

論文要旨

氏名 ブラッシュ鳥越 智美

論文題目（外国語の場合は、和訳を併記すること。）

日本人中学生を対象とした英語力の測定に関する研究

—Elicited imitation test の活用とその示唆を中心として—

論文要旨（別様に記載すること。）

- （注）
1. 論文要旨は、A4版とする。
 2. 和文の場合は、4000字から8000字程度、外国語の場合は、2000語から4000語程度とする。
 3. 「論文要旨」は、CD等の電子媒体（1枚）を併せて提出すること。

論文要旨

世界中で第二言語習得研究が進む中、日本における外国語、特に英語の運用能力の向上に関する関心は年々高まりを見せている。日本の学校教育で法的拘束力を持つ学習指導要領が、最近では平成 20 年に改定され、中学校では平成 24 年度から授業時間が週 3 時間から 4 時間に、また、指導語数は 900 語から 1200 語に増加した。日本の英語学習環境は、特別な場合を除き、概ね学校の英語の授業を主な学習の場としており、限られた授業時間内で運用能力を高めるためには、教師の力量のみならず、効果的な教授法の選択、また、適切なテスト等の評価方法が重要である。英語を学び始めたばかりの中学生は、間違いを繰り返しながら正確性と流暢性を身につけて行くことが予測される。彼らの英語産出能力を学校現場で、効率的に行えるテストとして、本研究では、誘出的模倣 (Elicited Imitation、以後 EI) テストについて、実証的研究を行った。以下、本論文の内容を、構成に沿ってまとめる。

序章においては、研究の背景、本研究の動機、本研究に至る経緯、本研究の意義、本研究の目的について述べた。本研究では EI を、学習者の口頭産出能力テスト法として取り扱う。EI は、聞こえてきた文をできる限り正確にリピートする、というテストである。EI の魅力は何よりもこのシンプルさであろう。4 技能の中でも、話す力のテストは、準備から評価に至るまでに多くの時間を要し、例えばインタビュー形式の口頭産出能力テストでは、インタビュアーの確保が難しく、自由スピーチテスト形式の場合、評価者が評価したいターゲット文法のデータが十分得られるとは限らないという問題点がある (秋山, 2000)。その点、EI は、研究対象となる文法について焦点化してデータ収集ができる点が注目され (Chaudron et al., 2005)、Chaudron et al. (2005) や Graham et al. (2008) は、EI が他のスピーキングテストに比べて、テスト時間が短く、テスト方法がリピートのみというシンプルさという点では実用性が高いテストであると主張している。また、繰り返すだけというテスト手順から、受験者はストレスをあまり感じずに行うことができるために、データ量においても、他の口頭産出能力テストより多くの産出量が期待できるという利点がある。

しかしながら、第二言語習得分野における EI の先行研究では、大学生や成人が研究対象となることが主流で、中学生を対象とする研究はまだあまり行われていないというのが実情である。中学生のリピート活動について、大田 (2009) は、中学 3 年生が 5 語以上の英文を正確にリピートできなかったという国立教育政策研究所の調査結果を伝えている。この調査結果をもとにすれば、中学生を対象にした EI テストは、実施が困難であることが予測された。本研究では、刺激文 (問題文) を聞いて、すぐリピートするものを、IR (Immediate Repetition)、刺激文を聞いて、数秒ポーズを置いてからリピートするものを EI と区別して表記し、IR テストと EI テストの両方を実施した。EI テストは、研究 I では中学 2 年生の 24 名を対象に、研究 II では中学 3 年生の 6 名を対象に実施し、3 名には絵有り EI テストを、残り 3 名には絵無し EI テスト

を実施するケーススタディ形式を用いて研究を行った。研究Ⅰの対象者と、研究Ⅱの対象者は重複せず、異なるメンバーで研究が行われた。

本研究の独自性は、主に3点である。独自性の1点目は、先行研究ではあまり見られない、中学生を対象にEIテストを行い、EIテストが認知的に発達段階にある中学生にとってEIテストが口頭産出能力を測定するテストとして実施可能であるかという点について明らかにしようとした点、独自性の2点目は、絵を提示するEIテスト（以後、絵有りEIテスト）と絵を提示しないEIテスト（以後、絵無しEIテスト）の両方を行い、2種のテストの比較とともに刺激再生法データから、絵の影響について分析した点である。独自性の3点目は、参加者のテスト中の思考内容についての検証を試みた点である。Erlam (2006)を始め、多くの先行研究において、EIが音の記憶テストにならないように、テスト受験者は文の意味に意識を向けることが重要視されるが、どの程度参加者の意識が文の意味に向いていたのかについてはあまり議論されていない。EIテストを「リスニング中」「ポーズ中」「リピート中」の3つに分けて、参加者が意味に意識を向けていたか、その他にもどんなことを考えたり行ったりしているのか、テスト参加者からの発話データを分析して明らかにすることを試みた。

本研究の研究課題（リサーチクエスチョン、以後RQ）は3点である。

- RQ1: 日本人中学生は7語以上の英文をEIテストでどの程度正確にリピートできるのか。
また、文法、語彙、流暢性という3つの異なる観点で採点した場合、能力の違いによってどのような結果が得られるか。
- RQ2: 絵が有る場合と、絵が無い場合のEIテストの結果は、どのように異なるか。
・絵は、参加者の理解・保持・産出においてどのような影響を及ぼすのか。
・絵が無い場合、参加者の理解・保持・産出がどのように行われていたか。
- RQ3: EIテスト中の参加者の思考・意識は、どのような状況であるか。
・リスニング中、ポーズ中、リピート中の思考・意識はどのような違いがあるのか。

第1章の先行研究においては、第二言語習得（Second Language Acquisition、以後SLA）分野の先行研究の概要として、EI研究の動機となる英語指導法であるフォーカス・オン・フォーム（Focus on Form、以後FonF）、インプット仮説や気づき仮説、第二言語習得に関わる理論、スピーキングの言語処理理論、認知情報処理に関わる理論、認知プロセス検証法として刺激再生法や思考発話法の内観法、そして本研究の中心的研究であるEIテストについてレビューした。Long (1991) はFonFを、「言語活動中の学習者の注意がコミュニケーション内容に焦点が向いている状況下で、学習者の言語形式への躓きをきっかけに、言語形式に付随的に指導の焦点を変える指導法」（p. 45-6）と定義している。刺激再生法は、内観法の一つで、作業後に、活動時の録画や書いたものを見ながら、活動中に考えていたことについて口頭で発話を行なうため、回顧法とされる。一方、思考発話法は、同じく内観法であるが、作業の最中に考えている内容

について口頭で発話を行なうため、内省法とされる。FonF は指導法、思考発話はデータ収集法であるが、どちらも、活動中は文の内容に学習者の意識が向いている、という点で共通する。Duff et al. (2009, p. 346) は、EI テスト中のテスト受験者は、刺激文を学習者自身の中間言語文法によって再構築する ("reconstructive") と説明している。中間言語とは、第二言語の学習者が持つ言語知識である (白畑他, 2009)。本論文の EI の先行研究においては、英文を聞いてすぐにリポートするのではなく、数秒のポーズを置いてリポートするスタイルが EI である点、EI テストが「音の記憶テスト」と批判される点を克服するため、Mackey & Gass (2007) や Erlam (2006) が主張するテスト設計や実施における配慮点についてレビューした。また、本研究のリサーチデザインである混合研究法については Creswell (2003) を中心に、ケーススタディを含む質的研究のリサーチデザインは、Yin (1984) と Stake (1995) を中心にレビューした。妥当性については、妥当性の歴史的変遷と EI の妥当性について述べた。印南によれば、使用するテストが小テストレベルの low-stakes のテストか、あるいは大学入試で使用されるような high-stakes のテストかで妥当性検証の方法異なるとし、stakes が高くなるほど、内容妥当性・結果妥当性に加えて、狭義の構成概念妥当性 (「測定したい能力をテストが測定する程度」印南 (2003)) や併存妥当性・予測妥当性・信頼性が検証されていると分析している。質的研究の妥当性については、グレッグ他 (2007) の 3 つの方法、トライアングレーション、メンバーチェック、厚い記述 [高木 (2011) を参照] について述べ、内的妥当性 (トライアングレーションとメンバーチェック) と外的妥当性 (厚い記述)、さらに、採点方法の妥当性、内容妥当性、構成概念妥当性、基準妥当性についてもまとめた。

第 2 章、予備的研究においては、第 1 章で取りあげた FonF 指導法や認知プロセスの検証法、また、EI テストに関する予備的研究の実際について説明した。これらの予備的研究の中でも特に、EI の予備的研究は、本論文の EI 研究の問題文の長さやスピード、語句等の作成に関する点や、ポーズやリポートの制限時間に関するテスト手順、具体的な採点方法と採点基準の詳細において、実際にテストを現場で活用するために示唆を与えるものであり、本研究の妥当性、特に、内容的妥当性の確保においても一助となるものであった。

第 3 章、研究方法においては、「研究方法の概要」で、本研究が研究 I と研究 II で構成され、リサーチデザインは量的分析と質的分析を組み合わせる混合研究法、量的・質的データを同じタイミングで収集する並立トライアングレーション (Creswell, 2003) の位置づけをとること、さらに、仮説を設定しない探索的研究である (Punch, 2009) ことについて述べた。研究 I で参加者 24 名中 3 名、研究 II では参加者 6 名に対して行ったケーススタディは、ある特定の事例に対して理解を深める目的で行う固有ケーススタディ (intrinsic case study) (Stake, 1995) [Baxter and Jack (2008) を参照] である。RQ1 と RQ2 を混合研究法で、RQ3 を質的研究法で進めることを示した後、IR・EI テストの目的、研究の参加者、本研究における分析データ、刺激

文の作成意図や構成、採点方法、妥当性の検証方法についてまとめた。24名を対象に行った研究Ⅰで分析されたデータは、1. NRTの結果、2. 英語環境アンケート、3. 文法知識テスト、4. IRテスト、5. 絵無しEIテストの結果、6. 内容理解の自己評価の結果、7. IRテスト直後のインタビュー結果とEIテスト直後の感想コメントの7点である。NRTとは集団基準準拠検査、全国標準学力検査 (Norm Referenced Test) で、集団基準準拠検査とは、McNamara (2004) によれば、「得点を解釈するのに、受験者間で比較するという枠組みを採用し、個々人の得点が、他者の得点とりわけ得点の頻度分布 (受験者集団でその得点を取るものがどのくらいいるか) に照らして解釈されるテスト (p.76)」である。NRTは「聞く」「話す」「読む」「書く」技能を観点別に評価し、妥当性が高いテストと評価される。本研究でNRTとEIとの相関分析を確認しながら研究を進めることは、EIテストに関わる分析結果の妥当性を確保する点で、重要であると思われる。

研究Ⅱは、絵有り と 絵無しの2種類のEIテストを行うことで、絵が与える影響について明らかにすることが目的であり、絵の作成や3つの観点の評価に時間的な限界があったため、絵無しと有り各3名、合計6名を対象としたケーススタディとして行った。絵を活用した目的は、EIテストが音マネのテストにならないように、学習者の意識を文の内容に向けさせるためである。しかしながら、本研究の絵の使用目的は理解を助けることではないため、マルチメディア学習理論 (Mayer & Moreno, 2003) にもとづき、絵と刺激文音声を同時に与えることをせず、一つ一つ、別々に与えることで、絵の提供による促進効果を挙げないように試みた。また、絵の提示時間も、1.5秒と、短い時間だけ提示し、絵が消えてから、刺激文音声を聞かせた。各3名は、NRTを基準に「高め」「中程度」「低め」が選ばれた。研究Ⅱで分析されたデータは、1. NRTの結果、2. 英語環境アンケート、3. EIテストの結果、4. EIテスト直後の刺激再生法の4点である。量的研究として、NRTの結果、絵無し・絵有りEIテストの結果を、また、質的研究として、英語環境アンケートとEIテスト直後の刺激再生法で得られた発話を分析した。また、6名の中から2名を対象に詳細なケーススタディを行った。

本研究における妥当性は、印南 (2003) の利害関係 (stakes, 以後、stakes) の視点によれば、EIテストは小テストよりも stakes は高いが、高校入試や大学入試ほど大きくないため、内容的妥当性と結果妥当性を中心に妥当性の確保ができたかどうかを検証した。内容的妥当性とは、テスト内容に関する妥当性であり、結果妥当性とは、所謂、波及効果である。内容的妥当性の確保は、予備的研究を行って綿密な計画のもとテスト計画を行ったこと、テスト方法やテスト問題の決定において専門家の助言を得ることや複数の先行研究で妥当性を検証することで行った。また、結果妥当性の確保は、アンケート、観察、インタビューで行った。また、テスト後の様子、インタビュー中、IRテストとEIテストまでの様子の観察も本研究の表面妥当性 (テストが測定したい能力を測定すると受験者やテスト実施者が感じる程度 (印南, 2003)) を補強

する手法として用いた。質的研究の妥当性の検証方法は、トライアングレーション、メンバーチェック（参加者にデータに間違いがないか確認すること）、厚い記述を活用した。評価者内信頼性を確保するため、文字起こしと採点を2回ずつ行い、評価者間信頼性は、IRテストにはKR-20を（ $r=0.98$ ）、絵無しEIテストにはクロンバック α を適用し、比較的高い信頼性が確認された。（文法： $\alpha=0.78$, $M=19$, $SD=4.7$ 、語彙： $\alpha=0.8$, $M=17.5$, $SD=5.1$ 、流暢性： $\alpha=0.8$, $M=18$, $SD=5.6$ ）評価における信頼性を確保するために、IRテストとEIテストの録音データの文字起こしと採点作業は、筆者と英語教師歴4年の第2評価者で行った。評価者内信頼性の検証のため、1回目から2~4か月後に2回目の文字起こしと点数化作業を筆者自身が行った。評価者間信頼性を確保するために、IRとEIの録音データの総数の10%を第2評価者が点数化し、評価者間の一貫性をカッパ係数で検証し、高い信頼係数が得られた（研究I syntax: $K=0.78$, vocabulary: $K=0.91$, fluency: $K=1$ 、研究II syntax: $K=0.65$, vocabulary: $K=0.61$, fluency: $K=1$ ）。

第4章、研究Iの分析結果では、24名を対象に実施したIRテストと絵無しEIテストの分析結果を示した。IRテストの目的は、問題文の長さやスピードが異なる状況で、EIテストに比べて4秒間のポーズが無いIRテストでのリピートを、中学生がどの程度正確にできるのか確かめるためであった。採点法には、正しくリピートされた単語数をカウントし、正解率を算出した。EIテストの採点は、EIが口頭産出能力を測定するテストと位置づけ、syntax（文法）、vocabulary（語彙）、fluency（流暢性）の3観点について Gaillard（2014）や Gaillard and Trembley（2014）が行った0~6点の7段階評価で採点した。採点基準は Gaillard（2014）と Gaillard and Trembley（2016）を参考に、文の半分以上が正しくリピートされれば4~6点、正確さや発話の流暢さが最大半分を下回れば、1~3点という「50%基準」（50% criterion）（Ortega, 2002; Chaudron et al., 2005; Ortega et al., 2005）が基本となった。分析の結果、IRテストは7語以上になると正解率が低下すること、IRテストはNRTや文法知識テストとの相関がかなり強いこと（ $r=0.62\sim0.67$ ）、能力「高め」と「低め」間に有意差があることが確認された。絵無しEIテストの分析では、NRTの4観点（聞く、話す、読む、書く）とのスピアマンの順位相関、能力間の比較を行い、その結果、syntax, vocabulary, fluencyの全観点でNRTの「話す力」との相関が最も高く、syntaxとfluencyではNRTの「総合」や「聞く力」との相関も高かった。「読む力」と「書く力」との相関は、全観点で相関係数が低かったことは、EIテストが「聞く」「話す」活動を伴うテストであって、「読む」「書く」活動とは関係性があまり考えられないことから、事実と矛盾しない結果となった。能力別にNRTとEIとの相関分析を行った結果、「高め」のfluencyを除けば、能力別においてもNRTの「話す力」とEIとの相関が高いことが確認され、その他の技能との相関は低かった。NRT「話す力」との相関が最も高かったのは能力が「中程度」のsyntax（ $r=0.44$ ）、「低め」のvocabulary（ $r=0.50$ ）とfluency（ $r=0.42$ ）であった。このように能力によって傾向が異なる結果となった。

観点毎に、能力間の差があるかどうかについて、ノンパラメトリックのクラスカルウォリス検定で分析を行った結果、**syntax** と **fluency** では $P < 0.01$ 、**vocabulary** は $P < 0.05$ と、3つの観点全てにおいて有意差が確認された。多重比較の結果、**syntax** では「高め」と「中」の間、「高め」と「低め」間に有意差が確認され ($P < 0.05$)、**vocabulary** と **fluency** では「高め」と「低め」間において有意差が確認された。また、スピード「遅め」(80wpm)と「標準」(120wpm)を分けて、能力間の比較を行った結果、「遅め」3問に関して、**syntax** と **vocabulary** に有意差が確認された ($P < 0.05$)。「標準」3問に関して、**syntax**、**vocabulary**、**fluency** 全てにおいて有意差が確認された。**vocabulary** と **fluency** では、高めグループと低めグループ間において有意差が確認された ($P < 0.01$)。

質的研究として実施した IR 直後に1名ずつ行ったインタビューの発話分析から、文の長さやスピードに対する参加者の感じ方、テスト中の思考・意識内容が明らかになった。IR 中は単語(24名)、語順(6名)、文法・文全体・イメージ各4名、文の内容3名であった。文の内容に意識を向けていた3名中2名が能力「高め」に属し、1名は「中間」に属していた。スピードに関しては、「遅め」が9名と最も多く、次に「標準」の7名、「速め」は4名であった。「遅め」を聞きにくい理由は「最初を忘れる」「記憶が続かない」「リズムがつかみにくい」「理解しにくい」とし、「遅い」スピードが必ずしも聞きやすいわけではないことが明らかになった。

ケーススタディでは、IRで1単語もリピートできなかった参加者、IR、EIテストで概ね80%程度の正解率であった参加者、小学校で英会話を6年間継続した経験を持つ参加者についてりテスト結果やインタビュー結果を元に詳しく分析を行った。

第5章、研究IIの分析の結果では、NRTとの相関、能力間の比較分析、絵ありと絵無しの比較分析、3つの観点の相関関係、絵無し(研究I)と絵あり(研究II)の比較、絵無しテスト間の比較分析を行った。また、刺激再生法の発話データを質的に分析した。NRTとのスピアマンの順位相関分析の結果、NRTの「聞く」「話す」「読む」「書く」全ての技能と **syntax**、**vocabulary**、**fluency** の全ての観点で $r=0.8$ 以上のかなり強い相関が確認された。能力間の比較をクラスカルウォリス検定で行った結果では、**syntax**、**vocabulary**、**fluency** の各観点毎に能力差を見た場合、有意差は見られなかったが ($P > 0.1$)、観点で分けずに総合的な得点を能力間で比較すると、絵ありグループ、絵無しグループ、また6名全員の場合も、能力間の有意差が確認された ($P < 0.05$)。絵ありと絵無しの比較をノンパラメトリックマンホイットニーのU検定で行った結果、**syntax**、**vocabulary**、**fluency** の全ての観点での評価において有意差は確認されなかった。質的分析の結果、能力「高め」の参加者は、リピート前の4秒が記憶保持に負担となったこと、「中」や「低め」は、文より単語に集中していたこと、内容理解には難しさを感じていたことが明らかになった。テスト中の思考内容は、参加者全員が「単語」を意識していた。4秒の間やリピート中に文の「意味」に思考を向けていたのは「高」の2人だけで、2人はリピート中、「意味」「単語」

「語順」「文法」（語順は文法に含まれるが、英語科の授業では区別することがある）の全てに意識を向けていたことが明らかになった。

文・ポーズ・リピート時間の感じ方については、能力「高め」の参加者たちは、聞いてリピートするという一連の流れを考慮したコメントをした。絵についての感じ方は3名で異り、絵が文全体の内容や文の保持に作用することや単語力や聞く力が不足する場合は却って混乱を招くこと等が明らかになった。テスト文のスピードの感じ方は能力間で異なることが分かった。能力「高め」の参加者達は、問題文のスピードよりも問題文を聞いてからリピートするまでにかかる時間に注目していたことがわかった。ケーススタディは能力「高め」の2名について、詳しく分析し、「文の長さ」が語数というよりリスニングにかかる時間という感覚を持っていること、テスト中は一連の時間の経過の中で、常に文の「意味」について意識が向いていたことが分かった。

第6章、考察では、3つのRQについて順に考察を行った。RQ1とRQ2は量的・質的分析の結果を統合して考察した。RQ1：「日本人中学生は7語以上の英文をEIテストでどの程度正確にリピートできるのか。文法、語彙、流暢性という3つの異なる観点で採点した場合、能力の違いによってどのような結果が得られるか。」については、研究IとIIの能力間の比較分析の結果を基に考察した。研究IのEIテストではsyntax、vocabulary、fluencyの全ての観点において能力間に有意差があった。特にsyntaxは、どの能力間においても有意差があり、vocabularyとfluencyでは「高め」と「低め」間に有意差があった。研究IIでは、3つの観点において能力間に有意差は見られなかったが、絵有り・絵無しを除外した比較では有意差が確認された。研究IIの絵有りグループの3名間、絵無しグループの3名間にも有意差が見られた。以上のことから、EIテストが参加者の英語能力を弁別するテストである可能性が見えてきた。

RQ2：「絵が有る場合と、絵が無い場合のEIテストの結果は、どのように異なるか。絵は、参加者の理解・保持・産出においてどのような影響を及ぼすのか。絵が無い場合、参加者の理解・保持・産出がどのように行われていたか。」についての考察から、能力が高い学習者の場合は、リスニングからリピートに至るまで絵は補助的にプラスに作用するが、能力が低めの学習者の場合は、必ずしも絵のイメージが文の理解や産出に作用しないことが判った。

RQ3：EIテスト中の参加者の思考・意識は、どのような状況であるか。リスニング中、ポーズ中、リピート中の思考・意識はどのような違いがあるのか、についての考察から、3つの時間的経過点での思考・意識内容は、英語能力によって異なり、「高め」の参加者は、最終目的であるリピートに備えて、意味、単語、文法（語順）と多くの要素に意識を向けているのに対し、「低め」の参加者は、絵を見ても見なくても単語だけに集中する傾向があったことを述べた。

終章においては、主な発見と理論への示唆、SLA/FLA研究分野における示唆、英語教育にお

ける本研究の示唆、研究の限界と今後の研究について述べた。本研究における主な発見は、中学生を対象とした EI テストの実施が概ね可能であるという点、リピート文として、正解率に影響する語数は 7 語以上である点、絵は必ずしも文の内容理解を補助しないこともある点である。また、EI テストが中学校の口頭産出能力テストとして使用できることの可能性についても示唆した。さらに、能力「高め」の参加者は、リスニング中、4 秒のポーズ中、リピート中に文の意味を意識しながら活動しているが、能力「低め」の参加者は、リスニング中からリピートまで「単語」のみを考えている傾向があること、意味に意識を向けたままの状態で行うことができるのは、比較的高い口頭産出能力を持ち合わせている場合であること、能力「低め」の参加者は、絵を見ない場合、文の意味に意識がほとんど向かないか、絵を見ても絵のイメージと文の理解が繋がらないことについても述べた。

理論への示唆として、絵が意識の方向をコントロールする以外に、参加者の産出を助けるというマルチメディア理論や二重符号化理論に基づく問題性を回避するために、絵の提示と音声の提示を、同時に提示するのではなく、別々に提示することで、対処したことについて述べた。

SLA/FLA 研究分野における本研究の示唆として、日本人中学生 2、3 年生の初級レベル学習者が、5 語以上のリピートを聞いてすぐ、あるいは数秒間のポーズを置いた状況でも、ある程度リピート再生は出来ること、4 秒間のポーズ中には、リピートまでに覚えたことを忘れないように、英語を繰り返して心の中で言うなどして記憶に留めておくための独自の対策を行っていたことについて述べた。リピートや EI ができるか、できないか、という単純な問題ではあるが、EI テストを口頭産出能力として実施することが可能かどうかを問うことは、中学生を対象にした EI テストの研究があまり進んでいない現状から考えれば、重要であると考えられる。本研究では、計画、問題作成、データ収集、文字起こし、採点作業、分析を全て手作業で行ったが、先行研究にはコンピュータで問題作成と採点・分析を自動的に行う研究が進められてきている (Cox & Davies, 2012; Graham et al., 2008) 点を述べた。自由に学習者に発話させるという自由発話形式の口頭産出テストと比べて、研究者が確認したい文法項目について、的を絞って発話を誘出させるという点では EI は優れているテストであること、EI における絵の使用に関する示唆にも言及した。また、EI は、口頭産出能力が伴わない場合は点数が伸びないことから、EI テストが学習者の口頭産出能力をある程度反映するテストである点、刺激再生法の分析から明らかになった EI テスト中の思考・意識の内容は、中学生を対象にした EI 研究自体が管見ではあまり見られない EI 研究分野において、意義がある点について述べた。

英語教育における本研究の示唆として 3 点述べた。まず、口頭産出能力テストとしての EI の利用において、EI テストが参加者の英語能力を弁別する可能性が見えてきた点である。このことは、学校現場におけるスピーキング能力の測定方法の拡大という点でも教育的な示唆があると思われる。リピート活動自体、馴染みのある活動であるため、他のテストで感じるような

緊張が与える影響 (McNamara, 2004, p. 20) を抑えたり、テスト手順のシンプルさは、テスト受験者も評価者もストレスを感じずに行えるという実用性が高さは今後の英語教育の中で更に注目され、活用されるべきテスト方法であろう。全国学力・学習状況調査 (全国学力テスト) における中学校の「英語」の実施について、文部科学省は 2016 年 6 月 15 日に、「全国学力・学習状況調査における中学校の英語の実施に関する中間まとめ」を公表した。平成 31 年度から英語 4 技能の学力検査を開始すると発表した。しかし、「聞く」「読む」「書く」の 3 技能に加え、「話す」技能については、教員による対面式で行うインタビューやペアーで行うテストを 10 分程度で実施することが検討されている。本研究で使用した EI テストは、対面式ではあるがインタビューでなく、ペアでもなく、1 人でテストが受けられる点でストレスはかなり下がり、能力「低め」の参加者であっても発話データ量も多いことが期待される。現在、まさに、中学校でスピーキングテストの必要性が叫ばれる中、EI テストは現場のニーズに合ったテストとして寄与できる可能性があると言えよう。指導法としての EI の利用について、大田が、理解の段階から産出の段階への橋渡しとなると主張する (大田, 2010, p. 48) ように、意味内容と文法事項に意識を向けながら行うリピート練習を日常的な指導に取り入れることの有効性について述べた。インタビューや刺激再生法の発話データから、問題文のスピードは、確実に参加者の聞きとりの程度に影響していることが分かったことは、第二言語の教師にとっても示唆的な問題であり、学習対象となる言語の音声スピードは、英語の授業やリスニングテストの際には重要な配慮点の 1 つであると言えることに言及した。

研究 II をケーススタディとして行ったため、参加者数が 6 名と少なかった。絵有り・絵無しグループに能力が「高め」「中程度」「低め」が 1 名ずつであるため、個々人が、いわば、グループや能力毎の代表という立場であった。個人の点数の変動や、刺激再生法による発話の内容分析等、詳細に行うことが出来る反面、得られた分析の結果を一般化するという点では限界があった。また、刺激文の作成においては、既習単語であり、聞きなれない単語よりも、教室でよく聞かれる単語であること、単語レベルが同程度のものであることを考慮したが、IR テストの動詞の再生結果から、同じ動詞間で再生率に差が出た原因として、人名の聞き取りや語の馴染み度の低い単語での躓きが影響する可能性が明らかになった。今後は、参加者の人数を増やし、語の馴染み度の統制についてさらに綿密に行いたい。

本研究が行った、探索的研究によって、検証すべき問題が明らかになってきた。今後は、参加者の人数を増やし、できる限り条件を統制し、さらに精度の高い条件統制を整えて、EI テストのより詳細な実施手順や採点方法について、また、テスト中の情報処理の詳細について研究を進展させていきたい。