

弁別力の高い推論質問の解明: 読解テストへの RI-Val モデルの応用

研究者: 茨城県/筑波大学大学院 在籍 西 聖

《研究助言者: 竹内 理》

概要

本研究では、RI-Valモデル(Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)の受動的な活性化プロセスの仮定に基づいて推論質問を9つに分類し、国内大規模英語読解テストにおける推論質問の出題状況と弁別力の高いテスト作成に貢献する推論質問のタイプを検証した。

調査1では、英検、TOEFL、センター試験の読解セクションにおける推論質問の出題状況を分析した。結果、英検では推論質問の出題がほとんど見られなかった一方で、TOEFLではアカデミックな文章を読む際に必要な単語統合に関する推論が、センター試験では概要や要点の把握、また書き手の意図を読み取るのに必要な因果的先行詞や感情に関する推論が有意に多く問われていることが明らかになった。

続く調査2では、9つの推論質問の中から推論テストの成績の高い/低い受験者を弁別することができる推論質問のタイプを解明した。結果より、照応参照は安定して推論能力の有無を弁別できること、また因果的先行詞は質問タイプ自体の難易度が高くないため推論能力の有無の弁別には適していないことが示された。

1 はじめに

令和4年度から完全実施される高等学校における新学習指導要領の「読むこと」においてはコミュ

ニケーションを行う目的や場面、状況などに応じて、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できる力を養うことが示されている(文部科学省, 2019)。つまり、英文読解においてはテキストレベルの情報を理解するだけでなく、テキストに書かれている情報から読み手の背景知識を活用してそのテキストの概要や要点、また書き手の意図まで把握することが求められている。さらに言い換えれば、読み手にはテキストに記述されている情報を越えた、より深い理解が求められていると言える。

こうしたテキストのより深い理解を達成するのに必要不可欠な処理が推論である。推論とはテキスト情報に自分の背景知識を合わせて心の中にテキストのイメージをつくることと定義され、そのはたらきはテキスト情報に自分の知識を合わせて本文中に書かれていない情報を導くことであるとされている(McNamara & Kendeou, 2011)。この推論は複数の読解モデルにおいても共通に仮定される処理であり、読み手は推論を行いながら読みを進めることが明らかになっている(川崎, 2014)。しかしながら、読み手は常に正しい推論を導くことができるわけではない。RI-Valモデル(Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)の仮定によると読み手は文を読んだ際に推論を生成し、符号化している文との繋がりを構築する。しかしながら、この処理過程は受動的に行われるため、文脈上課ったコネクションが作られてしまい、正しい推論を導けな

い可能性があることが明らかになっている。そのため、推論は高次な読解技能の一つと考えられ、優れた読み手を特徴づける技能であると言われる(吉田, 2001)。

このように、推論はより深いテキスト理解を達成するために必要不可欠な、また高次の読解処理である。そのため、読解テストにおいて読み手の推論について問うことは必須であり、そうした推論質問は優れた読み手とそうでない読み手を弁別することができる可能性を秘めている。

しかしながら、推論質問に関する研究は事実確認質問などと比べて少なく、実際に大規模英語テストにおいてどのような推論質問が出題されているのか、またそれぞれの推論質問は読み手の読解能力を適切に弁別できているのかどうかについては明らかになっていない。そこで、本研究ではRI-Valモデル(Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)の受動的で自己制御不能な活性化プロセスの仮定に基づいて、読み手の中で受動的に活性化、統合、確認されていると考えられる推論質問をCook and O'Brien (2017)を参考に9つに分類した。そして、その中でどのようなタイプの推論質問が出題されやすいのか、またその中でどのようなタイプの推論質問が弁別力の高いテスト作成に貢献するのかを2つの調査を通して検証する。

2 先行研究

2.1 読解と理解

読解とは、情報を獲得するための最も有効な手段であると言われており、特に教育の場面では、読解理解の成功が特に重要な役割を果たす。

読解における理解とは一貫した心的表象を構築することである(Kintsch, 1998)。心的表象とは読み手の心内に保存されるテキスト内容のことを指し、テキストの字義レベルの理解から読み手の背景知識と統合されたより深いレベルの理解まで3段階に分けることができる:(a) 表層的記憶(the surface memory), (b) 命題的テキストベース(the propositional textbase), (c)

状況モデル(the situation model; van Dijk & Kintsch, 1983)。この中で状況モデルはテキスト情報と読み手の背景知識を合わせたテキスト全体の理解を指し、一般的に最も深い理解であると言われる。

2.2 英文読解におけるプロセス

読み手の読解中のプロセスを明らかにするためこれまでに多くの読解モデルが提唱されている。先述のように多くの読解モデルでは、文章を理解する際に読み手の心内に心的表象が形成されると仮定する。しかしながら、それぞれの読解モデル間ではその仮定に大きな違いも存在する。本節では本研究で扱うRI-Valモデルを中心に3つの読解モデルについて取り上げ、英文読解におけるプロセスについて概略する。

Kintsch (1988, 1998) による構成・統合(Construction-Integration; CI)モデルは、現在までもその仮定が大きくは覆されておらず、多くの読解モデルの基礎となっている。このモデルは単一テキストだけでなく、複数テキストの読解モデルにもその影響を与えており、その影響は計り知れない。このモデルでは、読解の基本プロセスを構成と統合の2つに分割している。文章を読むとその読んだ文に関連するテキスト中の情報や関連知識の情報が文章や背景知識から再活性化され(構成)、ネットワークを通じて活性化が確定すると概念の活性化が安定する(統合)。この構成と統合のプロセスは、自動的で自己制御不能であるとされており、活性化の段階においては文章に無関係な知識も活性化されることが知られている。

O'BrienやMyersらはKintschのCIモデルを発展させ、記憶ベース・モデルを提唱した(Albrecht & Myers, 1995; Albrecht & O'Brien, 1993; Guéraud, Harmon & Peracchi, 2005)。このモデルでは焦点となる文とは離れた位置に存在する情報がどのように再活性化されるのかを説明するために、共鳴過程(Resonance)を導入した。読み手が文を読むと、ワーキングメモリ内の情報から長期記憶内の全情報にシグナルが送られ、読み手が現在読んでいない情報や背景知識に蓄積されている情報が再活性化される。こ

の再活性化が安定すると、全情報の中で最も活性化された要素がワーキングメモリに加えられる。共鳴過程は高速で受動的な処理であるとされており、共鳴過程においては文章に無関係な知識も活性化される。これは構成・統合モデルの構成過程に類似している。

Cook や O'Brien らは記憶ベース・モデルをさらに発展させ、RI-Val モデルと呼ばれる新たな読解モデルを提唱した (Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)。このモデルでは文章理解のプロセスに確認処理という第3のステージを仮定し、読解を共鳴 (Resonance)、統合 (Integration)、確認 (Validation) の3つのオーバーラップするステージに分割した。文を読むと、まず読んだ情報に関連する全ての情報が長期記憶から再活性化される (共鳴)。これは記憶ベース・モデルで提唱されたプロセスと同様である。その後再活性化された情報がワーキングメモリの中の情報と統合されコネクションを作る (統合)。この処理は読んでいる情報と再活性化された情報が最小閾値を超えるとすぐに行われてしまうため、統合過程では読解している情報に関連するもの実際には誤った情報がコネクションとして統合されてしまう可能性がある。そして、この統合過程で構築されたコネクションが先行知識に照らし合わせて正しいかどうかを判断する処理が確認処理である (確認)。確認処理も統合処理が閾値を超えなければ作動せず、それぞれのプロセスが自動的に制御不可な形で行われることが明らかになっている。

2.3 英文読解における推論

前述したように、推論とはテキスト情報に自分の知識を合わせて心の中にテキストのイメージをつくることと定義され、そのはたらきはテキスト情報に自分の知識を合わせて本文中に書かれていない情報を導くことであるとされている (McNamara & Kendeou, 2011)。

推論は大きく分けて橋渡し推論と精緻化推論に分類することができる。橋渡し推論はテキストを読み進める上で必須の推論であり、本文中の意味的なギャップを埋める推論である。一方で精緻化推論はテキストを読み進める上で必須

ではなく、背景知識を用いて本文中に書かれていない情報を導入する推論を指す。例えば、単語と単語を結びつける単語統合や、He が指すものを明らかにする照応関係の推論は橋渡し推論にあたり、次に起こる出来事を予測するような予期的推論は精緻化推論にあたる。また、推論にはその方向による分類方法もある。向きの観点からは、(a) 読んでいる内容とそれ以前の内容を結び付ける推論である後3向き推論、(b) 読んでいる内容について情報を補足する直交的推論、(c) 読んでいる内容から次にくる内容を予測する前向き推論の3つに分類される (van Dijk & Kintsch, 1983)。

では、実際の読解プロセスの中で推論はどのように説明できるのか。RI-Val モデル (Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b) では自己制御不能な再活性化プロセス、つまり共鳴処理によって導かれる推論として、以下の9つの分類を提案している。

- a) 単語統合 (word-to-text integration): テキストの単語や内容を先行している情報と結び付ける推論
- b) 照応参照 (anaphoric references): 照応形 (注1) を先行した情報と結び付ける推論
- c) 因果的先行詞 (causal antecedents): ある出来事とその因果的先行詞を結び付ける推論
- d) 具体例 (category exemplars): 本文では述べられていない具体例について推測する推論
- e) 道具 (instruments): 本文では言及されていないある動作に用いられた身体の一部や道具、手段を推測する推論
- f) 論理的議論 (logical arguments): 本文では言及されていない論理的な議論を経たのちの結果を推測する推論
- g) 予期的推論 (predictive inferences): 本文では言及されていない、予測される出来事を推測する推論
- h) 感情 (emotion): 本文中の出来事と感情を反映する些細な情報を結びつけて、本文中には明示的には述べられていない感情を推測する推論
- i) 感覚運動 (sensory motor): 物語文中の出来事や行動が含意する感覚運動を推測する推論

これらの推論は全て自己制御不能な共鳴プロセスによって活性化が可能であり、活性化、状況モデルへの統合のプロセスはRI-Valモデルの観点から説明可能である。推論は読み手が情報の符号化を行い先行文脈や背景知識から情報の再活性化が行われると同時に生成される。生成された推論はワーキングメモリ内にとどまっている情報と統合されてコネクションを作る。その後、活性化された推論のコネクションはワーキングメモリ内の情報に対して確認される。その際、より強いマッチングを産出した推論は状況モデルに統合され、そうでないものは捨てられる。そのため、たとえ読み手が文脈や背景知識に照らして誤った推論をしてしまったとしても、文脈や背景知識からより強いサポートを受ければ、その推論は状況モデルに統合されてしまう可能性がある。読解中に文脈上誤った推論が生成されてしまうことがあるのは、このような理由からである。

2.4 教育現場における推論質問

このように推論は読解理解に必要な不可欠なものであり、読解中に読み手は一部の推論を受動的に行っていることがわかっている。では、実際の教育場面において、推論に関わる質問はどのように用いられているのか。まず、英語の授業内で行われる推論発問の効果について見ていく。

田中(2010)は授業内での推論発問について、テキスト情報をもとにテキストには直接書かれていない内容を推論させるものであり、テキストの中でも不明瞭な部分に意識を向けさせ読みの目的を意識させる契機になること、また答えが一つに決まってしまう事実質問に比べて深いテキスト理解を問うことができることを明らかにしている。また、大下・広瀬(1997)は推論発問を行った群と読後感想の記入を行った群では、推論発問の方がテキスト内容をより詳細に記憶していることを示している。このようなことから推論発問は読解指導にテキストに対する理解やテキスト情報の記憶を促進する効果があると言える。しかしながら同時に、田中(2010)や横田(2017)では能力や知識が一定ではない様々な学習者が集う教室では、読み手に深いテキスト理解を問う推論質問は使いづらく、授業内で活用するには、

多くの困難点があることを明らかにしている。

では、読解能力を測定するテストでは推論に関わる質問はどのように用いられているのか。先述のように一般的に明示的な情報を読み取る技能が低次の技能とされるのに対し、推論は高次の読解技能の一つと考えられ、優れた読み手を特徴づける技能であると言われることもある(吉田, 2001)。つまり、推論質問は優れた読み手とそうでない読み手を弁別することができる可能性を秘めている。しかしながら、清水(2005)では、英検、TOEFL、センター試験のそれぞれの大規模読解テストで用いられている読解質問のタイプは異なり、用いられている推論質問の割合はあまり高くないことが明らかになっている。また卯城(2009)は、英語教員は答えが複数出ることを恐れ、定期テストなどの小規模のテストにおいても読み手の推論を問う質問の出題は避けられる傾向があると指摘している。特に精緻化推論は読み手の背景知識を利用してテキストに書かれていない情報を推測するため、読み手の経験や特性によってその内容に差が生まれてしまう可能性がある。

2.5 本研究の目的と概略

推論質問は読み手の深い読みを問うことができるため、優れた読み手とそうでない読み手を弁別することができる可能性を秘めている。しかしながら、実際に国内の大規模テストではどのような推論質問が出題されているのか、また具体的にどのような推論質問が推論テストの成績の高い/低い受験者を弁別することができるのかについては明らかになっていない。

そこで本研究では、RI-Valモデル(Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)の受動的で自己制御不能な活性化プロセスの仮定に基づいて、読み手の中で受動的に活性化、統合、確認されていると考えられる推論質問をCook and O'Brien(2017)を参考に9つに分類した。そして、その中でどのようなタイプの推論質問が出題されやすいのか、また弁別力の高いテスト作成に貢献するののかを2つの調査を通して検証する。

調査1では、大規模英語テストの読解セクショ

ンで用いられている推論質問を収集し、その種類を分類することによって、国内の大規模テストの読解セクションにおける多肢選択式推論質問の現状と出題方法を把握する。調査2では調査1で分類した推論質問の中から推論テストの成績の高い/低い受験者を弁別する推論質問を解明する。

3 調査1

3.1 目的

調査1は、日本の大規模英語テストの読解セクションで用いられている推論質問を収集し、RI-Valモデル(Cook & O'Brien, 2014; O'Brien & Cook, 2016a, 2016b)の受動的で自己制御不能な活性化プロセスの仮定に基づいて推論質問を9つに分類し、国内の大規模テストにおける多肢選択式推論質問の現状と出題方法を把握することを目的とする。具体的には、英検2級、TOEFL、センター試験の読解セクションで用いられている推論設問を分類し、テスト間での推論質問のタイプに有意な差は見られるのか、もしそうであればそれぞれのテストにどのような特徴が見られるのかを検証する。調査1における検証課題 (Research Questions: RQs) は以下の通りである。

RQ1-1

国内大規模英語読解テストの種類と出題されている推論質問のタイプに統計的な差はみられるか。

RQ1-2

それぞれの国内大規模英語読解テストで出題される推論質問にはどのような特徴の違いがみられるか。

3.2 方法

3.2.1 使用するテスト問題の収集

本調査では国内大規模英語読解テストとして、英検2級、TOEFL(注2)、センター試験(注3)の3つを選択し、その長文読解セクションで出題されている多肢選択式質問を収集した。その中でも

資料の活用が問われる設問や複数文章読解を必要とする長文セクションは除外し、以下の質問を最終的な分析対象とした。

- 1) 英検2級: 2018年度第3回から2020年度第2回までの全6回の試験における3B, 3Cの設問, 全54問
- 2) TOEFL: The Official Guide to the TOEFL test (5th edition) に掲載されているTOEFL Practice Test 1, 2の4つの長文問題における全ての設問, 全52問
- 3) センター試験: 2017年から2020年までの全4回の試験における4A, 5, 6Aの全設問, 全56問

3.2.2 分析で使用する指標

推論質問は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と定義される(清水, 2005)。本研究では、読み手が受動的に行っていると考えられる推論質問をCook and O'Brien (2017)に基づいて9つに下位分類した。いずれにも当てはまらないものはその他の質問に分類した。基準は以下の通りである。また以下に示す具体例はCook and O'Brien (2017), Graesser, Singer and Trabasso(1994), Lea and Mulligan (2002)を参考に作成した。

a) 単語統合(word-to-text integration)

定義: テキストの単語や内容を先行している情報と結び付ける推論について問う質問

具体例: "While Cathy was riding her bike in the park, dark clouds began to gather, and it started to storm. The rain ruined her beautiful sweater." を読み、The rainの意味や関連する情報について問う質問

b) 照応参照(anaphoric references)

定義: 照応形を先行した情報と結び付ける推論について問う質問

具体例: "He heard the news that his father was dying. It makes him go to the hospital." を読み、itの指す内容を問う質問

c) 因果的先行詞(causal antecedents)

定義: ある出来事とその因果的先行詞を結

び付ける推論について問う質問

具体例：“In his haste and abstraction he stuck a pickle fork into his right eye…” という部分を読んだときに、なぜ彼はフォークを目に刺したのか問う質問

d) 具体例(category exemplars)

定義：本文では述べられていない具体例について推測する推論について問う質問

具体例：“When two characters were on a picnic and were attacked by bugs”を読み、bugsが何かを問う質問

e) 道具(instruments)

定義：本文では言及されていないある動作に用いられた身体の一部や道具、手段を推測する推論について問う質問

具体例：“… the Man to Whom Time was Money had to sustain life by fishing from the end of a wharf ”という部分を読み、男が何を手にもっているかを問う質問

f) 論理的議論(logical arguments)

定義：本文では言及されていない論理的な議論を経たのちの結果を推測する推論について問う質問

具体例：“According to the recipe they could not use both fresh jalapenos and ground pepper in the chili.” “Well,” said Eric looking in the fridge, “let’s see what we have handy. We have got plenty of fresh jalapenos — let’s use them!” he said.”という文章を読み、結果的に料理の中に入っている香辛料を問う質問

g) 予期的推論(predictive inferences)

定義：本文では言及されていない、予測される出来事を推測する推論について問う質問
“The actress fell from the 14th story.”を読み、女優がどうなったのかを問う質問

h) 感情(emotion)

定義：本文中の出来事と感情を反映する些細な情報を結び付けて、本文中には明示的には述べられていない感情を推測する推論について問う質問

具体例：“He heard the news that his father was dying. It made him go to the hospital.”を読み、彼の感情を問う質問

i) 感覚運動(sensory motor)

定義：物語文中の出来事や行動が含意する感覚運動を推測する推論について問う質問

具体例：“The hero defeated enemies with a ray attack.”を読み、ヒーローの動きについて問う質問

3.2.3 手順

設問は英語教育学を専攻する大学院生3名によって、「3.2.2 分析で使用する指標」に記載した9つのいずれかに分類された。いずれにも当てはまらない項目はその他の質問に分類された。はじめに3名それぞれが162個の設問の30%を分類した。その際の一致率は83.7%であった。不一致項目に対して不一致協議を行った後に、それぞれが分担して残りの70%の評定を行った。手順は以下の通り(図1)。



■ 図1: 調査1の実施手順

3.3 結果

表1, 2からわかるように, 国内大規模テストの種類と出題される推論質問の関係には明らかな

違いが見られた。特にTOEFLとセンター試験では全体の設問の30%~40%が推論質問であったのに対し, 英検2級では推論質問は一問も出題されていなかった。

■表1: 英検, TOEFL, センター試験の質問タイプの項目ごとの分類

質問タイプ	英検	TOEFL	センター試験
a) 単語統合		項目62 (8) 項目68 (1) 項目76 (9) 項目81 (1) 項目82 (2) 項目89 (9) 項目98 (5) 項目103 (10)	項目117 (6A/2) 項目130 (6A/1) 項目142 (5/4) 項目144 (6A/1) 項目156 (5/4) 項目160 (6A/3)
b) 照応参照		項目71 (4)	
c) 因果的先行詞		項目56 (2) 項目100 (7)	項目112 (5/2) 項目113 (5/3) 項目114 (5/4) 項目125 (5/1) 項目139 (5/1) 項目141 (5/3) 項目146 (6A/3) 項目153 (5/1)
d) 具体例		項目57 (3)	
e) 道具		項目75 (8)	
f) 論理的議論		項目63 (9) 項目87 (7)	項目147 (6A/4) 項目152 (4A/4)
g) 予期的推論		項目94 (1)	項目110 (4A/4) 項目124 (4A/4)
h) 感情			項目126 (5/2) 項目129 (5/5) 項目157 (5/5)
i) 感覚運動			項目128 (5/4)
その他	項目1 (3B/1) 項目2 (3B/2) 項目3 (3B/3) 項目4 (3B/4) 項目5 (3C/1) 項目6 (3C/2) 項目7 (3C/3) 項目8 (3C/4) 項目9 (3C/5) 項目10 (3B/1) 項目11 (3B/2) 項目12 (3B/3) 項目13 (3B/4) 項目14 (3C/1) 項目15 (3C/2) 項目16 (3C/3) 項目17 (3C/4) 項目18 (3C/5) 項目19 (3B/1) 項目20 (3B/2) 項目21 (3B/3) 項目22 (3B/4) 項目23 (3C/1) 項目24 (3C/2) 項目25 (3C/3) 項目26 (3C/4) 項目27 (3C/5) 項目28 (3B/1) 項目29 (3B/2) 項目30 (3B/3) 項目31 (3B/4) 項目32 (3C/1) 項目33 (3C/2) 項目34 (3C/3) 項目35 (3C/4) 項目36 (3C/5) 項目37 (3B/1) 項目38 (3B/2) 項目39 (3B/3) 項目40 (3B/4) 項目41 (3C/1) 項目42 (3C/2) 項目43 (3C/3) 項目44 (3C/4) 項目45 (3C/5) 項目46 (3B/1) 項目47 (3B/2) 項目48 (3B/3) 項目49 (3B/4) 項目50 (3C/1) 項目51 (3C/2) 項目52 (3C/3) 項目53 (3C/4) 項目54 (3C/5)	項目55 (1) 項目58 (4) 項目59 (5) 項目60 (6) 項目61 (7) 項目64 (10) 項目65 (11) 項目66 (12) 項目67 (13) 項目69 (2) 項目70 (3) 項目72 (5) 項目73 (6) 項目74 (7) 項目77 (10) 項目78 (11) 項目79 (12) 項目80 (13) 項目83 (3) 項目84 (4) 項目85 (5) 項目86 (6) 項目88 (8) 項目90 (10) 項目91 (11) 項目92 (12) 項目93 (13) 項目95 (2) 項目96 (3) 項目97 (4) 項目99 (6) 項目101 (8) 項目102 (9) 項目104 (11) 項目105 (12) 項目106 (13)	項目107 (4A/1) 項目108 (4A/2) 項目109 (4A/3) 項目111 (5/1) 項目115 (5/5) 項目116 (6A/1) 項目118 (6A/3) 項目119 (6A/4) 項目120 (6A/5) 項目121 (4A/1) 項目122 (4A/2) 項目123 (4A/3) 項目127 (5/3) 項目131 (6A/2) 項目132 (6A/3) 項目133 (6A/4) 項目134 (6A/5) 項目135 (4A/1) 項目136 (4A/2) 項目137 (4A/3) 項目138 (4A/4) 項目141 (5/2) 項目144 (5/5) 項目146 (6A/2) 項目148 (6A/5) 項目149 (4A/1) 項目150 (4A/2) 項目151 (4A/3) 項目154 (5/2) 項目155 (5/3) 項目158 (6A/1) 項目159 (6A/2) 項目161 (6A/4) 項目162 (6A/5)

その上で、推論質問とテストの種類に統計的な差異があるかどうかを検証するため、計162個の質問項目に対して、その関連をカイ2乗検定で分析した。その結果 $\chi^2(18) = 41.34$, 漸近有意確率 $p < .01$, Cramer's $V = .36$ と中程度の有意な関連が示唆された。クロス集計表において、期待値が0未満のセルが存在したため、Fisherの正確確率検定を参照した。その結果も、 $p < .01$ と1%水準で有意であった。

そこで調整済み残差を確認したところ、その他

の質問は英検で極端に多く ($z = 5.0, p < .01$), センター試験で極端に少なかった ($z = -3.5, p < .01$)。また、推論質問では a) 単語統合質問は英検で極端に少なく ($z = -2.8, p < .01$), TOEFL で有意に多いこと ($z = 2.1, p < .05$), c) 因果的先行詞質問は英検で極端に少なく ($z = -2.3, p < .05$), センター試験で有意に多い ($z = 3.1, p < .01$) ことが明らかになった。さらに、h) 感情質問はセンター試験で有意に多くみられる ($z = 2.4, p < .05$) こともわかった。

■表2: 英検, TOEFL, センター試験の質問タイプごとの項目数と割合

質問タイプ	英検	TOEFL	センター試験	合計 (%)
a) 単語統合	0	8 (15%)	6 (10%)	14 (8%)
b) 照応参照	0	1 (2%)	0	1 (1%)
c) 因果的先行詞	0	2 (4%)	8 (13%)	10 (6%)
d) 具体例	0	1 (2%)	0	1 (1%)
e) 道具	0	1 (2%)	0	1 (1%)
f) 論理的議論	0	2 (4%)	2 (4%)	4 (2%)
g) 予期的推論	0	1 (2%)	2 (4%)	3 (2%)
h) 感情	0	0	2 (4%)	3 (2%)
i) 感覚運動	0	0	1 (2%)	1 (1%)
推論質問	0 (0%)	16 (31%)	21 (37%)	37 (24%)
その他の質問	54 (100%)	36 (69%)	35 (63%)	125 (76%)
合計	54	52	56	162

3.4 考察

統計的な結果を踏まえ、以下のような考察が可能である。

英検2級では、ほとんど全ての質問がその他の質問に分類された。この結果は清水(2005)によって示されたものと同様である。この結果について以下のような可能性が考えられる。英検の受験者は小学生から社会人まで多岐にわたり、英検は英語学習の初期段階から生涯にわたる継ぎ目のないスコア尺度としての役割を果たしている。先にも述べたように精緻化推論をはじめとする読解中の推論は文脈の内容だけでなく読

み手の背景知識にも依存するため、導かれる回答は読み手の特性によって異なる可能性があることが知られている(卯城, 2009)。そのため、推論質問は時に選択肢の妥当性の観点から指摘を受ける場合がある。このような推論質問の欠点を考慮し、英検では大規模テストとしての設問の妥当性を担保する必要性があることから、推論質問を出題していない可能性が考えられる(注4)。

一方で、TOEFLとセンター試験では全体の設問の30%~40%が推論質問であった。この結果は清水(2005)のものと大きく異なる。清水では、TOEFLでは14%、センター試験では56%の割合

で推論質問が出題されていることが明らかになっている。このような先行研究における結果との相違が見られた理由として以下のような可能性が考えられる。

- (1) 分類方法の違い: 本研究では Cook and O'Brien (2017) に基づいて推論質問を9つに分類した。一方で清水では Graesser et al. (1994) に基づき推論質問を8つに分類している。この分類方法の違いが本研究と先行研究の結果の違いにつながっている可能性がある。
- (2) サンプル数: センター試験における結果の違いに関して、本調査1では56項目の設問に関して分析を行った。一方で、清水は9項目に対してのみ分析を行っている。センター試験で見られた推論質問の割合に20%ほどの開きがある理由として、研究間でのサンプル数の違いが関係している可能性がある。
- (3) テスト出題傾向の変化: TOEFL においてはテスト自体の出題傾向が変化した可能性も考えられる。清水が分析しているのは TOEFL (Educational Testing Service, 1999) であった。一方で本調査1では The Official Guide to the TOEFL test (5th edition; Educational Testing Service, 2020) を分析対象としている。両者の発行年数の間には20年ほどの開きがあり、その間 TOEFL のテスト自体が読み手の推論能力の測定に重きを置くように変化してきた可能性がある。

先述のように、全体として TOEFL とセンター試験では30%から40%の割合で推論質問が出題されることが明らかになったが、出題されている具体的な推論質問の質問タイプには違いがあるのだろうか。結論から言うと、それぞれのテストの間で出題される推論質問のタイプには差が見られた。特に TOEFL ではテキストの単語や内容を先行している情報と結び付ける推論について問う単語統合質問が有意に多く出題されていた。その理由として以下の可能性が考えられる。

- (1) 先行研究では TOEFL の実施目的は「実際の大学での生活を反映した状況やタスクにおいて、英語を母語としない人のコミュニカ

ティブな言語能力を測定すること」(筆者訳; Jamieson, Jones, Kirsch, Mosenthal & Taylor, 2000, p. 10) と述べられている。また、TOEFL を実施している Educational Testing Service (2021) は、「TOEFL iBT® の読解セクションは大学などのアカデミックな環境で使用されるようなマテリアルをどの程度読んで理解できるのかを評価するためにデザインされた」(筆者訳) と記載されている。このようなことから、TOEFL では主に大学などのアカデミックな環境におけるリーディング力またはコミュニケーション能力の測定を目的としていることがわかる。このような理由から TOEFL では、物語文などで多く問われる g) 予期的推論質問や h) 感情質問ではなく、アカデミックな文章の読解に必要な不可欠な a) 単語統合質問(符号化している情報と先行情報の繋がりを推測する質問) が多く出題されている可能性がある。

- (2) また、Educational Testing Service (2021) には「回答に必要な全ての情報はパッセージの中に書かれている」(筆者訳) と言った記載がある。一般的に精緻化推論はテキスト理解に必ずしも必要とはされない一方で、a) 単語統合などの文章内の意味的ギャップを埋める推論はテキスト理解に必要な不可欠であるとされている (van Dijk & Kintsch, 1983)。このような理由から TOEFL では a) 単語統合質問が有意に多く出題されている可能性もある。

最後にセンター試験は大学入学の際の共通テストとしての意味合いが強く、「大学に入学を志願する者の高等学校段階における基礎的な学習の達成の程度を判定する」試験として実施されている(文部科学省, 2012)。センター試験には実際に高等学校の範囲内から出題されているかどうかを評価する、外部評価分科会も存在する。

高等学校学習指導要領(文部科学省, 2019) の第8節第1款目標(2)では、「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」できる力を養うことが明示されている。そのため、センター試験の出題意図の一つにも、学習指導要領にも記載があるような話し手や書き手の意図などを的確に理解

できているかを測定することが盛り込まれている。こうした話し手や書き手の意図の的確な理解を測定するために、センター試験では文章内で because などと明示的に示されていない因果的なつながりを問う c) 因果的先行詞質問や、人物の感情を問う h) 感情質問が意図的に多く出題されていると考えられる。これはセンター試験という試験が持つ役割や構成概念の妥当性の観点から見ても適当である。

3.5 調査1のまとめ

調査1では、英検2級、TOEFL、センター試験の読解セクションで用いられている推論設問を9つに分類し、テスト間での推論質問のタイプに有意な差は見られるのか、もしそうであればそれぞれのテストにどのような特徴が見られるのかを検証した。

結果より、テスト間で用いられる推論質問のタイプには統計的な違いが見られた。具体的には英検では推論質問の出題がほとんど見られなかった一方で、TOEFLとセンター試験では全体の30～40%程度を推論質問が占めていた。また、その中でもTOEFLでは符号化している情報と先行情報の繋がりを推測する a) 単語統合が、センター試験では c) 因果的先行詞や h) 感情が有意に多く問われていることが明らかになった。さらに、それぞれのテスト実施の目的が3つの大規模テスト間の推論質問のタイプの異なりに影響を与えていることが示唆された。

続く調査2では、こうした9つの推論質問の中でどの質問がよりよく学習者を弁別することができるかを調査した。

4 調査2

4.1 目的

調査2では調査1で分類された9つの推論質問の中から推論テストの成績の高い/低い受験者を弁別することができる推論質問のタイプを解明する。具体的には、TOEFL で出題されている4つ

のテキストを使用し、9つの読解質問を各2つずつ計18つの推論質問を作成する。そして、その推論質問の各項目における(1)項目難易度、(2)点双列相関係数、(3)項目弁別力指数を古典的テスト理論の観点から明らかにする。調査2におけるRQは以下の通りである。

RQ2

どのようなタイプの推論質問が推論テストの成績の高い/低い受験者を弁別することができるのか？

4.2 方法

4.2.1 協力者

調査には茨城県内の国立大学に通う日本人大学生・大学院生49名が参加した。サンプルは便宜的に抽出した。参加者は男性22名、女性27名であり、平均年齢(標準偏差)と範囲はそれぞれ21.88(2.128)であった。また、それぞれの専攻は人文学、工学、教育学、心理学、医学、国際関係学など多岐にわたり、全ての受験者は少なくとも中学高校の6年間は英語を学習していた。さらに、受験者の読解熟達度は事前アンケートの大規模英語テストのスコアによって、CEFRにおいてB1からC1レベルであると推定された。

4.2.2 マテリアル

調査で使用したテキストはThe Official Guide to the TOEFL test (5th edition; Educational Testing Service, 2020)に掲載されているTOEFL Practice Test 1, 2の中の4つの長文テキストである。テキスト間のレベルを一定にするため、マテリアルで使用される単語は新JACET 8000 リスト(Committee of revising the JACET basic words, 2016)のレベル5以下の単語に統一した。また、レベル6以上の全ての単語はレベル5以下の単語に置き換えるか注釈を付与した。マテリアルの語数と難易度は表3に記載している。

■表3: 調査2における各テキストの特性

テキスト名	A	B	C	D
語数	590	557	705	674
FRE	35.7	43.4	55.6	42.9
FKGL	13	11.9	10.5	13.6

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

設問については調査1で分類した9つの推論のタイプが各2つずつ、計18題の推論質問を出題した。そのうちTOEFL Practice Test 1, 2の中で実際に使用されていた質問は10題で、残りの8題については英語教育学を専攻する大学院生がそれぞれの推論質問の定義と具体例を参

考に作成した。表4には項目と出題テキスト、それぞれの推論質問のタイプ、作成の有無を記載している。

■表4: 各項目における質問タイプと自作の有無

項目	出題テキスト	推論質問のタイプ	作成の有無
1	A	c) 因果的先行詞	
2	A	d) 具体例	
3	A	a) 単語統合	
4	A	f) 論理的議論	
5	A	d) 具体例	自作
6	B	i) 感覚運動	自作
7	B	h) 感情	自作
8	B	b) 照応参照	
9	B	e) 道具	
10	B	g) 予期的推論	自作
11	C	a) 単語統合	
12	C	e) 道具	自作
13	C	f) 論理的議論	
14	C	b) 照応参照	自作
15	D	g) 予期的推論	
16	D	i) 感覚運動	自作
17	D	c) 因果的先行詞	
18	D	h) 感情	自作

4.2.3 手順

図2は調査2の実施手順である。COVID-19の感染拡大による影響を考慮し、調査はZoomを用いて非対面で行われた。初めにZoomのセッション

を開き参加者には調査の目的と注意点を説明した。その際に推論テストをオンライン上で配布し、参加者は各自設問に回答した。参加者ごとのテストへの回答時間が結果に与える影響を考慮し、

テストへの回答時間は60分とした。なお、問題用紙への書き込みは不可とし、参加者にはモニター

上で設問に回答することを指示した。



■図2: 調査2の実施手順

4.2.4 分析

実験実施上の手順に問題があった3名のデータを除外し46名のデータに基づいて分析を実施した。古典的テスト理論に基づき、それぞれの項目に対してそれぞれ統計ソフトRのlme4パッケージを用いて古典的項目分析を行った。項目難易度を測定するために項目難易度指標、項目弁別力を測定するために点双列相関係数と項目弁別力指標を使用した。

項目難易度の指標である項目難易度指標は正答した受験者の受験者全体に占める割合を指す。この指標は0から1までの値を取り、0に近づくほどその項目が参加者にとって難易度が高いことを示す。ブラウン(1999)より.29以下の項目を参加者にとって不当に難しすぎる可能性がある項目、また.71以上の項目について易しすぎる項目と判定した。

一方で項目弁別力の指標である点双列相関係数は項目の得点と合計点の相関係数を計算するもので、-1から1までの値をとる。ある項目で値が1に近づくほど項目得点と合計点の相関は強い。点双列相関係数における分析では渡部(2012)より.19以下を不良項目、伊東(2011)より.40以上を最良項目と判定した。

また、項目弁別力指標はその項目によってどの程度成績上位者と下位者が区別できるかの指

標になる。具体的には全体の正答数が上位3分の1の正答率から下位3分の1の正答率を引いたものがこの指標になる。項目弁別力指標は、大友(1996)に基づき、.40以上が最良項目、.19以下が不良項目として判定した。

信頼性指標としては、クロンバックの α 係数を算出した。 α 係数は、「すべての項目が同じ対象を測定しているなら、項目間の内的整合性(内的一貫性)が高まることをもとに信頼性の下限値を示す指標」としてよく用いられる(小野寺・山本, 2004, p. 192)。

4.3 結果

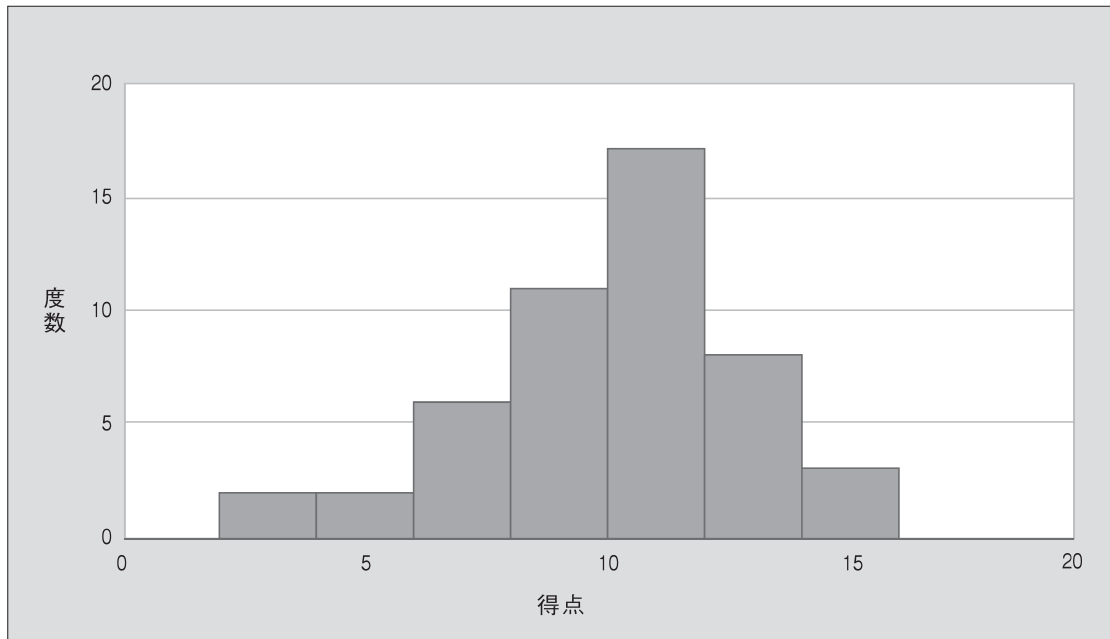
4.3.1 記述統計, 得点分布

テストの得点分布を確認するために表5に推論テストの記述統計を、図3にヒストグラムを示す。

推論テストは満点18点であったが、平均値10.61がテストの中央値に近く、SD, 最大値, 最小値がそれぞれ2.71, 16, 4であることから、床効果と天井効果は見られなかったといえる。また、歪度と尖度の値も基準の範囲内であり、極端に正規性を欠く分布ではないことがわかった。

■表5: 調査2の推論テストにおける記述統計

N	M	SD	最大値	最小値	range	歪度	尖度
46	10.61	2.71	16	4	12	-0.46	-0.11



■ 図3: 調査2の推論テストにおける得点分布

■ 表6: 調査2の項目ごとの項目難易度, 点双列相関係数, 項目弁別力指標

項目(タイプ)	項目難易度	点双列相関係数	項目弁別力指標
1. c) 因果的先行詞	.78	.33	.13
2. d) 具体例	.72	.20	.19
3. a) 単語統合	.78	.20	.19
4. f) 論理的議論	.17	.37	.33
5. d) 具体例	.37	<u>.40</u>	.38
6. i) 感覚運動	.63	<u>.42</u>	<u>.48</u>
7. h) 感情	.59	-.07	-.21
8. b) 照応参照	.60	<u>.48</u>	<u>.48</u>
9. e) 道具	.50	.31	<u>.43</u>
10. g) 予期的推論	.50	.31	<u>.43</u>
11. a) 単語統合	.28	<u>.41</u>	<u>.44</u>
12. e) 道具	.91	.36	.24
13. f) 論理的議論	.73	<u>.41</u>	.25
14. b) 照応参照	.57	<u>.46</u>	.37
15. g) 予期的推論	.26	.31	.33
16. i) 感覚運動	.89	.34	.12
17. c) 因果的先行詞	.72	.36	.30
18. h) 感情	.59	<u>.50</u>	<u>.60</u>

注) 塗りつぶしは本研究の基準における不良項目, 下線は最良項目を指す。

4.3.2 信頼性

テストの信頼性とはテストの内的な整合性を指す(齊田, 2011)。つまり、信頼性はテストの各項目間にどの程度の一貫性があるのかを指し示す。

本調査で実施した18項目に対して信頼性を検討したところ、クロンバックの α 係数は.61であった。この値はテストの内的一貫性としては十分な値と言えない。しかしながら、本調査では9つの推論のタイプを各2つずつ、計18題の推論質問のみを出題している。本調査においてはこの項目数の少なさが、内的一貫性の値の不十分さにつながった可能性が高い。

4.3.3 古典的項目分析

表6には各項目の項目難易度、点双列相関係数、項目弁別力指標を示している。塗りつぶしが不良項目、下線が最良項目である。

項目難易度に関して、.17から.91までの値が見られ、.30-.70を逸脱している項目が10確認された。項目4, 11, 15は.29以下であるため参加者にとって難しすぎたが、項目1, 2, 3, 12, 13, 16, 17は.70以上であるため参加者にとっては易しすぎる問題であったと言える。

点双列相関係数による弁別力指数分析では、不良項目にあたる項目は7のみであった。一方で、最良項目とされる.40以上の値をとる項目は7つ存在した(項目5, 6, 8, 11, 13, 14, 18)。また、上位33%の得点と下位33%の受験者の正答率を比較した項目弁別力指標では不良項目とされる項目が5項目(項目1, 2, 3, 7, 16)存在した。一方で、最良項目とされる項目は6項目(項目6, 8, 9, 10, 11, 18)であった。

4.4 考察

調査2の目的は調査1で分類した9つの推論質問の中から推論テストの成績の高い/低い受験者を弁別することができる推論質問のタイプを説明することであった。本セクションでは推論質問のタイプごとの考察を行う。

まず、読み手の橋渡し推論を問う3タイプの推論質問について考察する。まず、項目3と項目11からなる質問がa)単語統合である。結果より、項

目難易度が低かった項目3では読み手を弁別することができていなかった。しかし、項目難易度が高い項目11は弁別力が高いことが示された。つまり、推論テストの点数が高い読み手は、符号化している情報に適切に関連する情報を前の文脈から再活性化、統合しながら読解を進めるため、項目の難易度が高かったとしても正答することができた一方で、推論テストの点数が低い読み手は項目難易度の高い単語統合質問の回答には困難を抱えた可能性が示唆された。このことから比較的項目難易度の高い単語統合質問が読み手の推論能力を弁別可能であることが示される。しかしながら、この結果は単語統合質問の難易度は問題によって大きく異なることも同時に示唆している。そのため、テストを実施する際には選択肢や問題の調整を行い、項目難易度が最良となるような問題を作成することが必要である。

続いて項目8と項目14からなるb)照応参照質問である。結果よりどちらの項目でも項目難易度は最適値(.50)に近く、項目の弁別力も高いことが示された。照応参照質問は代名詞や単語、フレーズなど前出の情報に言及した照応形を先行した情報と結び付ける推論について問う質問(Cook & O'Brien, 2017)である。結果より推論能力が高い読み手は目標となる代名詞や指示語、いわゆる照応形を符号化する際に、受動的な再活性化プロセスの中で適切な先行情報を活性化することができるが、そうでない読み手は誤った情報を活性化してしまう可能性が高いことが示唆された。そのため、この照応関係について問う質問が読み手の推論能力を弁別する質問となる可能性が考えられる。

次に項目1, と項目17からなるc)因果的先行詞質問である。どちらの項目においても項目難易度は低く、弁別力も最良とまではいえることができなかった。因果的な一貫性はL2の読み手にも知覚されやすく、読み手は読解中に因果的一貫性を維持しながら読解を行なっていることが示されている(Ushiro et al., 2018)。このように、因果的先行詞を問う質問は推論能力があまり高くない読み手にとっても難易度が高くないため、L2学習者が正答を導きやすく、それぞれの項目における弁別力も最良とは言えなかったと考えられる。

ここからは精緻化推論である。項目2と項目5

からなる d) 具体例質問, つまり本文では述べられていない具体例について推測する推論について, 項目難易度が低かった項目2は弁別力が低く, 改良の必要性が示された。一方で項目難易度が適切だった項目5は点双列相関係数の指標で弁別力が高いことが示された。この結果は, 項目難易度が適切であれば具体例質問は推論能力の有無を十分に弁別できる可能性を示唆した。しかしながら, 項目難易度によって弁別力に差が生じるため, 常に具体例質問が推論能力を弁別可能であるということとはできない。

続いて項目9と項目12の e) 道具質問について, 項目弁別力指標より項目難易度が最適であった項目9では最良項目であることが示されたものの, 項目12では項目難易度が易しすぎたため弁別力は最良であるとは言えなかった。道具質問は本文では言及されていないある動作に用いられた身体の一部や道具, 手段を推測する推論について問う質問である。このタイプの質問に関しても d) 具体例質問と同様に, 項目によって問題の難易度が変化する可能性が示唆された。

項目4と項目13の f) 論理的議論はどちらも項目難易度が適切とは言えなかった。項目4は難易度が高く, 項目13は難易度が低いことが示された。また, 弁別力も高いことが示されたのは項目14の点双列相関係数でのみで, 全体としては論理的議論質問が読み手の推論能力を弁別する質問になるとは言えなかった。さらに, g) 予期的推論(項目10, 15)に関して, 項目10では項目難易度が最適, また項目弁別力指標で最良項目であると示されたが, 項目15では項目難易度が低く, 弁別力が高いとは言えなかった。

精緻化推論にあたる c)~g) では同じ質問のタイプであっても項目によって難易度に差が見られた。また, 弁別力も1つの項目で最良となることはあっても, 点双列相関係数・項目弁別力指標の双方で最良となる項目や質問タイプは見られなかった。精緻化推論は一般的にテキストを読み進める上で必須ではなく, 背景知識を用いて本文中に書かれていない情報を導入する推論であると言われる(van Dijk & Kintsch, 1983)。そのため, 精緻化推論においては導かれる推論自体が読み手の特性や経験によって大きく左右され, 読み手が本来は本文から推測できないような内容を,

自身の背景知識に基づいて再活性化させてしまう可能性が大きい(Cook & O'Brien, 2017)。本調査においても, 参加する読み手が文脈よりも自身の背景知識に基づいて推論を活性化させたため, 質問のタイプよりもそれぞれの項目に応じて難易度や弁別力が大きく変化し, 安定した弁別力の測定には至らなかった可能性がある。

最後に h) 感情と i) 感覚運動を取り上げる。項目7と項目18に当たる h) 感情質問では非常に興味深い結果が得られた。どちらも項目難易度は同じ値(.59)であったが, 項目7は点双列相関係数・項目弁別力指標どちらの指標も明確に不良項目であることを示した一方で, 項目18では点双列相関係数・項目弁別力指標どちらの指標も明確に最良項目であることが示された。感情に関する推論は体の認知的な枠組みとの関わりが指摘されており, 実際に顔の筋肉を緊張させることで感情に関する推論の活性化が阻害されることが明らかになっている(Havas, Glenberg, Gutowski, Lucarelli & Davidson, 2010)。このように感情に関する推論の活性化は読み手の身体的な認知の枠組みとの関わりが強く, 正答を導けるかどうかは推論能力ではなく読み手の環境や感情に左右されてしまう。そのため, 本調査においても安定した弁別力の測定には至らなかった可能性がある。

項目6, 16に当たる i) 感覚運動質問は, 物語文中の出来事や行動が含意する感覚運動を推測する推論について問う質問を指す。項目6では項目難易度が最適値に近く, 点双列相関係数・項目弁別力指標どちらにおいても弁別力が高いことが示された。一方で, 項目16では項目難易度が低く, 項目弁別力指標では不良項目であると明らかになった。この結果より, d) 具体例質問や e) 道具質問と同様に, 感覚運動質問においても項目難易度が適切であれば推論能力の有無を十分に弁別できる可能性が示唆された。

全体として, 本調査2では推論質問のタイプによって項目難易度や弁別力が変化するというよりも, 項目によって項目難易度や弁別力が変化するという傾向が見られた。具体的に, 項目難易度が0.30-0.70の間に存在し難易度が適切であると言える設問においては, 項目7を除いて1つ以上の指標で弁別力が最良であることが示された。

一方で、項目難易度の指標が0.30-0.70の間に収まっておらず難易度が適切とは言えない推論質問では弁別力が最良とは言えない傾向が見られた。この結果は、項目難易度が適切であれば概ね推論質問としての弁別力が高く、そうでなければ弁別力が低くなることを意味するだけでなく、L2学習者は概ねどのタイプの推論質問であっても適切な推論の再活性化が可能であることを示唆している。

しかしながら、一部の推論質問のタイプでは弁別力に差が見られることも明らかになった。b) 照応参照は項目難易度が常に最適値に近く、点双列相関係数・項目弁別力指標の値もがともに安定して高かったため、推論能力の有無を安定して弁別することができる可能性がある。一方で、c) 因果的先行詞は項目難易度が低く、点双列相関係数・項目弁別力指標も最良とは言えないため、推論能力を弁別するのに適していない可能性がある。それでも、各2つの項目で同様の結果が見られたのは橋渡し推論に属する照応参照、因果的先行詞のみで、精緻化推論や感情、感覚運動に関する推論質問では一部を除いて質問タイプというよりも項目の難易度に応じて弁別力が変化する可能性が明らかになった。

4.5 調査2のまとめ

調査2では46名の参加者によるデータに対して古典的項目分析を行い、調査1で分類された9つの推論質問の中から推論テストの成績の高い/低い受験者を弁別することができる推論質問のタイプを検証した。

全体的な結果として、推論質問の弁別力は質問タイプによって変化するというよりも、どのタイプの推論質問であっても項目難易度が適切な項目においてテストの得点を弁別する弁別力が高くなることが示された。一方で、b) 照応参照は安定して推論能力の有無を弁別できる可能性があること、またc) 因果的先行詞はどの読み手にとっても容易に回答できてしまうため、推論能力の有無の弁別には適していない可能性があることが示された。また、RI-Valモデルの観点から、b) 照応参照においては目標となる代名詞や指示語などの照応形を符号化した際に推論能力の高い読

み手は文脈内から適切な情報を活性化できるが、推論能力の低い読み手は文脈から誤った情報を活性化しやすい可能性があることが示唆された。

このことから、教育現場において推論質問を用いたテストを作成する際には、項目難易度が適切な項目を出題することで能力の高い/低いを適切に弁別できる可能性があること、また推論質問として照応参照質問を採用することで学習者の能力を効果的に弁別できる可能性があることが明らかになった。

5 結論と今後の展望

本研究では2つの調査を通じて、大規模テストにおける推論質問の出題の現状と、多肢選択肢問題においてどのような推論質問を出題することが望ましいのかについて検証した。以下に本研究から得られた主要な結果をまとめる。

調査1では、RI-Valモデルの仮定に基づき、読み手が読解中に受動的に活性化していると考えられる9つの推論(Cook & O'Brien, 2017)に基づいて推論質問を9つに分類し、大規模英語テストの読解セクションではどのような推論質問が出題されているのかについて検証した。結果より、それぞれのテスト間で用いられる推論質問のタイプには統計的な違いが見られた。またそれぞれで出題される推論質問には以下のような特徴が見られた;(a) 英検の読解セクションの問題では推論質問の出題がほとんど見られなかった、(b) TOEFLでは文章の難易度が高いアカデミックな文章を読む際に必要不可欠な単語統合推論が多く見られた、(c) センター試験では概要や要点の把握、また書き手の意図を読み取るのに必要不可欠な因果的先行詞や感情などが有意に多く問われていることが明らかになった。また、それぞれのテスト実施の目的が3つの大規模テスト間の推論質問のタイプの違いに影響を与えていることが示唆された。

続く調査2では、調査1で分類した9つの推論質問の中から推論テストの成績の高い/低い受験者を弁別することができる推論質問のタイプを解明した。結果より、照応参照は安定して推論能

力の有無を弁別できる可能性があることが示された。この結果は推論能力が高い読み手は目標となる代名詞や指示語を符号化する際に、受動的な再活性化プロセスの中で適切な照応詞の活性化が可能であるが、そうでない読み手は誤った照応詞を活性化し、誤った情報が心的表象に統合されてしまう傾向が強いことを示唆した。また因果的先行詞は質問タイプ自体の難易度が高くなく、正答率が高くなってしまいうため推論能力の有無の弁別には適していない可能性があることが示された。因果的な繋がりにはL2の読み手にとっても認識しやすく、読み手は常に因果的繋がりを意識していることが示されている(Ushiro et al., 2018)。そのため、推論能力の高い読み手も低い読み手も適切な推論を活性化しやすく、能力の弁別に優れていないことが示唆された。しかしながら、全体の傾向で見ると推論質問の弁別力は質問タイプによって変化するというよりも、個々の項目によって変化する傾向が見られた。特に精緻化推論や感情、感覚運動に関する推論は読み手の背景知識や特性に依存する部分も大きいいため、項目難易度を安定させることが難しいことも同時に明らかになった。

本研究から得られた教育的な示唆は、主に2点ある。1点目は読み手の推論能力を測定する際の照応参照質問の有用性である。調査1より、大規模英語テストの読解セクションにおける照応参照質問の出題数は全152項目中1項目のみであり、国内の大規模リーディングテストにおいては代名詞や単語、フレーズなど前出の情報に言及した照応形を先行した情報と結び付ける推論の能力は問われない傾向にあることが明らかになった。一方で調査2より、照応形が指す情報について尋ねる質問は安定して読み手の推論能力を弁別できる可能性が示唆された。これらの結果を踏まえ、国内の大規模テストにおいては、こうした照応形を先行情報と結び付ける質問をより多く取り入れることで、学習者の推論能力を適切に弁別することが可能になる。

2点目として精緻化推論や感情、感覚運動に関する推論質問出題の可能性である。調査2ではこうした推論質問の中で項目難易度が最適な項目で、読み手の弁別力も高くなる傾向が明らかになった。つまり、導かれる推論が読み手の背景知

識に依存する可能性がある推論質問でも、項目難易度が適切であれば読み手をよりよく弁別できる可能性が示唆された。一方で、同時にこうした推論質問を出題する際には事前に項目難易度を測定し、最適な質問のみを選んで出題することが必須である。

以上のように本研究は、大規模英語読解テストにおける推論質問の可能性を検証した。これまでの先行研究において本研究のように推論質問を分類し、それぞれの弁別力を検証したものは非常に限定的であるため、本研究は読解テストにおける推論質問の可能性を検証した萌芽的な研究であると考えられる。しかしながら一方で、本研究には注目すべき限界点も存在する。

まず、調査2における手法上の課題点である。古典的テスト理論においては結果が受験者集団に依存するという標本依存性の問題と、テストの得点がテストに含まれる項目に依存する項目依存性の問題が指摘されている。そのため、本調査2においてもこうした限界点があることに留意する必要がある。本調査における結果をより強固なものにするためには、今後サンプル数を大きく増やし項目反応理論を用いた検証を行っていくことが求められる。

また、調査2ではTOEFLで使用されていたマテリアルを援用し、調査実施者が一部問題を作成した上で調査を行った。TOEFLで実際に使用されていた設問と新たに作成した設問の項目困難度に有意差がなかったため、結果への影響は大きくなかったことが示唆されるものの、設問の妥当性については今後検証の余地がある。

さらに調査2では各推論タイプにつき2項目を設定し検証を行なった。しかしながら、本調査ではこの問題数の少なさが結果に影響を与えた可能性も否定できない。実際に4.3.2のセクションからも分かるようにテストの内的信頼性の値は十分とは言えなかった。そのため、今後は本研究の知見を一般化するために、サンプル数だけでなく項目数も増やした検証を行うことが求められる。

注

- (1) 照応形とは、前出の情報に言及した代名詞や単語、フレーズのことを指す(Cook & O'Brien, 2017, p. 3)。
- (2) 本研究における TOEFL は、Educational Testing Service が出版している公式ガイドブックに掲載されている問題を指すため、本報告書では TOEFL iBT や TOEFL IPT 等の表記は使用していない。
- (3) 本研究におけるセンター試験は、問題入手や調査実施時期の都合により2020年度から実施されている大学入学共通テストではなく、2019年度まで実施されていた大学入学者選抜大学入試センター試験を指す。
- (4) 本調査1では英検のセクションの中でも4択の選択肢式問題に焦点を当てており、空所補充問題を調査対象としていない。そのため本調査のみをもって、英検では推論質問が出題されないと結論づけることはできない。

謝辞

本研究を発表するにあたり、今回このような大変貴重な機会を提供してくださいました公益財団法人 日本英語検定協会関係者の皆様方、また選考委員の先生方に厚く御礼を申し上げます。特に調査実施や報告書作成に関して有益なご助言・ご指導をいただきました竹内理先生には大変感謝申し上げます。また、筑波大学大学院の卯城祐司先生には、本研究の構想から、実施、執筆に至るまで懇切丁寧なご指導をいただきました。この場を借りて心より感謝申し上げます。さらに、調査実施上のアドバイス等親身に相談に乗っていただいた小木曾智子さん、佐々木大和さん、小室竜也さんをはじめとする先輩方、また同期の岡野紗綾加さん、佐藤連理さん、小林慎太郎さん、さらに後輩の皆様にも厚く御礼を申し上げます。最後になりますが、本研究を実施するにあたり多くのご助言とサポートをいただいたすべての皆様に深く感謝申し上げます。

参考文献(*は引用文献)

- * Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1459-1468. <https://doi.org/10.1037/0278-7393.21.6.1459>
- * Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1061-1070. <https://doi.org/10.1037/0278-7393.19.5.1061>
- * J. D. ブラウン. (1999). 『言語テストの基礎知識:正しい問題作成・評価のために』. 東京:大修館書店.
- * Committee of revising the JACET basic words. (Ed.). (2016). *The new JACET list of 8,000 basic words*. Tokyo: Kiriara Shoten.
- * Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, 51(1-2), 26-49. <https://doi.org/10.1080/0163853X.2013.855107>
- * Cook, A. E., & O'Brien, E. J. (2017). Fundamentals of inferencing during reading. *Language and Linguistics Compass*, 11(7), e12246. <https://doi.org/10.1111/lnc3.12246>
- * Educational Testing Service. (1999). *Test of English as a Foreign Language practice tests workbook (Vol. 2)*. Princeton: Educational Testing Service.
- * Educational Testing Service. (2021). *TOEFL iBT® Reading Section, ETS Home*. Retrieved September 2, 2021 from <https://www.ets.org/toefl/test-takers/ibt/about/content/reading/>
- * Educational Testing Service. (2020). *The Official Guide to the TOEFL test (5th edition)*. Princeton: Educational Testing Service.
- * Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395. <https://doi.org/10.1037/0033-295X.101.3.371>
- * Guéraud, S., Harmon, M. E., & Peracchi, K. A. (2005). Updating situation models: The memory-based contribution. *Discourse Processes*, 39(2-3), 243-263. <https://doi.org/10.1080/0163853X.2005.9651682>
- * Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological science*, 21(7), 895-900. <https://doi.org/10.1177/0956797610374742>
- * 伊東祐郎. (2011). 「項目バンクによって広がるテスト開発の可能性」. 『日本語教育』, 148, 57-71. https://doi.org/10.20721/nihongokyoiku.148.0_57
- * Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework*. Princeton: Educational Testing Service.
- * 川崎恵理子. (2014). 『文章理解の認知心理学』. 東京: 誠信書房.
- * Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182. <https://doi.org/10.1037/0033-295X.95.2.163>

参考文献(*は引用文献)

- * Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- * Lea, R. B., & Mulligan, E. J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(2), 303-317. <https://doi.org/10.1037/0278-7393.28.2.303>
- * McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, 4(1), 33-46. Retrieved from <https://iejee.com/index.php/IEJEE/article/view/212>
- * 文部科学省. (2019). 『高等学校学習指導要領 (平成30年告示)』. 京都: 東山書房.
- * 文部科学省. (2012). 「大学入試センター試験への移行」. https://www.mext.go.jp/b_menu/hakusho/html/others/detail/1318394.htm (2021年10月2日閲覧)
- * O'Brien, E. J., & Cook, A. E. (2016a). Coherence threshold and the continuity of processing: The RI-Val model of comprehension. *Discourse Processes*, 53(5-6) 326-338. <https://doi.org/10.1080/0163853X.2015.1123341>
- * O'Brien, E. J., & Cook, A. E. (2016b). Separating the activation, integration, and validation components of reading. In B. Ross (Ed.), *Psychology of learning and motivation* (pp. 249-276). New York: Academic Press.
- * 旺文社. (編). (2020). 『2020年版 英検2級 過去6回全問題集』. 東京: 旺文社.
- * 小野寺孝義・山本嘉一郎. (2004). 『SPSS 辞典: Base編』. 京都: ナカニシヤ出版.
- * 大下邦幸・広瀬由美子. (1997). 「読解力を高めるための効果的方法: 発問のあり方を中心として」. 『福井大学教育実践研究』, 22, 265-280.
- * 大友賢二. (1996). 『項目応答理論入門: 言語テスト・データの新しい分析』. 東京: 大修館書店.
- * 齊田智里. (2011). 「第2章 英語学力測定論」. 石川祥一・西田正・齊田智里 (編), 『テストと評価—4技能の測定から大学入試まで』 (pp. 30-58). 東京: 大修館書店.
- * 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.
- * 田中武夫. (2010). 「英文読解における推論発問をもとにした協同学習について」. 『山梨大学教育人間科学部紀要』, 11, 182-192.
- * 卯城祐司. (2009). 『リーディングの科学: 「読めたつもり」の謎を解く—』. 東京: 研究社.
- * Ushiro, Y., Hosoda, M., Nahatame, S., Mori, Y., Suzuki, K., Tada, G., ... & Mandokoro, R. (2018). Understanding protagonist, causal, and intentional links during EFL narrative reading. *ARELE: Annual Review of English Language Education in Japan*, 29, 81-96. https://doi.org/10.20581/arele.29.0_81
- * van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- * 渡部倫子. (2012). 「R を利用した日本語ブレースメントテストの分析: 古典的テスト理論による分析」. 『広島大学日本語教育研究』, 22, 25-32.
- * 横田敦子. (2017). 「物語文読解における主題に関わる推論発問の活用」. 『聖学院大学総合研究所 Newsletter』, 27(2), 50-55.
- * 吉田真美. (2001). 「テキストを超えた理解: 読みにおける推論過程」. 門田修平・野呂忠司, 『英語リーディングの認知メカニズム』 (pp.149-169). 東京: くろしお出版.