

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—

茨城県／筑波大学大学院博士課程在籍 清水 真紀

概要

本研究ではリーディングプロセスの観点から、英検、TOEFL、大学入試センター試験のテスト構成を明らかにし、さらに上位・下位レベルの処理の概念も加え、これら異なるレベルの処理間にどのような関係があるかについて調べた。

まず、各テストの質問を、内容の観点から「パラフレーズ質問」、「推論質問」、「テーマ質問」、「指示質問」、「語彙質問」、「文章構造質問」の6種類に分類した結果、TOEFLが多様な種類の質問から構成されているのに対し、英検は明示的な情報について問う質問が多く、また、センター試験はこれらの中間とも言える特徴をなしていることがわかった（調査1）。さらに、「推論質問」、「テーマ質問」、「文章構造質問」を上位レベルの処理、そして「パラフレーズ質問」を下位レベルの処理とした場合、上位レベルの処理は下位レベルの処理よりも難しいことが示された。また上位・下位レベルの処理の相関は中程度であることがわかった（調査2）。

これらの結果から、質問内容は項目困難度に影響を及ぼす重要な要因の1つであると考えられること、そして幅広い技能を含めたリーディング能力を測定する場合にはさまざまな内容の質問をテストに含めることの重要性が示唆された。

1 本研究の目的

読解研究分野では、読解を複数のプロセスに分けてとらえた上で、それらプロセスとプロセスの間に、あるいはまた各プロセスと読解力との間にどのような関係があるかといった研究（e.g., Ikeno, 2002; Nassaji, 2003; 野呂, 1999）や、読み手は読解中に認知リソースをどのように各プロセスに割り当ててい

るかといった研究（e.g., Horiba, 1996）が盛んに行われている。その読解プロセスについて、Grabe（2000）は次のように述べる。

The central components of reading processing include the following: orthographic processing, phonological coding, word recognition (lexical access), working memory activation, sentence parsing, propositional integration, propositional text-model formation, comprehension strategy use, inference making, text-model development, and the development of an appropriate situation model (or mental model). (p. 230)

正書法処理、音韻の符号化から状況モデルあるいはメンタルモデルの構築にまでわたっている。また、Pressley and Afflerbach（1995）は、読解プロセスに関する先行研究のレビューを行い、読解前のプロセス、読解中のプロセス、読解後のプロセスの3つに分けて整理した。さらに、読解中のプロセスは、「前から後ろへと読み進める」、「テキスト中の重要な情報を特定する」、「意図的に推論を生成する」、「テキスト中の異なる部分を統合する」、「解釈する」の5つのリストから構成され、各リストは15から30の先行研究をもとにまとめられた。

本研究では、英検、TOEFL、大学入試センター試験のリーディングテストの質問文を読解プロセスの観点から、明示情報の理解、推論、パラグラフや文章のテーマの理解、代名詞の照応などの理解、語彙の理解、文章構造の理解のカテゴリ別に分類した場合に、リーディングテストによってどのような構成の違いが見られるかを明らかにすることを第1の目的とする。

そして、第2の目的は、質問を上記タイプに分類し、さらに上位レベルの処理、下位レベルの処理といった観点(1.1.3)も加えた場合に、それらの項目困難度に違いは見られるだろうか、そして、それら上位レベルの処理・下位レベルの処理に分類された質問の得点間にはどのような関係があるかについて調べることである。

1.1 リーディングテストと質問タイプ

1.1.1 テスト細目表と測定するリーディング能力

日本英語検定協会は、「実用英語」を「日常の社会生活に必要な英語」と定義し、問題の内容・評価についての基本的な考え方を挙げている。そのうち、リーディングに関連するものを以下に挙げる。

- (1) 試験問題はすべて実用性(practical)ということに基調をおいて作成されます。
- (2) 題材は、日常の身の回りのことから、級が上がるにつれて、新聞・雑誌・テレビやインターネット・ビジネスなどで使われる海外の話題も多く採用されます。

「実用英語」、「実用性」という語が見られ、さらに題材についても「新聞・雑誌・テレビやインターネット・ビジネスなどで使われる」とあることから、素早くそして的確に必要な情報を取り出して理解する能力及び文章の意図を理解する能力に重点を置いてテスト作成がなされていることがうかがえる。したがって、文章中の明示的な情報の理解を問う質問、主題について問う質問などが多いことが予想される。

一方、TOEFLについては、Educational Testing Service(以下ETS)(1997)で“The questions test information that is stated in or implied by the passage, as well as knowledge of some of the specific words as they are used in the passage.”(p. 12)との記述が見られることから、明示的な情報に関する質問の他、読解の際に生成される推論についての質問、語彙に関する質問が見られることが予想される。

そして、大学入試センター試験については、『高等学校学習指導要領』(文部省、1999)の「外国語」の「リーディング」との関連で見えていく。「2 内容(1)言語活動」で、「ア まとまりのある文章を読んで、必要な情報を得たり、概要や要点をまとめたりする」、「イ まとまりのある文章を読んで、書き手の意

向などを理解し、それについて自分の考えなどをまとめたり、伝えたりする」、「ウ 物語文などを読んで、その感想などを話したり、書いたりする」、「エ 文章の内容や自分の解釈が聞き手に伝わるように音読する」の4つが挙げられている。本研究との関連でこれらを整理すると、上記アには文章中に明示されている情報の内容について問う質問、テーマに関する質問などが該当し、イの「書き手の意向などを理解」には文章のテーマについて問う質問などが該当すると思われる。そしてウは、Graesser, Singer, & Trabasso (1994)のemotion of readerなど推論に関する質問に近いと考えられる。

文章理解は、状況モデルの構築、すなわち、読解した文章からの情報と読み手の持っている背景知識からの関連情報を統合することで作られる一連の出来事、行動、状態の一貫した心的表象を形成することととらえられることから(e.g., 堀場, 2001; Kintsch, 1998; Zwaan & Brown, 1996)、リーディングテストにおいても、文章中に明示される情報を理解しているかということに加え、適切に推論を生成できているかについても評価することが重要だと思われる。

1.1.2 解答時間とテスト形式

各テストの解答時間についてもここで検討しておきたい。というのも、費やされる時間とどれくらい多様な問題を出題できるかということは関係があると思われるからである。英検2級の筆記試験(リスニングは除く)は75分間で大問が4つである。うち2つが読解問題となっていることから、単純計算によると40分近い時間が費やされることになる。一方、TOEFLは、paper-and-pencil版のTOEFL Practice Test(ETS, 1999)を参考にすると、Section 3のReading Comprehensionに55分間の時間が設定されている。そして大学入試センター試験については、6問のうち4問が読解問題である。全体の解答時間は80分であるが、駿台予備学校(2004)は65分間をこれらの解答時間の目安として挙げる。以上から、大学入試センター試験、TOEFL Practice Test、英検の順にリーディング問題に費やされる時間が長いことがわかる。

また、リーディングテストの形式としては、(a)クローズテスト、(b)短答式、(c)多肢選択式、(d)マッチング、(e)空所補充、(f)真偽テスト(Alderson, 2000)などがあるが、本研究では質問内容を吟味し

ながらタイプ分けしていくため、(c) 多肢選択式の Question-Answer 問題のみを対象とする。

1.1.3 上位レベル処理と下位レベル処理

読解をさまざまなプロセスからとらえる考え方については先に Grabe (2000) について述べたが、ここではしばしば各プロセスに対して言われる下位レベルの処理 (lower level processes)、上位レベルの処理 (higher level processes) について説明を加えたい。Grabe では、下位レベルの処理として次のものを挙げる。また、統語解析 (syntactic parsing) や命題の統合 (propositional integration) もこれら正書法処理、音韻処理と並んで下位レベルの処理と位置付けている。

Low-level processing can be discussed in terms of three sub-component processes: The recognition of orthographic structure (recognizing line forms, letter shapes, letter group patterns), the recognition of morpheme structure, and the processing of phonemic information. (p. 231)

一方、上位レベルの処理は “working with larger units of information and information contributed by the reader” (Grabe, 2000, p.233) とされ、テキストモデル、状況モデルの構築などもこれに含まれるとする。そしてまた、これらの構築に関連するものとして推論生成やストラテジーの使用が挙げられる。

本研究では、必ずしも各プロセスを二分する立場をとるわけではないものの、これまでの研究からも共通理解が得られていると思われるこの Grabe (2000) の分類に倣うことにする。したがって、本研究の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質問は、統語解析や命題の統合と大きくかかわっていると考えられるから、下位レベルの処理に分類する。

また、文章構造の理解、テーマの理解の2つについては Grabe では直接に言及されていないものの、Graesser et al. (1994) でテーマの理解は推論の1つの種類として分類されていることから、そしてまた、文章構造の理解は文章全体、つまり “larger units of information” にかかわることから、本研究では上位レベルの処理に分類した。

ここで、上位レベルの処理、下位レベルの処理の

関連を見た研究として Ikeno (2002) を挙げる。Ikeno は日本人大学生を対象に複数のテストを実施し、テキスト構造予測、ワーキングメモリ (L1とL2)、下位レベル処理の効率度 (number matching [番号照合], L2 word matching [L2単語照合], L2 lexical semantic judgment [L2語彙意味性判断], L2 grammaticality judgment [L2文法性判断], L2 sentence verification [L2文検証課題]) との関連を調べた。Grabe の分類によるとテキスト構造予測は上位レベルの処理と位置付けられる。

下位レベルの処理、そして上位レベルの処理との組み合わせに焦点を絞って結果を見ると、テキスト構造予測と L2 lexical semantic judgment、テキスト構造予測と L2 sentence verification は、それぞれ中程度及び中程度に近い相関 ($r = .412, p < .05$; $r = .397, p < .05$) があったものの、それ以外の3つの下位レベルの処理とテキスト構造予測には有意な相関が見られなかった。この Ikeno の結果から、上位レベルの処理と下位レベルの処理との相関はあるとしても中程度であることが示された。したがって、本研究のように、すでにリーディングテストで用いられている質問を内容によって上位レベルの処理、下位レベルの処理に分類した場合にも、それらの間の相関も中程度以下であることが予測される。

なお、この Ikeno (2002) の他に、処理レベルに着目した研究には野呂 (1999)、Nassaji (2003) がある。しかし、野呂は下位レベルの処理 (音韻処理と正書法処理) と読解との関連を見た研究であるため、また、Nassaji は上位レベルの処理・下位レベルの処理の分類に関して Grabe (2000) 及び本研究と異なるため、結果は省略する。

そしてまた、上位レベルの処理、下位レベルの処理の項目の困難度に関しては Davey (1988) の研究が挙げられる。Davey は質問タイプの1つとして「解答の情報の位置」というカテゴリを設け、質問の解答が文章中に明記されているか、いないかの分類を行った。そして、このカテゴリと「もっともらしさ」のカテゴリ (不正解の選択肢がもっともらしい程度) との間に $.56 (p < .05)$ と比較的高い相関が見られたと報告している。

つまり、質問の解答が文章中に明示されていない場合には不正解の選択肢が受験者にとってもっともらしく感じられる度合いが高くなり、したがって項目の困難度が上がると考えられる。このことから、

「文章から導き出される推論」(推論質問)、「文章構造の理解」(文章構造質問)、「テーマの理解」(テーマ質問)などの上位レベルの処理の質問群では、「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」(パラフレーズ質問。2.1.2) 下位レベルの処理の質問よりも項目困難度は高くなると予想される。

1.2 リサーチ・クエスチョンと仮説

本研究ではリサーチ・クエスチョンを1つ、仮説を2つ立てた。なお、リサーチ・クエスチョンは調査1で、仮説は調査2で検証した。

リサーチ・クエスチョン：英検、TOEFL、大学入試センター試験のリーディングセクションの問題を内容により分類した場合、テスト構成に違いは見られるか。

仮説1：TOEFLで上位レベルの処理(「上位」)に分類される質問の得点は、TOEFLで下位レベルの処理(「下位」)に分類される質問の得点よりも、そしてまた、英検で下位に分類される質問の得点よりも低い。つまり、上位に分類される質問のほうが下位に分類される質問よりも項目困難度が高い。

仮説2：TOEFLで上位に分類される質問の得点とTOEFLで下位に分類される質問の得点との相関、及び、TOEFLで上位に分類される質問の得点と英検で下位に分類される質問の得点との相関は、中程度以下である。

2 調査1

2.1 方法

2.1.1 使用したテスト

次のテストのリーディング問題について、質問内容の観点から分類を行った。

- (a) 2004年度第1回英検2級4A～4C(資料1。「英検」)：4択の多肢選択式12項目。なお、3Aと3B(計8項目)もこれらと並んでリーディング問題であったが、空所補充式で本研究の分類になじまなかったため対象外とした。
- (b) TOEFL Practice Test A, Section 3(資料2。

「TOEFL」。ETS, 1999)：4択の多肢選択式50項目。

- (c) 2004年度の大学入試センター試験第4問のBと第6問のA(資料3。「センター試験」)：4択の多肢選択式9項目。なお、今回の分類になじまない以下の項目は除いた。
 - (1) 第3問：7項目
 - (2) 第4問のA：1項目
 - (3) 第5問：5項目
 - (4) 第6問のB：3項目

2.1.2 手順

各項目は、英語教育学専攻の大学院生2人によって、次の(a)から(f)に分類された。複数のカテゴリにわたるような場合には最も焦点が当たっているカテゴリに分類した(Lumley, 1993)。さらに、(b)推論質問はGraesser et al. (1994)に基づき下位分類を行った(注1)。なお、(b1)の原因を問う質問について、その因果関係がbecauseなどの語によって明示的に標示されている場合は(a)パラフレーズ質問に、それ以外は(b1)のcausal antecedentに分類した。

- (a) 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。語の重なりが多く見られたり、質問に解答する場合に必要なキーワードが含まれている場合はこれに該当する。明示情報に関する質問。
- (b) 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。下位分類は(b1)から(b8)のとおり。

- (b1) causal antecedent (因果的先行詞)：

“on a causal chain (bridge) between the current explicit action, event, or state and the previous passage context”と定義され、例えば“*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“The man was careless and mis-aimed his fork.”という推論のこと(p. 375)。

- (b2) superordinate goal (上位ゴール)：

“on a goal that motivates an agent’s intentional

action”と定義され、例えば“A Man to Whom Time was Money, and who was bolting his breakfast in order to catch a train...”という部分を読んだときに“The man wanted to get to work and earn money.”という推論のこと (p. 375)。

(b3) character emotional reaction (登場人物の感情的反応) :

“an emotion experienced by a character, caused by or in response to an event or action” (p. 375) と定義される。

(b4) causal consequence (因果的結末) :

“on a forecasted causal chain, including physical events and new plans of agents” と定義され、例えば “...on removing the fork the eye came with it” という部分を読んだときに “The man became blind in his right eye.” という推論のこと。ただし、登場人物の感情的反応は含まれない。その場合は、登場人物の感情的反応に分類する (本研究では b3 のカテゴリ) (p. 375)。

(b5) instrument (道具) :

例えば “...the Man to Whom Time was Money had to sustain life by fishing from the end of a wharf” という部分を読んだときに, “The man used a rod and reel (to fish).” という推論のこと (p. 375)。

(b6) subordinate goal/action (下位ゴール/行為) :

“a goal, plan, or action that specifies how an agent's action is achieved” と定義され、例えば, “...who was bolting his breakfast” という部分を読んだときに, “The man grasped his fork and moved it toward his mouth.” という推論のこと (p. 375)。

(b7) state (状態) :

“an ongoing state, from the time frame of the text, that is not causally related to the story plot” と定義される。“an agent's traits, knowledge, and beliefs; the properties of objects and concepts; and the spatial location of entities” もこれに含まれる (p. 375)。

(b8) emotion of reader (読み手の感情) :

“the emotion that the reader experiences when reading a text” と定義され、例えば “...on removing the fork the eye came with it” という部分を読んだときに “The reader is disgusted.” という推論のこと (p. 375)。

(c) 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 (注2)。

(d) 指示質問：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 (注3)。

(e) 語彙質問：語彙の意味について問う質問。文脈の情報がなくとも正解することができるもの (注4)。

(f) 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。

2.2 結果と考察

評価者の一致度は85%と高かった。不一致はディスカッションにより解決した。最も一致が見られなかったカテゴリは (b7) の state に関する推論質問で、その原因の1つに、因果関係があれば (b1) に、ない場合には (b7) に分類されるという判断の難しさがあったようだ。一方、(c) テーマ質問、(d) 指示質問、(e) 語彙質問、(f) 文章構造質問についてはすべて一致した (表1)。

さらに、カテゴリごとに、分類された項目数と全体に対する割合を見たのが表2である。

以上、分類を行った結果から、リサーチ・クエスチョン、「英検、TOEFL、大学入試センター試験のリーディングセクションの問題を内容により分類した場合、テスト構成に違いは見られるか」については、以下のことが明らかになった。

まず、TOEFL については、(a) から (f) のすべてのカテゴリに質問がわたっていること、そして各カテゴリに少なくとも8%の質問が含まれていること、中でも多かったのはパラフレーズ質問の32%だったということである。つまり、TOEFL はさまざまな内容について問う構成がとられていた。これは、1.1 で予測したものと一致していた。さらに、(c) テーマ質問 (e.g., “What is the author's main point in the first

■ 表 1 : 英検, TOEFL, 大学入試センター試験の質問タイプの分類

	英 検	TOEFL	センター試験
(a) パラフレーズ質問			
	項目35 (4A) 項目36 (4A) 項目37 (4B) 項目38 (4B) 項目39 (4B) 項目40 (4B) 項目41 (4C) 項目42 (4C) 項目43 (4C) 項目44 (4C) 項目45 (4C)	項目 1 (文章 1) 項目 4 (文章 1) 項目 7 (文章 1) 項目12 (文章 2) 項目15 (文章 2) 項目23 (文章 3) 項目27 (文章 3) 項目28 (文章 3) 項目29 (文章 3) 項目31 (文章 3) 項目32 (文章 3) 項目33 (文章 3) 項目36 (文章 4) 項目38 (文章 4) 項目40 (文章 4) 項目50 (文章 5)	項目34 (第 4 問) 項目35 (第 4 問) 項目36 (第 4 問) 項目37 (第 4 問)
(b) 推論質問			
(b1) causal antecedent	項目34 (4A)	項目43 (文章 5)	項目46 (第 6 問)
(b2) superordinate goal			項目44 (第 6 問) 項目45 (第 6 問)
(b3) character emotional reaction			項目43 (第 6 問) 項目47 (第 6 問)
(b4) causal consequence			
(b5) instrument			
(b6) subordinate goal/action			
(b7) state		項目 5 (文章 1) 項目13 (文章 2) 項目18 (文章 2) 項目21 (文章 2) 項目42 (文章 5) 項目46 (文章 5)	
(b8) emotion of reader			
(c) テーマ質問			
		項目11 (文章 2) 項目22 (文章 3) 項目35 (文章 4) 項目48 (文章 5)	
(d) 指示質問			
		項目 3 (文章 1) 項目 8 (文章 1) 項目14 (文章 2) 項目26 (文章 3) 項目37 (文章 4) 項目45 (文章 5)	
(e) 語彙質問			
		項目 2 (文章 1) 項目 6 (文章 1) 項目16 (文章 2) 項目17 (文章 2) 項目19 (文章 2) 項目24 (文章 3) 項目25 (文章 3) 項目30 (文章 3) 項目39 (文章 4) 項目41 (文章 5) 項目44 (文章 5) 項目47 (文章 5) 項目49 (文章 5)	
(f) 文章構造質問			
		項目 9 (文章 1) 項目10 (文章 1) 項目20 (文章 2) 項目34 (文章 3)	

(注) 下線部は, 評価者が独立して分類した場合に判断が分かれディスカッション後に一致した項目。

■ 表 2：各テストのタイプごとの項目数と割合

	英 検	TOEFL	センター試験
(a) パラフレーズ質問	11 (92%)	16 (32%)	4 (44%)
(b) 推論質問	1 (8%)	7 (14%)	5 (56%)
(c) テーマ質問	0 (0%)	4 (8%)	0 (0%)
(d) 指示質問	0 (0%)	6 (12%)	0 (0%)
(e) 語彙質問	0 (0%)	13 (26%)	0 (0%)
(f) 文章構造質問	0 (0%)	4 (8%)	0 (0%)
合計	12 (100%)	50 (100%)	9 (100%)

paragraph?”), (d) 指示質問 (e.g., “The word ‘this’ in line 8 refers to…”), (e) 語彙質問 (e.g., “The phrase ‘served as’ in line 6 is closest in meaning to…”), (f) 文章構造質問 (e.g., “Which of the following best describes the organization of the passage?”) は、質問の表現から、その分類が可能であり、分類の評価者間信頼性は100%の一致を見た。

一方で、英検の質問の多くは明示的な情報について問う質問であった。実用英語を基調にした出題であることから、一部、予測された結果ではあったものの、主題について問う質問も出題されるとの予測は当たらなかった。ただし、これに関しては以下2点を考慮する必要がある。

- (1) 解答時間との関係から、多様な質問内容を含めることに制約があったのかもしれない。
- (2) 空所補充式は分類対象としなかったため、この結果のみで主題を問う問題が出題されていなかったと結論付けることはできない。Yamashita (2003) は、空所補充形式の問題を解く際の受験者の発話プロトコルデータを収集した結果、テキストレベルの理解を反映したものであったことを報告している。

そしてセンター試験は、カテゴリは英検と同じく2つであったが、項目数より、TOEFL と英検の2つのテストの中間に位置付けることができる。また、推論に分類された質問に着目すると、TOEFL が (b7) state に偏っているのに対し、センター試験には (b1) causal antecedent, (b2) superordinate goal, (b3) character emotional reaction などの質問が見られた。特に (b3) の質問が出題されていたことは、1つに、今回のセンター試験の文章が物語文だったことと関係していると考えられる。というのも、「登場人物があ

る出来事・行為に対してどのような感情を抱いただろうか」と問うことは、説明文の場合でも十分に考えられることだが、物語文でより多くなると考えられるからである。

ただ、このセンター試験の分類結果についても、一部の問題を対象外としたことを考慮する必要がある (1.1.2)。対象外とした問題の中には、適切な談話標識を選ぶ問題や文の並び替えを行わせる問題があった。

続く調査2では、テスト構成に大きな違いが見られた英検と TOEFL の一部を実施し、さらに質問タイプに処理レベルの視点 (1.1.3) を加えて考察する。本研究では「(a) パラフレーズ質問」を下位レベル処理、「(b) 推論質問」、「(c) テーマ質問」、「(f) 文章構造質問」を上位レベル処理と位置付けた。また、分析する上で、問題数を考慮し、上位レベル処理のカテゴリはまとめて得点化して「(a) パラフレーズ質問」と比較した。

3 調査2

3.1 方法

3.1.1 対象者

日本人大学生、大学院生計204名が調査に参加した。ただし、使用したテストに不備のあった対象者4名を除く200名のデータを分析した。専攻は、生物学、人文学、心理学、国際関係学、情報学などさまざまであった。

3.1.2 使用したテスト

次の2種類のテストを使用した。

- (1) 2004年度第1回英検2級から4A (3項目)、4C (5項目) と2004年度第1回英検準1級から大問3 (3項目) (「英検」)：4択の多肢選択式計11項目。英検2級については、調査1でタイプ別に分類した。英検準1級の3項目については、あらためて同じ評価者2人で分類を行った結果、3項目とも「パラフレーズ質問」で一致した (表3)。
- (2) TOEFL Practice Test A, Section 3 (「TOEFL」。ETS, 1999)：4択の多肢選択式34項目。調査1で分類した50項目 (文章5つ) のうちの一部 (文章3つ) を対象とした (表4)。

■ 表3：調査2で使用した英検の質問の分類結果

	項目数	全体に対する割合	項目
(a) パラフレーズ質問	10	91%	英検2級から：項目35 (4A), 項目36 (4A), 項目41 (4C), 項目42 (4C), 項目43 (4C), 項目44 (4C), 項目45 (4C) 英検準1級から：項目35 (大問3), 項目36 (大問3), 項目37 (大問3)
(b) 推論質問	1	9%	英検2級から：項目34 (4A)
(c) テーマ質問	0	0%	—
(d) 指示質問	0	0%	—
(e) 語彙質問	0	0%	—
(f) 文章構造質問	0	0%	—
合計	11	100%	

■ 表4：調査2で使用した TOEFL の質問の分類結果（調査1より）

	項目数	全体に対する割合	項目
(a) パラフレーズ質問	12	34%	項目1 (文章1), 項目4 (文章1), 項目7 (文章1), 項目12 (文章2), 項目15 (文章2), 項目23 (文章3), 項目27 (文章3), 項目28 (文章3), 項目29 (文章3), 項目31 (文章3), 項目32 (文章3), 項目33 (文章3)
(b) 推論質問*	4	12%	項目5 (文章1), 項目13 (文章2), 項目18 (文章2), 項目21 (文章2)
(c) テーマ質問*	2	6%	項目11 (文章2), 項目22 (文章3)
(d) 指示質問	4	12%	項目3 (文章1), 項目8 (文章1), 項目14 (文章2), 項目26 (文章3)
(e) 語彙質問	8	24%	項目2 (文章1), 項目6 (文章1), 項目16 (文章2), 項目17 (文章2), 項目19 (文章2), 項目24 (文章3), 項目25 (文章3), 項目30 (文章3)
(f) 文章構造質問*	4	12%	項目9 (文章1), 項目10 (文章1), 項目20 (文章2), 項目34 (文章3)
合計	34	100%	

(注) *のカテゴリは本研究で上位レベルの処理とした。合計10項目 (29%)。

3.1.3 手順

一部の学習者に対しては授業時間内に実施したが、多くは授業時間外に実施した。TOEFL、英検の順番で、それぞれ33分間、25分間で実施した^(注5)。なお、実際の英検及び TOEFL の指示に従って、それぞれ問題用紙に書き込み可、不可とした。

3.1.4 データ分析

ピアソンの積率相関分析、反復測定による一元配置の分散分析を行った。統計処理はすべて SPSS 10.0 for Windows を使用した。

3.2 結果と考察

3.2.1 記述統計

本調査で実施した英検11項目の平均得点率は表5のとおり .667点 ($SD = 0.205$) であった。全体で11項目のテスト得点の α 係数は .635 であった。また、このうちパラフレーズ質問に分類された10項目の平

均得点率は .645点 ($SD = 0.215$) であった。

■ 表5：英検全体及びパラフレーズ質問に分類された項目の記述統計

	<i>M</i>	<i>SD</i>
全体 (11項目)	.667	0.205
パラフレーズ質問 (10項目)	.645	0.215

(注) 項目数で割って得点率に換算したもの。したがって満点は1.000。 $N = 200$ 。

一方、TOFEL について、34項目全体の平均得点率は .534点 ($SD = 0.284$) であった。質問タイプ別の結果は表6のとおりであった。全体で34項目のテスト得点の α 係数は .778 であった。

3.2.2 分散分析の結果

英検パラフレーズ質問 ($M = .645$, $SD = 0.215$), TOEFL パラフレーズ質問 ($M = .554$, $SD = 0.207$),

■ 表 6 : TOEFL 質問タイプ別記述統計

	項目数	M	SD
(a) パラフレーズ質問	12	.554	0.207
(b) 推論質問	4	.423	0.281
(c) テーマ質問	2	.548	0.357
(d) 指示質問	4	.714	0.252
(e) 語彙質問	8	.509	0.225
(f) 文章構造質問	4	.459	0.267
上位レベル処理(b+c+f)	10	.462	0.208
全体	34	.534	0.284

(注) 各カテゴリの項目数が異なっていたため、すべて得点率に換算した。したがって満点は1.000。

TOEFL 上位レベル処理の質問 ($M = .462$, $SD = 0.208$) の平均値の間に差があるか調べるため、一元配置の分散分析 (反復測定) を行った。なお、本調査で実施した英検11項目のうち、上位レベル処理に分類された質問はわずか1項目と数が少なかったため、英検上位レベルの処理という群は設けなかった。分析の結果、これら3つの平均値のいずれかの間に有意差があることが示された ($F(2, 398) = 70.579$, $p < .001$)。そして、eta squared (η^2) による効果量は0.355, と大きかった。

■ 表 7 : 一元配置の分散分析

Source	df	SS	MS	F	η^2
質問タイプ	2	3.331	1.665	70.579 ***	0.355
Error	398	9.391	0.024		

(質問タイプ)

(注) *** $p < .001$.

さらに、チューキー HSD の多重比較を行ったところ、5%水準でいずれの組み合わせにおいても差があることが示された。つまり、TOEFL 上位レベルの処理の質問、TOEFL パラフレーズ質問、英検パラフレーズ質問の順に有意に項目困難度が高いことが確認された。ここで重要なのは、TOEFL 上位レベルの処理の質問が、TOEFL と英検いずれのテストのパラフレーズ質問よりも困難度が高かったことである。

上位レベルの処理の質問は、解答が文章中に明示されていないことが多いことから不正解の選択肢が

よりもっともらしくなり、したがって困難度が上がるだろうといった予測に基づく仮説1「TOEFL で上位レベルの処理に分類される質問の得点は、TOEFL で下位レベルの処理に分類される質問の得点よりも、そしてまた、英検で下位に分類される質問の得点よりも低い」は、支持された。

なお、英検と TOEFL でパラフレーズ質問という同じカテゴリに分類された質問同士であっても項目困難度に差が見られたことについては、解答時間、読解文章の難易度、受験中にメモを取ることが許されていたか、いなかったかなどのいくつかの要因が関係していたかもしれない。今後、これらの要因を統制した研究が期待される。

また先行研究の Davey (1988) では、リーディング問題の項目困難度に比較的大きな影響を与える要因として、「解答の情報の位置」と「stem の長さ」が示された。「解答の情報の位置」とは、本研究で言う質問内容に相当するもので、この要因によって項目困難度の分散の12~27%が説明できることを明らかにした。一方、「stem の長さ」の要因では14~15%が説明できることを明らかにした。

また、Freedle and Kostin (1993) では、TOEFL のリーディング項目の困難度に影響を与える要因として、指示語の数、語彙、関連する情報が文章中で占める位置など、12の要因を挙げ、このうち11が項目困難度と関連し、さらにテキストと正解の選択肢の語彙的な重なり具合、文の長さ、パラグラフの長さ、修辞構造、否定の使用、指示語の使用、文章の長さの7つの要因で分散の58%を説明できるという結果が報告された。

これらからも、本研究で扱った質問の内容という要因は、項目困難度に影響を与える要因のうちの1つにすぎない。しかし、本調査結果からも、また Davey の研究結果からも、この質問内容は項目困難度との関係で特に重要な要因の1つであると言える。

3.2.3 相関分析の結果

TOEFL でパラフレーズ質問に分類された12項目、英検でパラフレーズ質問に分類された10項目、TOEFLで上位レベルの処理に分類された(推論質問、テーマ質問、文章構造質問)10項目、それぞれの得点間の相関は、いずれも中程度だった(表8)。

まず TOEFL で、上位レベルの処理の質問と下位レベルの処理の質問(パラフレーズ質問)の得点の

■ 表 8 : TOEFL パラフレーズ質問・英検パラフレーズ質問・TOEFL 上位レベル処理の質問の相関

	1	2	3
1. TOEFLパラフレーズ質問(12項目)	--		
2. 英検パラフレーズ質問(10項目)	.464*** (.348-.566)	--	
3. TOEFL上位レベルの処理(10項目)	.465*** (.349-.567)	.479*** (.364-.579)	--

(注) *** $p < .001$. $N = 200$. 括弧内の値は信頼係数95%区間の信頼限界。

間の相関係数は、 $r = .465$ ($p < .001$) であった。一方、TOEFL の上位レベルの処理の質問と英検の下位レベルの処理の質問（パラフレーズ質問）の得点の相関係数は、 $r = .479$ ($p < .001$) であった。このことから、仮説2「TOEFL で上位に分類される質問の得点と TOEFL で下位に分類される質問の得点との相関、及び、TOEFL で上位に分類される質問の得点と英検で下位に分類される質問の得点との相関は、中程度以下である」については、中程度の相関であったことから、支持されたと言える。

さらに、説明率によると、TOEFL の上位レベル処理の質問と TOEFL パラフレーズ質問では一方の要因によって他の要因の分散を説明できるのは約22%であり、そしてまた、TOEFL の上位レベル処理の質問と英検パラフレーズ質問では約23%である。

本研究で分類対象とした英検の問題の多くがこのパラフレーズ質問に集中していたこと（表2）を考えると、これらの値はより重要な意味を持つ。

リーディングの中でもある特定の能力のみを測定する場合は別として、幅広い技能を含むリーディング能力を測定したい場合には、質問文や選択肢をパラフレーズすることで正解を導き出せる質問の他、適切に推論を生成しているか、全体的な文章の構造を理解しているか、また、パラグラフや文章全体のテーマを理解しているかといった上位レベルの処理を確認する質問もテストの中に含めていくことの重要性を示している。

この他、同じタイプである TOEFL パラフレーズ質問と英検パラフレーズ質問で $r = .464$ ($p < .001$) という中程度の相関しか見られなかったことは、予想とは異なる結果だった。説明率によると残りの約78%については他の変数によって説明されることを意味する。これについては、3.2.2 でも述べたように、質問内容以外の要因による影響があったのかもしれない。今後さらに研究を行う必要がある。

なお、本研究では、下位レベルの処理を TOEFL と英検で数多く見られたパラフレーズ質問ととらえ

て分析及び解釈を行ったが、Grabe (2000) も指摘するように、下位レベルの処理はこれにとどまらず正書法処理や音韻処理なども含む幅広い概念である。したがって、今後、各処理間の関係をより詳細に検証する場合には新たな研究が必要である。

また、調査2の限界点としては、TOEFL パラフレーズ質問、英検パラフレーズ質問、TOEFL 上位レベル処理の質問の項目数にばらつきが見られたことである。項目数によっても平均値、標準偏差、相関係数などが異なってくることを考えると、今後はそれらを考慮したデザインで調査を実施し、結果を確認する必要があると思われる。

4 結論

本研究では、第1に英検、TOEFL、大学入試センター試験のリーディング問題の特に多肢選択式の Question-Answer 問題を対象にして、内容の観点から分類した場合にテスト間にどのような構成の違いが見られるかについて調べた。カテゴリはパラフレーズ質問、推論質問、テーマ質問、指示質問、語彙質問、文章構造質問の6種類を用いた。分類の結果、3種類のテストともパラフレーズ質問を多く含んでいたものの、パラフレーズ質問が全体に占める割合、そしてこれら6種類のカテゴリへの散らばり具合を見た場合には、テストによる違いが見られることが明らかになった。さらに、推論質問を多く含むテストであっても、さらに下位分類を行うと違いが見られることがわかった。なお、今回は手法上の理由から、リーディング問題であっても空所補充形式や真偽テストなどは対象にできなかった。今後もし可能であればそういった項目も含めることで、各リーディングテストの構成の違いといったものをより明らかにできるとと思われる。

第2に、上記カテゴリのうち、推論質問、テーマ質問、文章構造質問を上位レベルの処理、そしてパ

ラフレーズ質問を下位レベルの処理とし、それぞれの処理の質問の項目困難度について、また、これら得点間の相関について調べた。まず、項目困難度の結果から述べると、上位レベルの処理とされた質問の項目困難度は、下位レベルの処理とされた質問の項目困難度よりも高いことが示された。したがってこの質問内容の要因は、項目困難度に影響を及ぼす重要な要因の1つであると考えられる。そして、上位レベルの処理と下位レベルの処理の質問の得点間には、中程度の相関があることがわかった。

これについては、低い値であるとは言えないものの、80%弱は当要因では説明できないことを示すため、もし特定のリーディングの下位技能に限らず幅広い技能も含めたリーディング能力というのをテストで測定したい場合には、テスト作成上、さまざまな内容の質問を含めていくことの重要性を示してい

ると言えるだろう。

謝 辞

本研究の機会を与えてくださった(財)日本英語検定協会の皆様、選考委員の先生方に厚く御礼申し上げます。特に助言者の池田央先生には、中間報告や最終原稿に対して貴重なご助言をいただきました。また、筑波大学の卯城祐司先生、望月昭彦先生、同大学院博士課程の土方裕子さんに深く感謝申し上げます。また、本調査実施にあたっては、群馬県立太田高等学校の堀本一郎先生、岡島誠一郎先生、群馬県立伊勢崎東高等学校の多賀谷弘孝先生、群馬県立高崎高等学校の田中雅徳先生、群馬県立高崎北高等学校の中島剛先生、三重県立上野高等学校の岡本理佳先生をはじめ、多くの先生方に協力いただきました。ありがとうございました。

注

- (1) Graesser et al. の case structure role assignment (格構造役割付与) は、文法知識の測定になるので、今回は扱わなかった。
- (2) Graesser et al. は推論の一部と考えるが、本研究では別のカテゴリとした。さらに、この thematic の推論と Graesser et al. の言う author's intent (著者の意図) についての推論は区別するのが難しいと考えられたため、本研究ではこれらをまとめて(c)のテーマ質問に分類した。
- (3) Graesser et al. は推論の一部と考えるが、本研究では別のカテゴリとした。
- (4) Graesser et al. の instantiation of noun category (名詞カテゴリの例示) もこれに含めた。

- (5) 実施時間が長くなかったため、実施順序についてのカウンターバランスはとらなかった。ただし、TOEFL は速読が要求されることから、これらすべての項目に解答できない学習者も出てくることが事前に予想されたため、3つのバージョンを作成してカウンターバランスをとった。1つ目のバージョン (n=63) は文章1→文章2→文章3、2つ目のバージョン (n=68) は文章2→文章3→文章1、3つ目のバージョン (n=69) は文章3→文章1→文章2の順で配列された。必ずしもこの配列順序どおりに解答する必要はないが、各対象者にはいずれかのバージョンの冊子が配布された。

参考文献 (*は引用文献)

- * Alderson, J.C. (2000). *Assessing reading*. Cambridge University Press.
- * Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- * Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *Journal of Experimental Education*, 56, 67-76.
- * Educational Testing Service. (1997). *TOEFL test & manual* (1997 ed.). Retrieved November 30, 2004, from <http://ftp.ets.org/pub/toefl/678096.pdf>.
- * Educational Testing Service. (1999). *Test of English as a Foreign Language practice tests workbook* (Vol. 2). Princeton, NJ: Author.
- * Freedle, R., & Kostin, I. (1993). The prediction of TOEFL reading item difficulty: Implications for construct validity. *Language Testing*, 10, 133-170.

- * Grabe, W. (2000). Reading research and its implications for reading assessment. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida: Studies in Language Testing 9* (pp. 226-262). Cambridge University Press.
- * Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- * Horiba, Y. (1996). Comprehension processes in L2 reading. *Studies in Second Language Acquisition*, 18, 433-473.
- * 堀場裕紀江. (2001). L2リーディング研究の課題と可能性. 『言語科学研究』, 7, 43-63.
- * Ikeno, O. (2002). Text structure prediction in L2 reading and working memory. *JACET Bulletin*, 35,

- 105-116.
- * Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
 - * Lumley, T. (1993). The notion of subskills in reading comprehension tests: An EAP example. *Language Testing*, 10, 211-234.
 - * 文部省. (1999). 『高等学校学習指導要領』. 東京：大蔵省印刷局.
 - * Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *Modern Language Journal*, 87, 261-276.
 - * 日本英語検定協会. (2004). 「STEP 英検」. Retrieved November 30, 2004, from <http://www.eiken.or.jp/index.html>.
 - * 野呂忠司. (1999). 下位レベルの処理技能と外国語としての英語の読解力. 『中部地区英語教育学会』, 29, 23-30.
 - * Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.
 - * Statistical Package for the Social Sciences. (1999). SPSS for Windows (Version 10.0) [Computer software]. Chicago: Author.
 - * 駿台予備学校 (編). (2004). 『大学入試センター試験過去問題集英語』. 東京：駿台予備学校.
 - * Yamashita, J. (2003). Processes of taking a gap-filling test: comparison of skilled and less skilled EFL learners. *Language Testing*, 20, 267-293.
 - * Zwaan, R.A., & Brown, C.M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes*, 21, 289-327.

資料1：2004年度第1回英検2級（4A）から抜粋

*は正解。

To: Basil Black <basil-32fds@cordline.net>
From: Mary Finnegan <finnegan-m@sunnyhols.co.uk>
Date: June 13, 2004
Subject: Re: Vacation at the Park Hotel

Dear Mr. Black,

I was sorry to hear that you were not satisfied with your recent vacation to Portugal with Sunny Tours. You mentioned in your e-mail that your flight left London five hours late because of poor weather conditions in the south of England. While I understand that this must have been frustrating, I'm afraid we cannot offer a cash refund for the time you lost, as you requested. You'll find that our brochure clearly states that Sunny Tours does not compensate customers for airport delays.

I appreciate your comments about the Park Hotel. Feedback from customers about the hotels we deal with is always useful. I telephoned Mr. Figo, the manager of the Park Hotel, to pass on your concerns about their catering. He told me that their regular chef was on vacation during the period of your stay and a less experienced chef was on duty. That may explain the poor quality of the meals you had there. We will be sending one of our staff to the Park Hotel next month. If we find that the service provided there does not meet our normal high standards, we will remove the hotel from our brochure.

I hope the experiences you had on this occasion will not stop you from traveling with Sunny Tours again in the future.

Yours sincerely,

Mary Finnegan

Customer Relations Manager

Sunny Tours

(34) What is one reason that Mr. Black was not satisfied?

1. He could not stay at the Park Hotel.
2. His flight from England was delayed.*
3. The refund he received was too small.
4. The weather was bad in Portugal.

(35) Mary Finnegan

1. got an e-mail from the hotel's regular chef.
2. apologized for the regular chef's lack of experience.

3. told the hotel manager about Mr. Black's complaints.*
4. asked Mr. Black for some feedback about his vacation.

(36) Next month,

1. Sunny Tours will check the service at the hotel.*
2. Mr. Black will book another vacation at the same hotel.
3. Mr. Figo will return from his vacation in Portugal.

資料 2 : TOEFL Practice Test A の Section 3 (文章1) から抜粋

"Wow, I suppose that I'm a double winner: I got a scholarship and made a friend, too."

"You're not the only one," Angela remarked.

問1 How did Kate feel after Angela joined the swimming club?

- Line
(5)
1. Happy.
 2. Excited.
 3. Sad.
 4. Anxious.*

問2 Why did the coach talk to Kate before the trial races?

- (10)
1. He thought she should support Angela.
 2. He wanted to encourage her to swim better.*
 3. He wanted her to be afraid of Angela.
 4. He thought she should support Angela.

問3 Why did Kate want to help Angela?

- (15)
1. She wanted Angela to win.
 2. The coach insisted that she do so.
 3. Angela was a newcomer.
 4. She understood how Angela felt.*

(20) 問4 Why was Kate disappointed immediately after the final race?

1. She thought she would not be going to the Nationals.*
2. She thought Angela had lost.

3. She had decided to give up swimming.

4. She found that Angela was not her friend.

*は正解。

問5 What did Angela mean when she said, "You're not the only one"?

1. She knew that both she and Kate had won the race.
2. She believed that only Kate would get a scholarship.
3. She thought that she had made a friend, too.*
4. She guessed that she would be given more than one scholarship.

The growth of cities, the construction of hundreds of new factories, and the spread of railroads in the United States before 1850 had increased the need for better illumination. But the lighting in American homes had improved very little over that of ancient times. Through the colonial period, homes were lit with tallow candles or with a lamp of the kind used in ancient Rome — a dish of fish oil or other animal or vegetable oil in which a twisted rag served as a wick. Some people used lard, but they had to heat charcoal underneath to keep it soft and burnable. The sperm whale provided a superior burning oil, but this was expensive. In 1830 a new substance called "camphene" was patented, and it proved to be an excellent illuminant. But while camphene gave a bright light it

too remained expensive, had an unpleasant odor, and also was dangerously explosive.

Between 1830 and 1850 it seemed that the only hope for cheaper illumination in the United States was in the wider use of gas. In the 1840's American gas manufacturers adopted improved British techniques for producing illuminating gas from coal. But the expense of piping gas to the consumer remained so high that until midcentury gaslighting was feasible only in urban areas, and only for public buildings or for the wealthy.

In 1854 a Canadian doctor, Abraham Gesner, patented a process for distilling a

pitchlike mineral found in New Brunswick and Nova Scotia that produced illuminating gas and an oil that he called "kerosene" (from "keros," the Greek word for wax, and "ene" because it resembled camphene). Kerosene, though cheaper than camphene, had an unpleasant odor, and Gesner never made his fortune from it. But Gesner had aroused a new hope for making an illuminating oil from a product coming out of North American mines.

1. Which of the following is NOT mentioned as a reason why better lighting had become necessary

資料3：2004年度大学入試センター試験（第6問 Part A）から抜粋

4. Mary Finnegan will send a brochure to Mr. Black.

*は正解。

My confidence as a swimmer started to disappear the day Angela moved to our small town. At the time, some members of the town's swimming club, myself included, were preparing for the National Championships, which were just six months away. I had always been the best, and everyone thought that I would be chosen for the relay race. But now I had competition. There was only one place for the butterfly on the relay team, and we both wanted it.

For two weeks it was awful. Angela was always the star. She was faster than I, and her form was better, too. I was jealous and scared. My chances of being selected were disappearing fast. My fear caused me to be unfriendly to Angela. I refused to speak to her and never said anything good about her.

One day, however, our coach called me over and said, "Kate, I've got something to say to you. Your attitude is hurting your performance. I know you can change that. I'd like you to think about it."

When I arrived at the pool the next morning, I thought about what he had said as I was going through my warm-up. Angela and I were going to compete that morning, and only eight girls would enter the finals.

My thoughts were interrupted when a voice said, "Nervous?" It was Angela. "I don't like to talk before a race," I replied coldly.

"I get nervous, too," she said. Her voice didn't have the anger of mine, which surprised me quite a bit.

Angela and I competed in our separate trial races. I jumped into the pool and swam like a flying fish. My mind was clear, and I could think about only one thing: swimming well.

When the races were over, the judges announced that both of us were among the lucky eight who would be competing in the finals. Despite this good news, I noticed that Angela was sitting sadly alone. This puzzled me, but I thought that I knew how she felt and tried to be friendly to her.

"I don't talk before races, but I do talk after them. Sometimes it helps," I said.

Angela was silent for a while, but then she said, "I'm great in practices, but in competitions I just can't do well. It's like this all the time. I'm so worried about the finals."

Now I felt really bad. I realized how horrible I had been to Angela. I wanted to help her. I wanted to show her that I was sorry for my behavior of the past two weeks.

"Listen, I have an idea," I said. "Why don't we help each other prepare for the final race? We have two weeks to work on things."

"Good idea," said Angela.

For the next two weeks Angela and I worked together. I taught her how to deal with stress and how to train her muscles. She helped me with my form, and at the end of those two weeks we were the best of friends and respected each other as swimmers.

The day of the final race came and when the starter pistol was fired, I swam off like a dolphin. I thought about nothing but winning, but just before I reached the finish line, I thought of Angela and looked over into her lane.

As I was climbing out of the pool, I said to myself, "Oh no, what have I done?" I thought that I had dropped behind Angela and lost the race when I looked in her direction.

While we were waiting for the official announcement, the coach came running over to us. "Congratulations, girls! It was close, but Angela has won, and so have you, Kate!"

"What do you mean?" I asked.

"Well, Angela will be in the relay, but you swam so fast, Kate, that you, too, have won."

"Really?" I screamed. "I don't understand."

"I'll explain. The 21st Century Swimsuit Company is giving both of you their Future Swimmers Scholarship this year."