクロルールの構成及び一般化と定義して実験を行ってきた。削除、一般化、構成の3つのルールを含む新型のテストであってもスクリープロットにおいて2つの因子しか検出されなかったこと、加えて要約課題と新型テストにおいて、一般化と構成の発生率及び正解率の予測は削除に比べて有意に低いのに対し、一般化と構成の間には有意差がなかったことから、この二つの規則の類似性が示された。新型テストと旧型テストの信頼性の観点から、削除が設問の8割を超える従来よく用いられていたテストから、一般化あるいは構成に基づく設問を増やすことでこの信頼性を高められる可能性が示唆されているといえる。なお、新型テストにおける正答率で比較すると、新型テストの一般化は88%であるのに対し、構成は80%とやや低い値となっている。協力者の数を増やすことで、削除とは明確に異なる一般化、

構成というマクロルールが幅広い熟達度を持つ学習者の弁別に寄与するのか、さらなる検証が求められる。

## 5 結論と今後の課題

本研究では、日本人英語学習者の詳細情報から概要把握能力までを段階的に測ることができる多肢選択式テスト作成を目的として、情報の取捨選択・統合のプロセスであるマクロルールの持つ階層性を踏まえた設問と、従来の多肢選択式設問、そして要約課題を2つの調査を通じて比較・検証した。本研究から得られた主要な発見を以下にまとめると。

調査1では、英検の3級、準2級、2級、準1級という4つの異なる熟達度をもつ学習者を弁別するために設置された大規模テストの社会的な話題について述べられた長文課題を用い、学習者の理解がどのように測定されているのかを調べるために、マクロルールに基づく設問分類を行った。設問分類の結果から、級によって解答に必要なマクロルールが異なることが示され(RQ1)、「概要把握能力」として定義した一般化及び構成が設問の半分を超えたのは準1級のみであり、高校卒業程度とされる英検2級を含むそれ以外の3つの級では概要理解を測る設問は全体の3割に満たなかつた。その一方で、最も簡単であるとされていた英検3級において、構成設問が存在するなど、テキスト内に解答が明示的に示されていないという条件のもとで特定の尋ね方をすれば、概要を問う設問が比較的平易に実施できることが明らかになった。

続く調査2では、調査1で分類したマクロルール3種類に基づいて新たに5問4択式の多肢選択式テストを作成し、従来型として英検2級で最も多かった削除を4問、次いで多かった構成を1問含んだ旧型多肢選択式テスト、要約課題と比較することで新型テストの信頼性、妥当性及び要約との関係性を検証した。その結果、新型と旧型の多肢選択式テストを比較すると、母数が少ないため数値は低いものの新型テストの方がより高い信頼性を示した(RQ2-1)。これについては、学習者の弁

別をするためには削除だけを用いるのでは不十分であり、段階的に難易度を上げるために一般化や構成を用いた選択肢の作成が有効であることを示している。さらに、新型テストが要約課題に代替可能かどうかを調べるため、要約課題における各マクロルールの使用率と新型テストにおける各マクロルールの正解確率を比較したところ、要約課題ではほとんど概要把握を反映した一般化や構成を用いることができなかった学習者も、多肢選択式では正答できるなど、要約課題のみから学習者のテキスト理解を測る危険性が示された(RQ2-2)。

本研究の結果から、学習指導要領が謳う「思考力・判断力・表現力等」を測定する設問として、マクロルールを用いた多肢選択式設問の有用性が示された。これまでに広く使用してきた要約課題には、多肢選択式では補えない発信語彙の増強やアウトプットの機会といった要素が存在することは確かである。しかしながら、こと概要把握能力を検証する手段としては、学習者の要約産出方略に大きく影響を受けてしまうこと、採点時の信頼性や時間効率に課題が残ること、また語数を慎重に決定しなければ学習者が行ったはずの統合処理が見られず、複写や部分的な削除にとどまってしまう可能性を拭い去ることはできない。そこで、学習者の理解を測る手立てとして、特に

一般化や構成を取り入れた設問を使用することで、ペーパーテストでありながら学習者の下位的な処理だけでは正答できない設問を作成することが可能となるだろう。

最後に、本研究に残る限界点として、要約課題採点に伴う問題で協力者の人数をこれ以上増やすことができなかつたことが挙げられる。協力者の数を増やし、各種マクロルールの正答率や要約に見られる発生頻度をより詳細に検証することで、一般化と構成の違いや、マクロルールの使用に関する新たな知見が得られる可能性が存在する。受験者の特性をより一般化した調査を行うことによって、より弁別力に優れた、学習者の詳細情報理解から概要把握までを問えるテストの作成に寄与することが可能となるだろう。

## 謝 辞

本研究を実施する機会を与えてくださった公益財団法人 日本英語検定協会の皆様と選考委員の先生方、寺内一先生に心より感謝申し上げます。また、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なご指導をいただきました。最後に、設問分類・採点作業や調査にご協力いただきました皆様に厚く御礼申し上げます。

## 参考文献(\*は引用文献)

- \* Brown, A.L., & Day, J.D.(1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22, 1-14.
- \* Horiba, Y.(2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, 29(3), 223-267. [https://doi.org/10.1207/S15326950dp2903\\_3](https://doi.org/10.1207/S15326950dp2903_3)
- \* Keck, C.(2010). How do university students attempt to avoid plagiarism? A grammatical analysis of undergraduate paraphrasing strategies. *Writing & Pedagogy*, 2, 193-222. <https://doi.org/10.1558/wap.v2i2.193>
- \* Kintsch, W.(1988). The role of knowledge in discourse comprehension: a constructionintegration model. *Psychological Review*, 95(2), 163-182.
- \* Kintsch, W., & van Dijk, T. A.(1978). Cognitive psychology and discourse: Recalling and summarizing stories. In W. U. Dressler(Ed.), *Current trends in textlinguistics*. New York: de Gruyter.
- \* Kobayashi, M.(2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19, 193-220. doi:10.1191/0265532202lt227oa
- \* Mateos, M., Martin, E., Villalon, R., & Luna, M.(2008). Reading and writing to learn in secondary education: online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing*, 21, 675-697. <https://doi.org/10.1007/s11145-007-9086-6>
- \* Ministry of Education, Culture, Sports, Science and Technology.(2018). Chugakko gakushu shidou yoryo kaisetsu: Gaikokugo hen, eigo hen [The commentary to the curriculum guideline for lower-secondary school: Foreign language, English]. [https://www.mext.go.jp/component/a\\_menu/education/micro\\_detail/\\_icsFiles/afieldfile/2019/03/18/1387018\\_010.pdf](https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/18/1387018_010.pdf)
- \* 邑本俊亮(1992).「要約文章の多様性—要約産出方略と要約文章の良さについての検討—」『教育心理学研究』40(2), 213-223.
- \* Scoular, K.(1998). The influence of assessment method

## 参考文献 (\*は引用文献)

- on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472. doi:10.1023/A:1003196224280
- \* Segev-Miller, R.(2007). Cognitive processes in discourse synthesis: the case of intertextual processing strategies. In M. Torrance, L. Van Waes, & D. Galbraith(Eds.), *Writing and cognition*(pp. 231-250). Bingley: Emerald Group.
- \* Sherrard, C.(1989). Teaching students to summarize: Applying textlinguistics. *System*, 17,1-11.
- \* Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7-26. https://doi.org/10.1598/RRQ.24.1.1
- \* Ushiro, Y., Kai, A., Nakagawa, C., Watanabe, F., Hoshino, Y., & Shimizu, H.(2009). Effects of reading perspective on EFL learners' summary writing and importance rating. *Annual Review of English Language Education in Japan(ARELE)*, 20, 11-20.
- \* Ushiro, Y., Nakagawa, C., Kai, A., Watanabe, F., & Shimizu, H. (2008). Construction of a macroproposition from supporting details: investigation from Japanese EFL readers' summary and importance rating, *The Japan Association of college English Teachers(JACET)*, 47, 111-125.
- \* van Dijk, T.A., & Kintsch, W.(1983). *Strategies of discourse comprehension*. New York: Academic Press.
- \* Weigle, S. C., & Parker, K.(2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21, 118-133. https://doi.org/10.1016/j.jslw.2012.03.004
- \* Weigle, S. C., Yang, W., & Montee, M.(2013). Exploring reading processes in an academic reading test using short-answer questions. *Language Assessment Quarterly*, 10, 28-48. https://doi.org/10.1080/15434303.2012.750660
- \* Weijen, D., Rijlaarsdam, G., & van den Bergh, H.(2019). Source use and argumentation behavior in L1 and L2 writing: a within-writer comparison. *Reading and Writing*, 32, 1635-1655. https://doi.org/10.1007/s11145-018-9842-9
- \* Yang, H. C., & Plakans, L.(2012). Second language writers strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46, 80-103. https://doi.org/10.1002/tesq.6
- \* Yang, H. C. (2014). Toward a Model of Strategies and Summary Writing Performance. *Language Assessment Quarterly*, 11(4), 403-431. https://doi.org/10.1080/15434303.2014.957381
- \* Yu, G.(2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing*, 25, 521-551.