

# 実践と研究をつなぐライティング評価法

大 澤 真 也

(受付 2006 年 5 月 10 日)

## 1. は じ め に

ライティングの評価は、多肢選択型のように間接的にライティング能力を測定するものから、実際に産出されたテキスト、つまり運用を直接評価するものへと変化してきた。しかしながら、能力は目に見えないものであり推測せざるをえない。その目に見えないものをいかに正確に捉えるかが焦点となりライティングの評価法は変化してきている。

一方、ライティングの指導においては産出されたもの、言い換えればテキストにおける言語的特徴などを指導することから、ライティングを産出するプロセスに焦点を当てたプロセス・アプローチが特に大学英語教育においては主流になりつつある。このような状況を考えると、プロセス・アプローチに基づいて指導するのであれば、プロセスも考慮に入れた評価を行わなければいけないのは当然である。しかしながら、教育者と評価法を作成・検証する研究者の間には依然として深い溝があり、既存の英語ライティング評価法は適切なものであるかについては疑問が残る。そこで本論文では、既存のライティング評価法について要約したのち、日本の英語教育研究においてはどのような手法が採用され議論されているかについて概観する。その後、既存の評価法の問題点を指摘した上で、現在注目を集めつつあるポートフォリオ評価を中心として、研究と実践をつなぐ新たな評価法作成へ向けての提言を行うこととする。

## 2. ライティングにおける評価法

ライティング評価においては、現在運用を焦点にしたものが主流になっ

ているが、それらは主に (1) 全体的評価, (2) 分析的評価, (3) 特定要因の評価,あるいはそれらを組み合わせたものに分類することができる。全体的評価とは、各段階において達成すべき能力の記述がある場合も含め、評価者の印象に基づいて段階的に評価をする方法である。代表的なものとしては TOEFL のライティングテスト (TWE) の評価基準をあげることができる。この評価においては、タスクを適切に理解し課題を遂行しているか、主題を支える細部の適切さ、構成、言語使用の適切さ、文法・語彙の多様性の観点から 0 - 6 段階で総合的に評価を行う。次に分析的評価においては語彙、内容などライティング能力を構成していると思われる下位項目を設定し、それらを個別に採点する。よく言及されるものとしては Jacobs et al. (1981) による ESL Composition Profile を挙げることができる。ここでは、内容30点、構成20点、語彙20点、言語使用25点、綴りなどの構造5点の計100点満点で評価を行う。最後に特定要因の評価においては、それぞれのタスクに応じて基準及び焦点を変化させる。例えば構造の複雑さの発達に焦点を当てる場合には、t-units などの指標を用いて評価を行う。

ライティングの評価は状況や目的に応じて変化する多種多様なものである。具体的にはライティングの評価は、(1) クラス分けなどの管理、(2) 指導、(3) 研究のため、の3つの目的に応じて使い分けられる。Weigle (2002) によれば、研究などにおいて大規模の被験者を調査する場合には信頼性と実用性が重要になり、教室などにおいて評価を行う場合は評価項目の妥当性、実生活と関連しているか、プロセスを効率的にこなしているか、指導の効果がどれだけあったか、などの点が重要になる。一般的には、前者においては全体的評価、後者においては分析的評価が主に用いられる傾向にあるが、ライティングの評価は学習者の能力を正確に測定するため、というよりも評価を行う目的に適切なものを適宜選択していると言えるのかもしれない。

### 3. 日本における現状

日本においても、ライティング評価法を用いた研究が数多く行われてきている。そこでここでは、日本の英語教育研究における主要学会誌である JACET Bulletin, ARELE に投稿された過去 5 年間の研究を振り返り、(1) 全体的評価、(2) 分析的評価、(3) 特定要因の評価、の 3 つの観点から概観していく。

#### 3.1 全体的評価

全体的評価を用いる研究の傾向としては、大規模な実験や調査など評価法の効率性が求められる場合、また一般的なライティング能力を明確に定義することは難しいが、ライティング能力を表す 1 つの尺度として使用する場合などである。

例えば、宮迫 (2000a) は、学習者の英語ライティング能力を明らかにするために調査を行った。被験者は高校生 37 名で、彼 (女) らを実力テストによって上位、下位グループに分類し、一貫性 (coherence) をトピックの構造分析 (Topical Structure Analysis) を用いて調査した。具体的な手順としては、被験者は教室内で 45 分以内に辞書を用いなくて、好きなトピックを選択しエッセイを書いた。そして調査者、ALT の 2 人が全体的印象法に基づき 10 点満点で採点し、平均点を算出した。その結果、制限はあるものの、英語力の上位、下位及び全体的評価法に基づいたスコアとの相関が見られ、上位、下位グループの間には一貫性に関して差がみられることが明らかになった。また宮迫 (2000b) では収集されたライティングのサンプルを別の角度から分析し、ライティングのスコアと t-units の関連等が明らかになった。

伊藤 (2004) はアカデミック・ライティングの知識を持たない 262 名の大学生を対象に、L1 (日本語)、L2 (英語) ライティング、L2 熟達度の関連を探った。L2 熟達度は TOEIC、L1・L2 ライティングは TWE に基づき論説文

(argumentative) のものを書かせた。TWE に倣い、制限時間は30分、及び辞書の使用は不可とした。また、トピックによる影響を排除するため、L1・L2 におけるトピックは同一のものとした。ライティングは、2名の評価者の全体的評価の平均点を算出し、2点以上の差がついた場合には第3者によって判定された。日本語は Carson et al. (1990) に基づく5点満点、英語は TWE に基づく6点満点の基準を採用した。その結果、3者には相関関係があることが明らかになった。

これらの研究においてはタスクの制限や採点者間信頼性など、分析的評価と比べると曖昧になりがちな評価法の信頼性を高めることに主眼が置かれている。その上で、ライティング能力を全体的評価法を用いて定義し、特定要因の評価などを行うことで、全体的評価では見逃しがちなライティング能力の構成要素を探ろうと試みている。

### 3.2 分析的評価

ライティング能力は幾つかの下位項目から構成される複合体であると仮定すると、全体的評価を行うのではなく、分析的評価を行う方が望ましいのかもしれない。ここでは、全体的評価と比べて、小規模から中規模の被験者を対象に評価が行われる場合が多い。

広瀬 (2001) においては、説得文 (persuasive text) において L1・L2 の構成に違いが生じるかどうかを11名の大学生を被験者として調査した。被験者はまず CELT に基づいて英語力上位、下位グループに分類された。その結果、英語力はライティングの経験と関連があることがわかった。その後、タスクの刺激として、「読者欄に投稿すると仮定して」制限時間30分以内にライティングを行わせた。ライティングは英語母語話者2人によって、ESL Composition Profile に基づき採点され、2人の合計点を算出した。また被験者がタスクを遂行する様子を録画し、それを元にインタビューを行った。収集されたサンプルは Kubota (1992) に基づき主題の位置、巨視的な修辞パターン、要約部の有無の3つの観点から分析された。その結果、

L1 あるいは L2 の熟達度によって、L1・L2 において産出されたテキストには様々な共通点や相違点が存在することが明らかになった。

大嶋（2000）は、43名の高校生を対象に調査を行い、評価法自体に関する考察を行った。被験者は2つのタスクをそれぞれ15分で辞書を用いずに行うように指示された。収集したサンプルはタスクの理解度、構成、文間の連結性、統語および語彙の多様性に基づき、5点満点の全体的評価を行い、上位、中位、下位グループに分類した。その後、英語母語話者一名によって内容、修辭的構成、論理的連結性、文の容認性、語彙の多様性の5項目についてそれぞれ3段階（-1, 0, 1）で評価を行った。その結果、良い書き手は内容、語彙の多様性において優れていること、また修辭的特徴や語彙・統語的多様性がライティングの質を左右していることが明らかになった。

大嶋（2002）においては評価者の意思決定について調査が行われている。ここでは9名の被験者が参加し TWE の手法に倣い、比較、および論説文を書かせるようなものを刺激として与え、4つの異なったトピックに基づいて20分以内にエッセイを書かせた。エッセイはまず2名の評価者により、全体的評価法を用いて5点満点で採点された。その後、1名が適切な言語使用、文法や語彙、綴りなどの慣用法、内容の構成や提示の方法、トピックの展開の4つにおいてそれぞれ3点満点で評価した。また英語母語話者としてそれぞれの項目において英語母語話者らしいものか、あるいはそうでないかを評価した。その結果、評価者は言語的、またライティングに関する専門知識に基づいて評価を行うが、特に後者の影響を受けて評価を行っていることが明らかになった。

これらの研究においては、分析的評価のみを用いるのではなく、全体的評価あるいは特定要因の評価などと組み合わせて調査を行う場合が多い。分析的評価の特徴として、ライティング能力を幾つかの下位項目に分類しているため、日本の英語学習者の長所、弱点を明らかにすることができるという点が挙げられる。しかしながら、分析的評価のみを用いる場合は稀

である。このことは、分析的評価法の妥当性の無さを露呈しているようなものであり、分析的評価と言えども学習者のライティング能力を包括的には評価できていないとも言える。

### 3.3 特定要因の評価

特定要因の評価においては、評価者が関心を持っている要因に焦点を絞り測定する。例えば、川口 (2000) は40名の大学生を対象に、アンケートによって過去形の知覚を調査し、その上で過去形の産出を促すようなタスクを課し、過去形の知覚と産出の関連を探った。その上で、間違いの種類を分類している。その結果、知覚に関しては英語母語話者とあまり差が見られないが、日本人英語学習者は産出において過去形の代わりに現在形を多用するなどの差がみられた。

Leishman & Nishigaki (2001) はライティングの授業を受講する大学生17名を対象に1年間の授業履修前後にライティングを行わせた。トピックは各自が選び、比較対象のエッセイを書いた。教室内ではなく、家に持ち帰ってライティングを行う方法を採用したため、時間や辞書使用などの制限は無い。評価は2名の教師によって行われ、Nishigaki & Leishman (2001) で明らかになったライティング能力の発達に重要な項目について評価した。具体的な項目は、トピックの選択、構成 (A-D の4段階で評価)、パラグラフ (各パラグラフに何文含まれているか)、つなぎ言葉 (分類し数をカウント)、文体 (Biber (1988) の分類に基づきカウント) であり、分析の結果、幾つかの項目において向上がみられた。

矢亀 (2005) は学生同士の意見交換 (peer response) が学習者のプロセスに与える影響を探る目的で調査を行った。被験者はライティング授業を受講している26名の大学生で、彼 (女) らはパラグラフを書くプロセス等について指導を受けた。意見交換の場においては、それぞれが相手の作文の好きな箇所、もう少し詳しく述べれば良くなると思うことについてコメントした。コメントは先行研究に基づきカテゴリー化され、量的には語

数、T-units の長さを調査した。また質的なものとして推敲の種類、また語、節、句、文の修正頻度と種類が調査された。その結果、文を付け足すという行為が推敲においては最も多いこと、また指導者の助言がなくても効果的な推敲が行えることなどが明らかになった。

特定要因の評価においては、評価者が興味を持っているもの、主に良い書き手とそうでない書き手との間にある違いは何かについて、主にテキスト上の特徴に注目した研究が行われている。また少ないながらも、指導の効果やプロセスを測定しようとした研究も幾つか行われている。しかしながら、それは学習者の英語能力のごく一部分に限定した評価法であり、また質的に評価するか量的に評価するかなどは個々の研究の目的に大きく依存している。そのため、依然として信頼性および妥当性のある評価法が確立していないという問題点も存在する。

### 3.4 評価における問題点

日本の英語教育研究においては、どれか一つの評価法が単独で用いられる場合は稀で、それぞれの評価法における長所、短所を考慮に入れて2つ以上の評価法を組み合わせる用いる場合が多い。共通する点は、評価の信頼性および妥当性を高めようと試みているところであろう。これら2つの項目を達成するため、評価者間信頼性を高める、あるいは TWE など標準的とよばれるテスト形式を用いるなどの工夫が行われている。けれども、これらを達成するのは容易ではない。例えば、工藤・根岸（2002）は、採点者の人数、及び印象的採点法（6点満点）、全体的採点法、分析的採点法（ESL Composition Profile）の関連を調査した。その結果、ある程度の信頼性（約0.6以上）を得るためには、印象的採点法の場合は5人、全体的採点法の場合は6人、分析的採点法の場合は3人の評価者が必要であることが明らかになった。彼らの研究において収集されたエッセイの平均語数は84語であり、この語数で採点された点数が信頼できるものかどうかなど幾つかな問題点も存在するが、もしも彼らの結論が正しいものと仮定すれば、

多くの研究は成立しないことになる。なぜなら多くの場合においては、1人、多くても2, 3人の評価者によって評価が行われているからである。

また、いわゆる標準的なテストに従うことによって妥当性が高まるかという点にも疑問が残る。例えば、石川 (2005) は高校生54名を対象にタスクの複雑さの違いによって、彼 (女) らの産出するライティングの質に変化が生じるかどうかを調査した。ここで用いられたタスクはコマ割りされた漫画を描写するもので、2グループに分けられた生徒が行ったタスクの違いは、活動途中に漫画を見ることができる、あるいは見ることができない、という非常に些細なものであった。けれども、前者の方が後者に比べて産出されたライティングの質が高いという結果になっている。このような結果を考慮に入れると、厳密な意味での妥当性の達成は非常に困難であると言わざるを得ない。

#### 4. 実践につながる評価へ

これまで用いられてきた多くの評価法においては、大規模の調査を行うことを考慮に入れて信頼性、妥当性を高めることに努力が注がれていた。しかしながら、以上のようなことを考慮に入れると、信頼性、妥当性という概念あるいは定義自体が曖昧なものであるのかもしれない (Huot, 1996)。また大きな問題点の1つは、既存の評価法の多くは研究あるいは評価のための評価法であり、それらの結果に基づいて学習者への指導に生かすなど実践への配慮に欠けていることである。そこで、ここでは新たな評価法作成への第一歩として (1) 個別の状況に応じたタスク及び評価法, (2) プロセスを考慮に入れた評価法, の2点について考えてみたい。

##### 4. 1 個別の状況に応じたタスク及び評価法

これまでの多くの研究においては脱文脈的, すなわちできるだけ状況が運用に及ぼす影響を排除し, 結果を一般化できるように考えられてきた (Schoonen, 2005)。しかし一方ではライティング能力は状況に深く依存する



固有の能力であるという仮定に基づき、独自の評価法を作成する試みもある (Brown & Bailey, 1984; Novak, et al. 1996)。社会文化的な視点から考えれば、ライティングは状況に依存する社会的な行為であり、状況に応じて産出されるライティングの質も異なってくると考えるのが自然である (Enk, et al. 2005など)。このように考えると、ライティングの評価においては、過度に信頼性、妥当性を確立し一般化可能性を探るよりも、それぞれの状況に応じたタスクを作成し評価法を考える方が良いのではないだろうか。多くの研究で採用されているタスクは、他の活動と何のかかわりもない、また明確な目的のないものである。数少ない例外としては、JACET 北海道 SIG on CALL (2000) が行った調査が挙げられる。この調査においては、118名の被験者を対象に、10週にわたり、世界中の相手と email によってやり取りを行うというタスクを設定している。

このような、実生活に関連した明確な目的を持ったもの、あるいは他の技能と有機的に結びついたタスクを用い、そのタスクに適合させた評価法を用いる方がより適切にライティング能力を測定することができるのではないだろうか。

#### 4.2 プロセスを考慮に入れた評価

前述したように、ライティング指導においてはプロセスを重視したプロセス・アプローチが主流になっている。けれども、そのプロセスを評価する試みはあまり行われていない。評価ではないが、ライティングのプロセスを探る試みとして発話思考法を用いる試みは幾つか行われている。例えば、野呂 (2004) は、学習者の自己分析に基づいた教師のフィードバックおよび修正の効果、妥当性を検証した。被験者は24名の大学生で、クラスで扱ったパラグラフのパターンを用いてライティングを行うというタスクを与えられた。また被験者は、メタ認知発達のため、内容、構成、言語形式、全体的な質について自己分析するよう求められた。2ヶ月の授業期間の前後に、彼らは15分で1パラグラフを書く課題を与えられた。そしてラ

イティング終了後、すぐに自分のプロセスを振り返るように指示された。またメタ認知の知覚を探るために、アンケートも実施した。その結果、学習者の自己分析に基づいたフィードバックはライティングの運用およびメタ認知の発達に効果があることが明らかになった。

ライティングのような認知的活動においてはメタ認知、そして自己評価に基づいた推敲などの役割が重要であると考えられており、これらのプロセスを適切に判断できる評価法を確立することが望ましいと考えられる。

#### 4.3 ポートフォリオを含めた新しい評価法にむけて

以上のことから、研究と実践を関連づける評価法としては、状況及びプロセスを考慮に入れたものが適切であると考えられる。これら2つを考慮に入れた評価法として考えられるのは近年注目されつつあるポートフォリオ評価である (Jenkins, 1996; Hamp-Lyons & Condon, 2000)。Weigle (2002) によれば、ポートフォリオには様々な定義があるが、共通するキーワードは収集、省察、選択である。つまり、時間制限がある中で書かれた1つの作品だけを評価するのではなく、ポートフォリオは場合によっては草稿段階のものも含めて複数の作品を含んでいる。そのため、教室内で書いたもの、あるいは教室外で書いたもの、ジャンルやタスクなど様々な変数を考慮に入れた上で評価を行うことができる。次にポートフォリオにおいては学習者の省察及び選択が含まれる。教師が課した課題の全てを含むことももちろん可能だが、ポートフォリオにおいては学習者自身が完成度の高いと思うものを選び提出することができる。つまり、学習者は客観的に自分自身の作品を評価する必要がある。また多くのポートフォリオには省察エッセイと呼ばれるものが含まれており、学習者はなぜその作品を選んだか、自分の長所は何か、自分の能力はどのように発達したか、などをできるだけ詳しく説明する必要がある。このように学習者自身が自分自身の作品を客観的に見つめ直す活動を含むことによって、学習者はより効率的にライティング能力を発達させることができ、教師はより公平に学習者

の能力を評価することができる。

## 5. 新しい評価法

学習者のライティング能力は一般化できるものではなく領域固有で柔軟なものであるという前提に基づくと、より広範囲にわたって包括的な評価をしていく必要がある。その可能性の1つとして考えられるのがポートフォリオ評価であるが、その実施には幾つかの問題もつきまとう。

まず第一にポートフォリオ評価においては明確な基準が存在しない。ポートフォリオはあくまでも学習者の作品をファイルに綴じるという性質のものであるため、どのようなものをポートフォリオとして定義すれば良いのか曖昧である。特に外国語としての英語ライティングにおけるポートフォリオに関しては、峯石（2002）が実際に日本の大学英語教育においてポートフォリオを活用しその可能性を探っているが、まだ発展途上のものであると言わざるを得ない。

次に評価基準に関しても意見が分かれる。全体的評価法を用いるのか分析的评价法を用いるのかはそれぞれの状況に応じて適切に見極めなければいけないであろう。特に前者を用いた場合はそれほどでもないが、後者を用いた場合、学習者の能力を適切に評価できる一方、費やさなければいけない時間は増大してしまう。特にクラスサイズが大きい場合などにおいては、今後の課題であると言えよう。

ここでは研究と実践が結びついた評価としてどのようなものを行えば良いか具体的に考えてみたい。

### （1）タスクの実用性及び再帰性

タスクを設定する際に考慮するものとして、時間制限の有無、教室内／外、他の技能との関連性など様々な要素がある。学習者のライティング能力を包括的に判断するためには、単一ではなく複数のタスクを課す方が望ましい。また学習者の自律性を高めるため、peer feedback や電子掲示板を

利用した議論などを取り入れるのも良いであろう。またライティングを単独の活動として終わらせないために、リーディング活動等と組み合わせるのも面白いのではないだろうか (図 1)。このようにタスクを他の技能と結びつけたより実用的なものに、また 1 度の活動で終わるのではなく再帰的に行われるものにすることで、学習者の能力をより適切に測定できる評価法へつながると考えられる。

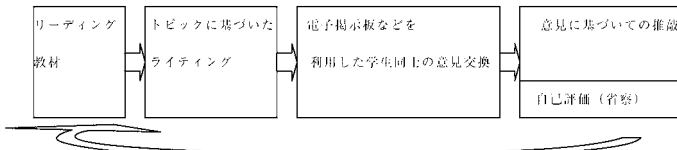


図 1：ライティングタスクの設定例

## (2) 学習者の自律性を高める工夫及びプロセスの評価

ライティングにおいてよく用いられる表現、文章構成等を明示的に教えることは可能だが、最終的には学習者自身が客観的に自分の作品を見直す、言い換えればメタ認知を適切に働かせることが最終的な目標となる。そのためには、タスクの設定や評価においてもそのことを考慮に入れなければならない。またプロセスを評価するためには発話思考法などを用いるのが最適かもしれないが、実現可能性の高いものとして学習者に自分のプロセスを省察させ、それを評価するという方法を取り入れるべきであろう。例えば、Jacobs et al. (1981) に基づき、内容、構成、言語使用、語彙、綴りなどについて学習者が自分の産出したテキストにコメントを加えるという方法が考えられる。

## (3) 評価基準の設定

全体的評価、分析的評価のどちらを採用するかは個々の状況に依存するが、いずれにせよ明確な評価基準を設定することが必要になる。ライティ

ング能力としては主に、言語的能力、談話的能力、社会言語的能力、そしてプロセスに焦点を当てた能力としては、タスクに対する理解度、推敲の成功度などを取り入れると良いであろう。これらの評価基準を学習者にも提示することで、学習者の自覚につながりライティング能力の発達を促すことができるかもしれない。

研究と実践をつなげる評価法としては主に以上の3点について考える必要がある。評価法自体の精度を上げるのではなく、学習者のライティング能力の発達に寄与できる評価法を考えることによって、実践および研究に有益な情報を与えることの出来る評価法へとつながるのではないだろうか。しかし、これらはあくまでも実践の場で試してはじめてより具体的な評価法へとつながるものであり、今後の研究課題としたい。

## 6. お わ り に

現在に至るまで評価法は大規模での調査を目的として、いかに信頼性、妥当性を高めていくかに焦点があてられてきた。しかしながら、実践者の視点からすればそれはあくまでも静的な能力を測定しているにすぎない。学習者の能力は日々、また状況に応じて変化、発達していくものであり、学習者の能力の発達を動的に評価するためには、異なった視点からの評価法が必要になると思われる。そのためには、多少の労力が必要にはなるが、ポートフォリオ評価は今後期待を持てる評価法の1つであると言える。今後は授業の観察、教師のニーズ分析、評価を実際に用いての信頼性及び妥当性の検討など、実用可能な評価法として確立していく必要がある。

## 参 考 文 献

- Biber, D. (1988). Variation across Speech and Writing. Cambridge University Press.
- Brown, J. D. & Bailey, K. M. (1984). A categorical instrument for scoring second

- language writing skills. Language Learning, 34 (4), 2-43.
- Carson, J. E., Carrell, P. L., Silberstein, S., Kroll, B., & Kuehn, P. A. (1990). Reading-writing relationships in first and second language. TESOL Quarterly, 24, 245-266.
- Enk, A. V., Dagenais, D., & Toohey, K. (2005). A socio-cultural perspective on school-based literacy research: Some emerging considerations. Language and Education, 19 (6), 496-512.
- Hamp-Lyons, L. & Condon, W. (2000). Assessing the Portfolio: Principles for practice, theory, and research. Hampton Press.
- Hirose, K. (2001). Persuasive writing in L1 and L2: A look at Japanese EFL students' rhetorical organization strategies. JACET Bulletin, 33, 43-56.
- Huot, B. (1996). Toward a new theory of writing assessment. College Composition and Communication, 47 (4), 549-566.
- 石川智仁. (2005). EFL ライティングにおける構造的複雑さの発達指標と熟達度の関係の検証：タスクに基づくアプローチ. JACET Bulletin, 41, 51-60.
- Ito, F. (2004). The interrelationship among first language writing skills, second language writing skills, and second language proficiency of EFL university students. JACET Bulletin, 39, 43-58.
- JACET Hokkaido SIG on CALL. (2000). Corpus-based analyses of E-mail by Japanese college students. JACET Bulletin, 32, 137-150.
- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V. and Hughey, J. (1981). Testing ESL Composition: A practical approach. Newbury House.
- Jenkins, C. B. (1996). Inside the Writing Portfolio: What we need to know to assess children's writing. Heinemann
- Kawaguchi, K. (2000). Japanese EFL learners' understanding and use of the past tense: A study based on questionnaire and analysis of students' narratives. ARELE, 11, 21-30.
- Kubota, R. (1992). Contrastive rhetoric of Japanese and English: A critical approach. Dissertation Abstracts International, 54 (5), 1709A.
- 工藤洋路・根岸雅史. (2002). 自由作文の採点方法による採点者間信頼性について. ARELE, 13, 91-100.
- Leishman, S. & Nishigaki, C. (2001). Style elements in foreign-language writing among Japanese college students. The proceedings of the fourth international conference on foreign language education and technology (FLEAT IV), 665-674.
- Nishigaki, C. & Leishman, S. (2001). Needs analysis, instruction, and improvement in writing class: Developing textual features in Japanese EFL college writing.

JACET Bulletin, 34, 57-72.

峯石緑. (2002). 『大学英語教育における教授手段としてのポートフォリオに関する研究』. 溪水社.

Miyasako, N. (2000a). Topical structure analysis of high school student compositions. JACET Bulletin, 31, 49-64.

Miyasako, N. (2000b). An analysis of high school student compositions in the sentence-based approach. ARELE, 11, 111-120.

Noro, T. (2004). A study of the metacognitive development of Japanese EFL writers: The validity of written feedback and correction in response to learners' self-analyses of their writing. ARELE, 15, 179-188.

Novak, J. R., Herman, J. L., & Gearhart, M. (1996). Establishing validity for performance-based assessments: An illustration for collections of student writing. The Journal of Educational Research, 89 (4), 220-233.

Oshima, H. (2000). On the writing ability of Japanese high school students - A native speaker's perspective on writing. ARELE, 11, 121-130.

Oshima, H. (2002). Assessing communicative writing proficiency - An expert rater's decision making when scoring EFL compositions. ARELE, 13, 81-90.

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. Language Testing, 22, 1-30.

Weigle, S. C. (2002). Assessing Writing. Cambridge University Press.

Yakame, H. (2005). The role of peer feedback in the EFL writing classroom. ARELE, 16, 101-110.

## Summary

### Assessing Writing: Bridging a gap between theory and practice

Shinya OZAWA

In most of the research on writing assessments, the focus has been on how we can establish validity and reliability of the assessments. For practitioners, however, it may not have any useful implications. As is often said, there is a gap between theory and practice. Specifically, practitioners need to have assessments that can evaluate learners' progress and that will lead to learners' autonomy. Recently, portfolio-based assessments have been drawing researchers' attentions. In this article, I categorized the assessments into (1) holistic, (2) analytic, and (3) primary trait assessments, and reviewed previous literatures. Then I outlined the principles of portfolio assessments, and suggested that we should take into account the (1) site-based and (2) process-oriented tasks and assessments. Finally, I showed one example of how the assessment that can bridge the gap between theory and practice would work.