

中学・高校における英語4技能の評価 －伝統的評価からICTによる評価まで－

How to Evaluate Four Communicative Skills in Junior and Senior High Schools:
From the Traditional Approach to ICT

林 伸 昭

教師は学習者の反応・興味・アチーブメントを基に自分の授業（指導）を評価する必要があるし、それにより教師が指導項目の生徒への定着の度合いを判定することができる。また生徒もこの評価プロセスに参加し、自らの学習レベルと同時に学習の効果を判断する。このように評価により、教師・生徒ともに自らの指導や学習に関する成功・失敗・反省の材料などを得ることができるのである。したがって評価は、生徒が自らの学習を改善するために、また、教師が自らの指導をより優れたものにしていくために欠かすことのできない活動となる。

キーワード：テスト、測定、評価、ICT、語彙、文法、4技能、

目 次

I はじめに

II 評価とは

1. 評価の定義
2. 評価の役割
3. 評価の方法（テスト）
 - 3.1 テストの種類
 - 3.1.1 客観テストと主観テスト
 - 3.1.2 集団基準準拠テストと目標基準準拠テスト
 - 3.1.3 直接テストと間接テスト
 - 3.1.4 適性テストと到達度テストと熟達度テスト
4. 評価の種類
 - 4.1 診断的評価と形成的評価と総括的評価
 - 4.1.1 診断的評価
 - 4.1.2 形成的評価

4.1.3 総括的評価

4.1.4 観点別評価と評定

4.1.4.1 観点別評価の定義と意義

4.1.4.2 観点別評価における評価規準と判定基準と評定

4.1.4.3 観点別評価の実情

Ⅲ 具体的評価法

1. 言語知識に関する評価

1.1 語彙知識の定義

1.1.2 受容語彙の幅（サイズ）の評価

1.1.3 発表語彙の幅（サイズ）の評価

1.1.4 語彙の深さの評価

1.2 文法知識の評価

1.2.1 文法知識の種類

1.2.2 受容的文法知識の評価

1.2.3 発表的文法知識の評価

2. 4技能の評価法

2.1 受容技能と発表技能

2.2 リスニングの学習

2.3 リスニングの評価

2.4 リーディングの学習

2.5 リーディングの評価

2.6 スピーキングの学習

2.7 スピーキングの評価

2.7.1 スピーキングの間接的評価

2.7.2 スピーキングの直接的評価

2.8 ライティングの学習

2.9 ライティングの評価

2.9.1 ライティングの間接的評価法

2.9.2 ライティングの直接的評価法

Ⅳ まとめ

I はじめに

教師は教育現場のあらゆる場面で、評価を行っている。生活指導面であれば、生徒（学生）¹の多様な行動（朝・帰りの挨拶の有無、怠学による欠席・早退・遅刻遅刻の回数、集会での態度、掃除への取り組み等々）、教科指導面であれば（授業中の私語・発言・居眠り、予習・復習の状況、テストの結果、ノートの記入状況等）で生徒を評価する。つまり、教師は意識してまた無意識のうちに何らかの形で生徒の行動に評価を加えている。ではこの「評価」とはどのようなものだろうか。本論文では、教科、なかでも外国語（英語）の評価に関して考察を行ってみたい。

II 評価とは

1. 評価の定義

酒井(2012)は、教育の目的は、「知識をつけること、できるようになることであり、したがって、教師が一定期間中に実施した授業の成果を何らかの方法で測る必要がある」と述べている。また、古谷(2013)は「評価は、教育目標の達成状況や教育活動の実態を把握する行為で、測定で出された数値を基に価値判断をすること」と定義している。

したがって、「評価」とは、「ある一定期間にわたって、教師が生徒に指導及び生徒が学習した内容が、どれだけ生徒に定着しているかを何らかの方法で測定し、その定着の度合いを判断すること」と定義できる。

2. 評価の役割

金谷(2009)は、評価の意義には以下の3点があると指摘している。まず、評価は生徒の学習活動を推進する。評価により生徒の学習達成状況を的確に本人にしらせることにことで、生徒はそれまでの自分の学習への取り組みを振り返る機会を与えられ、その反省を基に生徒は次の学習をより適切に行う環境を作ることができる。次に、評価は教師の指導も改善する。生徒の達成状況を的確に把握・分析することにより、それまでの教師自身の指導を振り返る機会を与える（生徒に指導した方法が適切であったかどうか；教師は、生徒は指導内容をよく理解していると思っていたのに、実際には生徒は理解できていなかった等）。このような機会を通じて教師はより適切な指導へと自己の指導法を改善することができる。最後に、評価は教育課程を点検してくれる。教師は、評価結果から教育課程が適切に実施されているかどうかを判断することができ、さらに、教科で作成した年間指導計画にのっとり、授業を始めとする教育活動がその学校（教科）の教育目標に向かって着実に進められているかを判断し必要に応じて軌道修正することができる。

生徒は評価の結果（具体的には、テストの結果や通信簿の5段階の評定）に一喜一憂するもの

であるが、評価は生徒・教師・学校（教科）に、それぞれが行ってきた学習（生徒）、指導（教師）、教科指導計画（教科・学校）に多様な情報を与えてくれる。

3. 評価の方法（テスト）

望月(2010)は、評価は測定とテストに基づいて行われると述べている。測定は行き当たりばったりでなく計画的で明確な手順と規則に従って生徒の能力・適性・動機等を数量化する方法である。テストは、測定の1つの方法で、生徒の能力・適性・動機等を数量化するために、そのサンプルを引き出すように設計されたものである。評価は測定した後に数量化された結果に価値判断をすることである。

つまり、評価をするためには、テストで生徒の能力・適性・動機等を数量化した値を出すのが、能力・適性・動機等をどう数量化すればよいかを考察するのが測定で、測定の結果得られた数量化した値が何を示すのかまたどう判断すればよいかを考察するのが評価ということになる。

3.1 テストの種類

3.1.1 客観テストと主観テスト

望月(2010)は、テストにどんな種類があるかを考える時に、最も大きな違いはテスト項目の形式が客観的に採点できるかどうか（採点者の主観が数量化に影響を及ぼすかどうか）であることを指摘している。江原(2015)は、代表的な客観テストとして、採点が誰でも同じ点数が出る多岐選択式テストや正誤法等をこのテストを挙げている。また、代表的な主観テストは採点者の主観で点数が与えられるテストには、英作文のタスク（自由英作文や制限英作文）、リーディングの要約（口頭と筆記）等がある。

3.1.2 集団基準準拠テストと目標基準準拠テスト

望月(2010)は、学力テストや入試でよく使われる相対評価は集団基準準拠テストであり、一方、現在、小・中・高において通知表などに使われている観点別評価とは目標基準準拠テストであると指摘している。2つ評価の大きな違いは、前者が他の生徒と比べ当該生徒の相対的な位置関係を調べるのに対して、後者は当該生徒の絶対的な能力を調べる点である。

3.1.3 直接テストと間接テスト

直接テストはダイレクトに「測定したい技能」を生徒に行わせるテストで（ライティングのテストであれば、あるトピックに関する実際に英文を自由に書かせる等）、間接テストはある能力を測定するためにその能力の基盤となっている下位能力を測定することによって「ある能力の代用」とする（発音のテストをペーパーで行う等）である。

3.1.4 適性テストと到達度テストと熟達度テスト

江原(2015)は一般的な言語適性を測定する MLAT (Modern Language Aptitude Test) 等が適性テストにあたりと記述している。

望月(2010)は到達度テストは、生徒が授業で学習した知識・技能をどれくらい身につけたかを測定するテストで、テスト範囲を決めて行う定期考査がこれに該当すると指摘している。

高橋・高梨(2007)は熟達度テストは生徒の学習経験や学習内容に関係なく英語力を測定するテストで、テストの内容や難易度は生徒が受けた教育カリキュラムとは関係がない、いわゆる実力テストと定義している。具体的には英語を使って何をどれだけできるかを測定するもので TOEFL、IELTS、TOEIC 等がこれに該当する。

4. 評価の種類

4.1 診断的評価と形成的評価と総括的評価

4.1.1 診断的評価

江原(2015)は診断的評価は、生徒の英語に関する知識・技能のレベルや問題点を教師が把握し、その後の指導方針に生かすための評価で、中学・高校・大学への入学時や、学年、学期の始めに行うことが多いと指摘している。

また、高橋・高梨(2007)も、診断的評価は生徒の英語力の強いところや弱いところ、形成済み項目や未形成の項目を明らかにするために評価と述べている。

4.1.2 形成的評価

望月(2010)は形成的評価は、指導過程の中で継続的に繰り返される評価方法で、現場では小テストがこれにあたり、教師にとっては授業の改革や改善に、生徒にとっては自分の学習の反省に役立つものと定義している。形成的評価をこまめに繰り返し行うことにより、生徒の学習状況及び教師の指導方法の良否が判断できるものと考えられる。

4.1.3 総括的評価

江原(2015)は総括的評価は一定の指導の後に（たいていは学期ごとに）生徒の学習の達成度を把握するための評価で定期考査がこれにあたり、学習者はこの評価で自分の長期にわたる学習の成果を知り、教師も長期にわたる自分の指導を省みて次の期間の授業の改善・改革への資料を得る評価と記述している。したがって、総括的評価は、「より長いスパンでの形成的評価」とも定義できる。

4.1.4 観点別評価と評定

4.1.4.1 観点別評価の定義と意義

金谷(2009)によれば、教科における評価・評定は、指導要録に記載され、記載内容は原則として情報公開の対象とされている。したがって、評価・評定は説明に耐える妥当・適切な手順により行われ、かつ信頼できるものでなくてはならない。学習指導要録の評定は、平成14年度から、それまでの「集団に準拠した評価(いわゆる相対評価)」から「目標に準拠した評価(いわゆる絶対評価)」に改められた。

北尾(2004)は、生徒の学力は彼らの有する知識だけでなく思考力や意欲などを含めた複数の観点から評価する必要があるが、そのような評価に用いられるテストを観点別評価テストと呼び、その測量値に基づく評価を観点別評価と呼ぶと定義している。

高橋・高梨(2007)は、学習指導要領の目標はそれを達成する具体的な施策(指導方法)の2つがセットになっていないとしばしば目標と学習結果にズレが生じことがあるが、そこでこのズレを埋めるために提案されたのが観点別評価であると指摘している。

望月(2010)は、中学校外国語科の目標は、コミュニケーションの基礎を養うことであり、「①外国語を通じて、言語や文化に対する理解を深めること、②外国語を通じて、積極的にコミュニケーションを図ろうとする態度の育成を図ること、③聞くこと、話すこと、読むこと、書くことなどのコミュニケーション能力の基礎を養うこと」の3つであると述べている。また、高等学校の外国語科の目標は、「①外国語を通じて、言語や文化に対する理解を深めること、②外国語を通じて、積極的にコミュニケーションを図ろうとする態度を育成すること、③外国語を通じて情報や考えなどを的確に理解したり適切に伝えたりする能力を養うこと」の3つから成り立っていると指摘している。

つまりこれらには、高橋・高梨(2007)も指摘しているように、「①外国語を通じて、言語や文化に対する理解を深めること、②外国語を通じて、積極的なコミュニケーション能力を築くこと、③4技能を統合して意思の伝達を図ることができるコミュニケーション能力を育成すること」の4つの要因が含まれており、このことから観点別評価に、「①コミュニケーションへの関心・意欲・態度、②外国語表現の能力、③外国語理解の能力、④言語や文化についての知識・理解」という4つの評価の観点を設けたわけである。

また、各評価の観点における評価基準は、A：十分満足できる状況、B：おおむね満足できる状況、C：努力を要する状況、となっている。

4.1.4.2 観点別評価における評価規準と判定基準と評定

金谷(2009)は、観点別評価を行う際、次の手順として、「評価規準」を設定する必要があると述べている。評価規準は、生徒の学習状況が目標に対してどの程度達成しているものかかを判定するための視点で、評価の各観点を適切に反映したものであることが求められる。この、評価規準に基づいて、A、B、Cの3段階の評価を決定するが、その際必要になるのが「判定基準」である。判定基準では、「具体的な評価規準」をさらに客観的に具体化することが求められる。

数値目標を示したり、チェックリスト的なものを用意して、その達成率により呈することが考えられる。また評定は、その教科の目標に対するある段階での総合的な評価である。外国語（英語）では、収集した評価資料を課ごとに総括しさらに学期ごとに総括して観点別学習状況の最終評価を決定する。最終評価は一定の基準により1～5の5段階で評価し、これが評定となる²。

4.1.4.3 観点別評価の実情

教育課程研究センター（2012）は、「小・中学校の学習評価では観点別学習状況評価の着実な浸透が見られるが、高等学校の学習評価では観点別学習状況の評価の趣旨を踏まえた学習評価を行い、授業の改善につなげる努力をしている学校がある一方で、ペーパーテストを中心として、いわゆる平常点を加味した成績付けのための評価にとどまっている学校があるとの指摘がある。高等学校においても、学校教育法や新しい学習指導要領を踏まえ、基礎的・基本的な知識・技能に加え、思考力・判断力・表現力等、主体的に学習に取り組む態度に関する観点に関する評価も行うなど、観点別学習状況の実施を推進し、きめの細かい学習指導と生徒一人一人の学習の確実な定着を測っていく必要がある」と指摘している。

この文章から判断する限り、大半の小・中学校では観点別評価とその結果に基づく評定が行われているが、高等学校においてはまだ観点別評価が十分に浸透していないという現状が認識できる。

そこで、本論では、タイトルの「中学・高校における4技能の評価－伝統的評価からICTによる評価まで－」に基づいて、以下、従来の相対評価も含めて4技能及び言語知識の具体的評価法³を考察していきたい。

III 具体的評価法

1. 言語知識に関する評価

1.1 語彙知識の定義

単語を知っているということには以下の多くの要素が含まれる：1) 発音、2) 単語の文法（活用）、3) 意味、4) コロケーション、5) 同意語、6) 反意語、7) 上位語、8) 下位語、9) 同語源語等々である。そして、多くの単語を知っている場合には「その生徒の知っている単語知識の幅（語彙サイズ）が広い」と言い、また、ある単語に関して「上記の2）～9）の多くを知っている時にはその生徒の単語知識は深い」と言う。さらに単語には読んだり聞いたりして理解できる「受容語彙」と、書いたり話したりできる「発表語彙」がある。

では、このような受容単語・発表知識の「広さ（幅）」と「深さ」をどのように測定し評価すればよいか。以下が代表的な評価法になるであろう。

1.1.2 受容語彙の幅（サイズ）の評価

望月他(2003)によれば、語彙サイズのテストとしては Nation (2001) の *The Vocabulary Levels Test*、Maera (1992) の *Check List Test*、望月 (1998) がある。望月 (2003) によれば、これらのテストはいずれも基となる語彙表から測定すべき単語のサンプルを採り、そのうちどれだけを知っているかの割合を、元の語数と掛け合わせて語彙サイズを測定する。

1.1.3 発表語彙の幅（サイズ）の評価

望月他(2003)によれば、Laufer and Nation (1999) は発表語彙を、学習者が自分の思いのままに使える発表語彙（自由発表語彙）と、他からヒントを与えられたり使うように強制された時にだけ使える発表語彙（統制的発表語彙）の2つに分類している。テストされる単語は文脈で提示され、最初の数文字が与えられている。生徒は文脈に合うように最初の数文字の後につづりを書いて単語を完成する。

1.1.4 語彙の深さの評価

望月他(2003)は、単語知識の深さの評価方法としては、*Vocabulary Knowledge Scale* (1993) と *Test to Measure Depth of Vocabulary Knowledge* (1998) が語知識の多様な側面を評価する優れたテストであることを指摘している。*Vocabulary Knowledge Scale* (1993) においては生徒は、「(1) 見たことがない；(2) この語は見たことはあるが、どういう意味かわからない；(3) この語は見たことがある → 同意語は～で、訳は～である；この語は知っている → 同意語は～で、訳は～である；この語を使って文を作れる → その単語を用いた文章の作成」というテストを受ける。*Test to Measure Depth of Vocabulary Knowledge* (1998) においては、生徒はある単語から連想して挙げられる語のネットワークを用い、語の連想を3つ（連合的、統語的、分析的）に分けて、それぞれの連想を選択させる形式の語彙テストが作成された。例えば edit という語が目標語なら連合的連想語は revise、統語的連想語は film や text、分析的連想語は publishing になる。

1.2 文法知識の評価

1.2.1 文法知識の種類

1.1の語彙知識の評価と同様に、文法知識の評価を行うためにも、ある文法項目（例：現在進行形）の認識や理解（受容的知識）に焦点を当てる場合もあれば、発表（発表的知識）に焦点を当てる場合もある。

1.2.2 受容的文法知識の評価

江原(2015)は多岐選択方式が受容的文法知識の測定を行うのに、つまり英語の文法を理解しているかどうかを評価するのに効果的であると指摘している。例えば、Yoko has been playing

soccer () ten years now. の () に入れる選択肢として、(a)since, (b)for, (c)while, (d)during を生徒に与える、正解は b であるが、正解以外の選択肢は錯乱肢と呼ばれる。

1.2.3 発表的文法知識の評価

江原 (2015) は文法能力の発表能力を測定するテストとして以下の 3 つを挙げている。

1) 穴埋め問題：「a, an, the を () に入れなさい」という指示を与えて、冠詞の発表能力を測定する One day () young girl met famous tennis player who had graduated from () same elementary school that she graduated. () girl was so excited to get () athlete's autograph ; 2) 書き込み問題：「ユミが言ったことを書いて文章を完成させなさい」という指示を与える。このテストでは生徒が過去形を使用する手掛かりとして文中の last month というフレーズを利用利用して、この文が過去のことを示す文であることに気づくことを期待している。Yumi: Hi, Naomi. Is anything wrong? You look sad. Naomi: Hi, Yumi. That's right. I feel terrible. I get a mail from my cousin in Canada. Yumi: Oh, she _____ last month, didn't she. Naomi: Yes, now she says she didn't enjoy her summer vacation in Japan ; 3) 和文英訳：「以下の文を英文にしなさい → 赤い帽子をかぶっている女性は祖母です」という指示を与えて、生徒が The woman (who is) wearing a red hat is my grandmother. という英文を作成することを期待する。このテストにより、「関係代名詞」や「現在分詞の形容詞用法」の生徒の発表能力が測定できる。

以上のような語彙・文法のテストを繰り返しこまめに実施し、教師は生徒の語彙の定着度、文法項目の理解度を評価していく必要がある。

2. 4技能の評価法

2.1 受容技能と発表技能

土屋 (2011) は、英語に限らずある言語に習熟するということは、最も普通の定義ではリスニング、スピーキング、リーディング、ライティングの 4 つの技能を獲得することであり、彼は 4 つの技能を「基本的言語技能」と呼んでいる。さらに、彼はこの 4 つの技能のうち、リスニングとリーディングは受容技能であり、スピーキングとライティングは発表技能であると定義している。この章ではこれら 4 技能の効果的な評価法を検討する。

2.2 リスニングの学習

秋山 (2011) は、リスニングは 4 つの言語技能の中で最も基本的な技能であると指摘している。それは、母語の習得に見られるように人はまずリスニングの言語活動を習得し、それからスピーキングや他の言語技能に進んでいくからである。また、会話においてコミュニケーションを成立させるためには話し手の言っていることを理解できなければならない。このようにリスニングは

言語習得やコミュニケーションにおいて基本的な技能であると秋山(2011)は指摘しているが、以下、このリスニングの言語技能をどう測定し評価すればよいかを検討する。

2.3 リスニングの評価

本節では、リスニングの言語能力(リスニング力)の測定法をランダムに提示していく。1) 多岐選択テスト法: 垣田他(1988)はある一定の長さのテキストを生徒に聞かせ、その内容に関する T-F Question, WH Question の設問と選択式の解答を作成し、選択肢の中から正解を1つ選ぶ多岐選択テストを紹介している。多岐選択式のテストでは多様な難易度で一度に多くの出題が可能であり、したがって高学力者も低学力者も挑戦できる幅広い問題を準備することができ、また、採点も迅速に行うことができるという利点も指摘している。多岐選択テスト法では、英文の構造上の知識・英文内容理解・推理能力・話者の意図を判断する力・感情や比喩表現の察知力をテストする問題を加えることで、高度なリスニング能力も測定できる; 2) ディクテーション: ディクテーションは聞き取りテストとして従来から広く用いられてきた。森山(2009)は、ディクテーションにも、テキストの中に設けられた空所を教師の音読や CD を聞いて埋める部分ディクテーション、教師が読んだ英語をできるだけ正確に再現するディクト・コンポ、聞き取った内容を持ち寄って元のテキストに復元するディクト・グロス、教師が発した英語を絵で描くディクテーション等、多様なディクテーションが考案されていると述べている。ディクテーションは、音声を聞き取りそれを文字に結びつけその復元した文字情報の正確さで生徒のリスニング力(聞き取り及び音声識別能力という本来のリスニング力に加えて、音声を表す文字の組み合わせに関する知識及び英語の構造上の知識、特に形態上の知識)を測定・評価することができる; 3) サマリーの作成: ある一定の長さのテキストを2~3回ほど聞かせ、その概要を英語もしくは日本語で書かせる。

このようなテストで生徒のリスニング力を測定すれば、生徒が単語段階から Connected Speech に至るまでの多様な音声をどれくらい聞き取れるかを測定・評価することが可能となる。

2.4 リーディングの学習

岡他(2004)はリーディングの学習は、「英文を日本語に訳すことが目的で訳せたらそれですべて終わりというものではない」と主張している。リーディングにおける読解活動の目的は、ただやみくもに英文を日本語におきかえるのではなく、書き手の意図をつかみ自分の目的にあった情報を獲得することである。訳は確認のひとつの手段にすぎず、英文をいちいち日本語に訳さずに自分が必要とする情報を得るためにスキミングやスキニング等のスキルを用いて直読直解をし、訳す作業としての読解から内容を読み取る読解へとシフトする必要がある。そして、英文中に書いてある事実から作者の意図を推測し、さらに作者の考を評価しながら読んでいく読解力を養成するというのが岡他(2004)の主張である。

以下、リーディングの言語技能(リーディング力)をどう測定し評価すればよいかを検討する。

2.5 リーディングの評価

垣田他(1987)は、何がリーディング技能の構成要素であるかを明らかにしなくてはテストを作成することはできないと指摘している。垣田他(1987)の考えに従えば、リーディングテストはリーディング技能に直接関与している要素に焦点をあてて問題を作成することになり、テスト問題で扱われて要素を訓練すれば生徒のリーディング力は理論上は向上することになる。

そこで、垣田他(1987)はリーディングテストに盛り込まれている問題項目を全体的に見回して以下の3項目を導き出した：1) 文字・記号にかかわる領域：文字の確認識別、音声と文字の対応関係の能力を見る知覚レベルの作業が中心となる；2) 語彙・文法構造に関わる領域：文脈中の任意の語彙の意味を母国語で尋ねたり、抜粋中の特定の文とか文の一部（個々の語あるいは語句）の文法上の機能が理解できているかどうかをチェックする。つまり、主として言語上の知識を点検する；3) 内容把握にかかわる領域：パラグラフ全体もしくはパラグラフの一部の意味が理解できているかどうかをテストする。生徒の総合技能としての読み取り能力をテストすることになる。

これらの3項目に基づいて、垣田他(1987)は以下の具体的なテストを考案した：1) 組み合わせ法：このテストでは語や文の再認がどの程度うまく行われるかを見る。生徒には類似した綴りの語や文を識別する能力が求められる。例えば「語の組み合わせ」では、与えられた語と同一の語が多岐選択肢のどれになるかを選ばせる。このテストにおいては、パワーポイントを用いて時間制限内にどれだけ数の類似した綴り字や語を識別できるかを測定することもできる。また、「文の組み合わせ」では、与えられた文と同じ文を多岐選択肢の中から探し出す；2) 正誤問題：リーディングにおいて最も広汎に用いられているテストである。このテスト形式は多岐選択法と異なり、多くの錯乱肢を必要としないので、問題作成が容易である。例えば、ある英文のパッセージを読んで、問題文がパッセージの内容と一致しているかどうかを生徒は判断する。それにより、生徒がパッセージの内容を正確に理解しているかどうかをテストできる；3) 単文・長文を用いた多岐選択法：単文や長文中の任意の箇所を選び、準備した選択肢のうちどれが前後の文脈関係から判断して適切かを判断させる・*The eyes are wonderful teachers - even musicians, who deal with sound, learn as much by (1 doing, 2 playing, 3 watching, 4 practicing) as by listening.* このタイプは一見語彙テストのように見えるが、選択肢を選ぶ際に文脈をふまえて解答する必要があるためリーディングテストとなる；4) クローズ法：クローズ法は、「読み手は文脈上の手掛かりをもとに、文中の欠落した箇所（空所）を修復できるはずである」という「ゲシュタルト心理学の不完全な形態を完全なものとして知覚しようとする人間の潜在的傾向」を基盤とした考えを応用し利用している。クローズ法では、ある程度の長さの英文の最初の英文を残して、その後の英文のn番目（大抵は、5～10番目）ごとの単語を空所にしていく。生徒は前後の文脈をよく理解して最も適切な単語を入れなくてはいけない；4) 要約課題：生徒は比較的長い英文を読み、重要なトピックセンテンスを見つけながら、削除、一般化、選択、構成等の要約規則を用いて要

約を作成する（卯城他，2012）。

上記のように、リーディング力に関しては、ボトムアップとトップダウンの両者の力を測定・評価する必要がある。

2.6 スピーキングの学習

土屋他(2011)は、スピーキングを行うためには、「言語知識」と「言語を使って話し、話し手が話した情報を同時に処理する能力」が必要であると指摘し、例えば前者には音声の知識（強勢、リズム、イントネーション等）や（コミュニケーションを行っている場や相手に）適切な語彙を選択して使える能力や、発した言葉が相手に理解できるように発話を組み立てることができる文法能力が含まれ、後者には「言語知識」を活用して相手が意図しているように反応して相手に話す能力、話し手の言っている事が理解できる能力、相手が言ったことに即座に反応することができる能力が求められると指摘している。

したがって江原(2015)が指摘しているように、スピーキングはリスニングやリーディングにより得られた情報や知識に基づいて進められるものであるがその際に話者は、話す内容（主張、意見等）、それを英語で表現するための知識（文法体系、語彙等）、口頭で発表するための音声（発音、抑揚等）、そして確認、質問、相づちなどのスピーキングをスムーズに進行させていく知識を持っている必要がある。さらに、持っているだけでなくそれを実際に使える能力を有している必要がある。スピーキングにおいては、スピーキングのための明示的知識をいかにスムーズに手続き知識に変えて利用できるが重要となるのである。

2.7 スピーキングの評価

垣田他(1989)はスピーキングのテストでまず問題になる事柄として、スピーキングの測定には、直接評定者が生徒と向かい合い英語の応答を等して測定する方法（直接測定法）とペーパーテストやクローズテスト等間接的な手段でスピーキングの測定を行う方法（間接測定法）があり、信頼性の面から言えば間接測定法の方が安定しているように思われるが、反面、間接測定法には妥当性に難がみられ直接測定法の場合には信頼性に難がある点を指摘している。

本論では、間接測定法と直接測定法の2つの測定法を考察する。

2.7.1 スピーキングの間接的評価

望月他(2010)は測定したい技能を間接的に測定するテストが間接テストで、例えば、スピーキングにおける発音の能力を測定するために、実際にその単語を発音させるのではなく単語のどこに第1アクセントがあるかを生徒に答えさせるのが間接テスト（測定法）であると述べている。

垣田他(1989)はスピーキングの間接テストとして以下のテストを考案している。1)ペーパーテスト：ペーパーテストでのスピーキング力の測定法には以下の5種類がある— (1)アクセント、

ポーズ、ストレス、イントネーション等を記号で記入させる・例としては、(1) Could rhymes with A. blood, B. food, C. wood で Could の下線部と同じ発音の箇所に下線部がある語を記号で答えさせる；(2) 試験管の質問に対して答えを書かせたり選択肢の中から選ばせる・質問 Did you come to school by bus? 答え・A. Yes, I did. B. Yes, I am. C. Yes, it is.；(3) あるまとまりのあるテキストを聞かせそれについての口頭質問に答えを書かせたり、選択肢の中から答えを選ばせる；(4) 本文も質問も印刷して与えておき、答えを書かせたり選択肢の中から選ばせたりする・Edward played ball this afternoon. It is six thirty now. He is at home. He is in the living room. *Mother*: Are your clothes wet? *Edward*: No, they (A). *M*: Do you want to eat now? *E*: (B) I'm hungry and thirsty. (C) I played ball this afternoon. *M*: Do you want some juice? *E*: Yes, please. (D) *M*: Yes, it is. It's cold orange juice. *E*: Thank you, Mother. It's very good. *M*: (E). *E*: All right, Mother. I'm ready. 選択肢 ア I'm tired, to. イ Let's eat now. ウ They're dry. エ Yes, thank you. オ It is orange juice?

垣田他(1989)は、ペーパーテストは問題を正しく解答できたとしても、実際の運用面で正しく発音(話)できなければ評価方法としての意義が薄れてくる。そのためペーパーテストは実際のスピーキングテストの代用にならず、テストはスピーキング評価にとどめるべきであるという主張があることも指摘している。

2.7.2 スピーキングの直接的評価

望月他(2010)は、直接テストは、測定したい技能を実際に生徒に行わせるテストでテストで用いるタスクもテキストもできる限りオーセンティックなものでなければいけず、スピーキング力を測定する場合には生徒に実際に発話させるのが直接テストになると述べている。

垣田他(1989)は以下のようなスピーキングの直接テストを考案している。1) 発音のテスト：発音のテストでは、個々の音素、単語や文のストレス、イントネーション等が評価の対象に含まれる。例えば評価の対象となる音素を含む事物を描いた絵を(ドレスの絵を /dres/ で /r/ を /l/ と)正しく発音できるかを測定する；2) 語彙のテスト：語彙テストでは以下のように、生徒の母語を利用することもできる(質問・『月曜を英語で何と言いますか?』→答え『Monday』、質問・『お父さんのことを英語で何と言いますか』→答え『Father』)；3) インタビュー：評価者と生徒が互いに向き合いインタビューを行うテスト形式では、基本的な定型レベルの英問英答から、あらかじめ準備したトピックにそって評定者が質問し生徒に応答を促す比較的高いレベルのものまで、学力に応じた実施方法が考えられる。インタビューに関してはALTが来校できない場合にはスカイプ等のICTを用いて評価することもできる；4) スピーチ：スピーキングテストのスピーチにはあらかじめトピックを決め、準備した文章を暗唱させる形式から、1枚あるいは連続した絵を生徒に提示し1～2分の程度の準備時間を与え、約2～3分学力に応じて単文を作らせたり内容の説明させるものがある；5) ディスカッション：生徒は特定のトピックに関する発話の中で

listener-speaker としての役割を同時に要求され、しかも自分の考えや意見をタイミングよく適切に表現することが求められる。

江原 (2015) はさらに高度な測定法としてディベートを挙げている。ディベートはある命題や提議について、生徒（もしくは生徒のグループ）が賛成と反対の立場をとり、お互いに質問し合ったり主張を述べ合ったりする。最後にジャッジ（通常は JTE か ALT が行う）が、どちらの立場がより論理的で説得力が主張をしたかを判定する。

以上のテストでは生徒が実際に話さなくてはいけないという点で、これらのテストはまさに直接テストとみなされるが、テストの信頼性を高めるためにはできれば JTE と ALT の 2 名で評価して、評定者間信頼性を高める努力をすることが望ましい。

2.8 ライティングの学習

JACET (2012) はライティングはスピーキングと同様、自分の考えを発信する発表技能であり、スピーキングが「話し言葉」である一方、ライティングは文字媒体を介在する「書き言葉」であることがその最大の違いであると述べている。その場での対応が求められるスピーキングに比べ、時間をかけ推敲を重ねられる点や、綴りが正しければ発音できなくても読み手に理解してもらえるという点でライティングの方がスピーキングよりも容易であると言える。しかし、スピーキングは表情、ジェスチャー、声調、相手の反応などの補助的手段（パラ言語）を利用できるが、ライティングは目の前にいない相手に文字だけを通して的確に自分の考えを伝えることを求められるという点で、スピーキングよりも高度な認知能力を必要とする活動である。またライティングは単なる紙上に書かれた文字あるいは個人の単独作業ではなく、読み手と書き手そしてそれぞれの社会的文脈が相互に作用しコミュニケーションする社会的行為であるとも言える。

2.9 ライティングの評価

江原 (2015) はライティング力を測定する最もよい方法は、スピーキングの力の測定と同じように直接にテストすること、つまり、生徒に実際に書かせることであると述べている。しかし、生徒全員のライティングの直接テストを採点するには教師の時間的制約もある。

そこで本節では、スピーキング力と同じくライティング力についても間接測定法と直接測定法の 2 つの測定法を考察する。

2.9.1 ライティングの間接的評価法

垣田他 (1989) はライティングの間接的評価方として以下の例を考案している。1) 書くことの機械的技能に関するテスト：句読法の測定（必要な句読法を用いて以下の文章全体を書き直させる → 指示文 In the following passage there is no punctuation. Write out the passage, putting in all punctuation and capital letters. 問題文 [lend me your pen please peter asked], [be

careful i said] ; 2) 書くことの文法技能と文体技能に関するテスト：各領域での文法テストは、文法構造の知識ではなく、それを書き言葉としてふさわしいように運用できる力を測定する必要がある 1 主語と動詞の一致 The design of the two new bridges (are / is) very unusual. 2 構造の対比 She enjoyed sewing, reading, and just (to sit / sitting) on the porch watching people go by. 3 代名詞の格 To my little brother and (I / me), Uncle John was the most wonderful friend. 等 ; 3) 書くことの判断技能に関するテスト：文法的には正しいが文脈には不適切な語を混ぜておき、文脈上最も適した文を選択させて文体についての感覚を得る →1 完成問題（多岐選択法） During his last year in office, his popularity with the voters (). A: descended downward swiftly B: quickly touched bottom. C: declined very suddenly. D: rapidly plunged under.

2.9.2 ライティングの直接評価法

垣田他(1989)はライティングの直接テストとして以下の項目を考えた。1) 場面・指示英作文：状況や場面を設定して、英作文内容と表現方法（レジスターや文体）を制限する。このテストでは、語法上の正確さよりも効果的な伝達が可能かどうかを見る → 絵や写真を見せて場面や状況を描写させる Directions: Look at the pictures. Write down what is happening. Use complete sentences. You have 2 minutes to write as much information as possible ; 2) 自由英作文：書く内容（スキーマの有無）が特定の生徒に有利に、あるいは不利に働くことは好ましくない。そこで、どの生徒にもわかりやすく身近な話題で、また興味が出て人に伝えたいような内容になるトピックを選択する必要がある。想像力とか獨創性を必要以上に求めるトピックではなく、生徒の有する語彙・文法知識を正しく使用し書きたい内容を読み手に上手に伝えられる程度のものがよい。自由英作文とはいえど、適切な時間制限と字数制限を設定することが必要となる。また、完成した英作文にフィードバックを与える場合には、ワード等のワープロファイルで提出させ、文法の間違ひはグリーン、綴り字の間違ひはイエロー、パラグラフ全体の構成が必要な場合にはそのパラグラフ全体をレッドでマーカーして訂正した原稿を生徒に再提出させながら最終原稿を完成させ、その原稿を評価するというプロセス・ライティングによる評価を行うこともできる。

直接テストでは、すでに述べたように、生徒のライティングの出来、不出来が「指定されるトピック」に左右されやすい。また採点に評定者の個人的主観が入りやすい。従って、1人で採点する場合はしっかりと採点基準を設定しておくこと、あるいは、スピーキングと同様にテストの信頼性を高めるためにはできれば2名以上で評価して、評定者間信頼性を高める努力をすることが望ましい。

IV まとめ

垣田他(1992)は、教室における評価の主たる目的は、生徒及び教師にそれぞれ自分達の学習及び指導の達成度を教えてくれるものであると述べている。評価は生徒の英語力の弱点や長所を教えてくれるし、また、教師にとっては自分達の指導方法を省みる機会を与えてくれる教授・学習の一翼になうものであり学習内容の習得の目的達成に向かう最終ステップであると言える。教師は学習者の反応・興味・アチーブメントを基に、自分の授業を評価する必要があるし、それにより生徒の学習の度合いを判定することができる。また生徒もこの評価プロセスに参加すれば、生徒は自らの学習レベルと同時に学習の効果を判断することもできる。このように評価により教師・生徒ともに、自らの成功・失敗・反省の材料などを得ることができるのである。

したがって評価は、生徒が自らの学習方法を改善するために、また、教師が自らの指導方法をより優れたものにしていくために欠かすことのできない活動なのである。

注

- ¹ 一般に小・中・高の学習者を「生徒」、高等教育の学習者を「学生」とよぶが、本論文では両者を併せて以後、「生徒」と呼ぶことにする。
- ² 詳しい算出法は、金谷(2009: 225~261)を参照のこと。
- ³ 本論文では、4技能の他にも、「英語に関する言語知識(語彙・文法・構文等)」の評価法も考察する。

参考文献

- 卯城 祐司他(2012)『英語リーディングテストの考え方と作り方』 東京：研究社
- 江原 美明他(2015)『基礎から学ぶ英語科教育法』東京：松柏社
- 岡 秀夫・赤池 秀代・酒井 志延(2004)『英語授業力教科マニュアル』 東京：大修館書店
- 岡田 圭子(2015)「評価とテスト」岡田圭子・江原 実明他(編著)『基礎から学ぶ英語科教育法』(pp. 269-293) 東京：松柏社
- 垣田 直巳・松村 幹男他(1987)『英語のリーディング』 東京：大修館書店
- 垣田 直巳・吉田 一衛他(1988)『英語のリスニング』 東京：大修館書店
- 垣田 直巳・岡 秀夫他(1989)『英語のスピーキング』 東京：大修館書店
- 垣田 直巳・沖原 勝昭他(1991)『英語のライティング』 東京：大修館書店
- 垣田 直巳・青木 昭六他(1992)『英語の評価論』 東京：大修館書店
- 金谷 憲(2009)『英語授業ハンドブック』 東京：大修館書店

- 北尾 倫彦 (2004)『新しい観点別評価問題集 東京：図書文化社
- 酒井 志延 (2012)「測定と診断」JACET SLA 研究会 (編)『新しい時代の英語科教育の基礎と実践』
(pp. 235-244) 東京：三修社
- 土屋 澄男・秋山 朝康・千葉 克裕・望月 正道 (2011)『新編英語科教育法入門』 東京：研究社
- JACET 教育問題研究所 (2012)『新しい時代の英語教育の基礎と実践：成長する英語教師を目指して』東京：三修社
- 高梨 康雄・高橋 正夫 (2007)『新・英語教育概論』 東京：金星堂
- 古谷 貴雄 (2013)「テスト作成・評価法」石田 雅近・小泉 仁・古谷 貴雄 (編著)『新しい時代の英語教育の実践：グローバル時代の人材育成を目指して』(226-239) 東京：金星堂
- 望月 昭彦 (2010)「測定と評価」『改訂版 新学習指導要領にもとづく英語科教育法』(pp. 170-181)
東京：大修館書店
- 文科省 国立教育研究所 教育課程研究センター (2012)『評価規準の作成、評価方法の工夫のための参考資料：高等学校 外国語』 東京：教育出版 望月 正道 (1998)「日本人学習者のための英語語彙サイズテスト」『語学研究所研究紀要』第 12 巻, 27-53.
- 望月 昭彦・相澤 一美・投野 幸夫 (2010)『英語語彙の指導マニュアル』 東京：大修館書店
- 森山 善美 (2009)『教室におけるリスニング指導』 岡山：大学教育出版
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Meara, P. (1992). *EFL vocabulary test*. Center for Applied Language studies, University College Swansea.
- Nation, I. P. S. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paribakt, T., & Wescher, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11,9-29.
- Read, J. (1998). Validating a test to measure depth of vocabulary Knowledge. In A. Kunnan (Ed.), *Validating in language assessment* (pp. 41-60). New Jersey: Lawrence Erlbaum.

