

英語の能力テスト開発の必要性

佐藤 史郎

英語科の教育目標と大学入試の関連

中学校学習指導要領による外国語科の目標は、以下の三つの要素から成っている。(高等学校の学校指導要領は以下の(1)及び(3)の傍線部「基礎的」という表現が削除されている以外は同じである。)

- (1) 外国語を理解し、外国語で表現する基礎的な能力を養う。
- (2) 言語に対する関心を深める。
- (3) 外国の人々の生活やものの見方などについての基礎的な理解を得させる。

(1)の外国語を理解するとは四技能のうち、「聞く・読む」事を意味し、外国語による表現とは、「話す・書く」技能を意味する。また、(2)の「言語」とは特に自国語の事であり、外国語と母国語の言語活動を通して、外国語と母国語の言語構造に於ける相違、及び発想の相違に気付かせる事により、母国語に対する関心を深める事が狙いとされている。(3)では、題材の中で外国の人々の生活(風俗・文化・習慣等)、及びものの見方に関して基礎的な理解を得させる事が主な狙いとされ、この事が同時に、他文化圏の人々の価値感を受け入れる事につながり、究極的には国際理解への橋渡しとなる事が期待されている。いずれにしても、(2)と(3)は言語活動・題材の中で副次的な効果を狙うものであって、あくまでも外国語科(以後、英語)の基本的な目標は、(1)で謳われているように、まず中学校に於いて、英語の四技能に関する基礎的能力を養い、高等学校に於いては、四技能の能力をさらに育て、伸ばす事にあると言えよう。

このような事を踏まえた上で、中学校・高等学校に於ける英語教育の

教育内容と入試（ここでは特に大学入試）の関連について考えてみたい。

まず、理想面から言うと、大学入試では高等学校で習得した四技能の能力（proficiency）をバランスよく測定するような入学試験が望まれるのは言うまでもない。しかしながら、特に、表現する能力のうち話す能力を測定しようとする場合、テストの妥当性・信頼性の維持も重要課題であるが、入試の場合特に経済性（客観性も含む）の面での困難さが、大きな障害となる事は明白である。この経済性の問題は技術的な問題であって、決して解決不可能ではないにしても、入学試験のように時間的制約を受ける場合は、極めて困難と言わざるを得ない。しかしながら、今後話す能力テストを開発しようとする場合、テストが話す能力を真に測定するものであるかと言う問題が生じる。即ちテストに於ける最も重要な要件である妥当性を構成する要素についての合意を得るべく、更に研究を深め、実験の中で検証していく事が、必要であろう。例えば、実際の口頭でのコミュニケーションが聞き手と話し手の間でキャッチボールの形で行なわれる事を考えれば、話し手からの一方的な情報伝達能力の測定のみでは妥当性に欠けるのではないだろうか。（実用英語検定試験の一級二次試験はこの要件をある程度満足させていると思われる。）

また、評価の面では、いかなるコミュニケーションも特定の社会言語的状况の中で行なわれる事から、スピーチレベルを考慮に入れた質問としての発話及び返答としての陳述の適切さ（appropriateness）を評価の中に組み入れる事は不可欠と思われる。さらに、発話内容が文法的に正しくてもネイティブスピーカーが恐らく発しないであろうような表現に対する評価基準の設定も、今後の課題であろう。理解する能力のうち、聞く能力の測定テストに関しては、国立大学主導の形で（共通一次試験の中で）、大学入試の一環に組み入れる事が現在真剣に検討されている。

聴解力テスト実施に伴う資金を始めとして、受験時に客観性を維持する為の技術的な問題等、様々な未解決な課題を抱えつつも、実施の方向へ動いている事は日本の英語教育の改善に大きく寄与するものであると

思われる。

大学に於ける入学試験は一定期間の学習の到達度合を測る為に、教師が作成し、評価を下す実力テスト (Achievement Test) と異なり、志願者が入学後、種々の授業科目で要求されるであろう英語の運用能力を予測しようとするものであると定義づけられる。従って、同一の高等教育機関でも学部・学科によって入学後、学生に要求する英語の運用能力の中味が異なれば入試問題の中味は各々の要求度を反映すべく異って良いはずである。しかし、ここで最も問題となるのは既に述べたように、ある大学または短大の学部・学科で英語の四技能に亘る基礎能力を測定しようとしても、音声面のテスト (聞く・話す) で妥当性・信頼性・経済性を備えたテスト形式が現在のところ存在しないという事である。聴解能力テストの大学入試への導入は、現在検討されている。しかし、話す能力テストの開発、更にその施行となると、どんなに研究が為されても、相当の困難を伴う事は十分に予想される。極めて限られた時間中に測定・評価を行わなければならないと言う制限が、付きまとうからである。

中学校・高等学校に於ける英語教育が、受験体制に組み込まれた形で、ともすれば技能面の基礎力の養成よりも知識面 (文法・語法・語彙・発音・アクセント等) の強化を主眼とした内容になりがちな事は、以前から多くの識者によって指摘されてきた。そこで、推薦入試の中で受け入れ大学の学部・学科が音声面の能力を学科試験もしくは小論文の成績と共に、合否判定の基準に組み入れることができるよう高等学校の調査書の中に英語の総合評価と共に音声面の評価の記載を義務づけたらどうであろうか。もちろん、音声面の評価をどの程度英語全体の評価の中に組み入れるかは、受け入れ校の学部・学科によって異なってよい。話す能力テストに関しては、教師が定期・不定期的に主観テストの形で実施し、その都度、評価結果を記載・集積し、最終的にその平均値を個人の話す能力の評価として算出する方法が妥当性の面で優れている事が既に報告されている。

教師によるペーパーテストを主体とした評価もそれ自体主観性が入り込む余地があるが、話す能力テストの場合は特に、問題作成及び評価の面で主観性が入り込む可能性が、他技能に比べて大きい。そこで、少なくとも高等学校の内部で学年毎に教師が一体となって、より高い妥当性・信頼性を維持できるような問題の作成及び評価の基準の設定に関して協議をする必要があるであろう。

いずれにしても、大学入試の一環として英語の音声面の評価を可否の判定基準に組み入れるとなれば、まず高等学校、ひいては中学校の教育現場に於いて、聞く能力・話す能力を養成すべく教育せざるを得なくなる事が予想され、英語教育が本来あるべき姿に向かって一步前進する事が期待される。

総合能力テスト開発の意義

大学入試に於いては、個々の技能を別々に測定する事は、既に述べたように、極めて困難と思われる。そこで、もし検証の結果、外国語の学習に関する潜在能力を完全ではないにせよ、可能な限り引き出し得るテスト形式が存在すると判明すれば、現行の入試問題の改善に多大な貢献をする事になるであろう。更に、テスト形式が総合能力を測るものである以上（従って、四技能を総合した真の潜在能力の測定される訳であるから）、受験生は特定の受験準備をする事は不可能で、表面的な英語の知識（文法・音韻・語彙）の習得のみでは不十分となり、日頃から習得した知識の技能面での運用能力の向上を目指し、努力する事の必要性を認識するであろう。

また、特に中等教育の場に於いては、英語の到達目標に対する到達度の測定（achievement）と同時に、各学習者が到達した度合が潜在能力（proficiency）として定着したかどうかと言う側面も測定する必要がある。つまり、特定の教師、教授法、言語材料のもとで学習した内容が真の能力に結び付く事の保証はどこにもないからである。

どの教師が用いても妥当性・信頼性・経済性を兼ね備えた技能別能力

テストが用意されていれば、能力定着の度合は、本来は測定可能なはずである。しかし、残念ながら、今のところ技能別の能力テストに関しては、技能によって程度の差こそあれ、今後更に研究・実証していかねばならない問題を少なからず抱えているのが実状である。そこで、その代替手段として、もし英語（他言語にも有効であればテストとしての有効性は更に高まる事になるが）の総合能力を結集させなければ高得点を得られないようなテスト形式が開発されれば学習者に対してより適確な評価を下す事ができるようになるであろう。そして、従来ともすると欠けていたと思われる特定の英語教授プログラム全体の評価（テストのより積極的な使用法）手段として、総合能力テストを使用する事も可能であろう。

クローズ・テストの有効性と実用性について

クローズ・テスト（Cloze Test）のクローズという言葉は、ゲシュタルト心理学の概念で、不完全なパターン（広い意味の情報部分が部分的に欠如している状態）を完全なものとしてみようとする人間の心理から由来している。クローズ・テストはこの概念を言語に応用して作成されるものである。従来、英語のテスト問題の中で用いられてきた穴埋め問題も広い意味ではこの範疇に入るが、クローズ・テストの場合は、従来の穴埋め問題のように穴埋めの方法が恣意的でなく、機械的に何番目かの単語（通常5～9番目）を消去して空欄を設けているところにその特徴がある。通常、空欄に対するヒントは全く与えず、被験者は空欄の前後のコンテキストから判断して適切な単語を記入する。（後に、空欄に対して選択肢を与える客観形式が登場する。）

例えば、次のような二つの文の中には三つの空欄が設けてあるが、被験者は空欄の前後から判断して、when, few, vacation を記入すれば得点が与えられる訳である。

Most people look forward to the summer, () they can take

a vacation for a () weeks. For Americans one of the favorite () spots is Northern Wisconsin, the land of a thousand lakes. (このような形式で空欄が50個所設けられている。因みにこの場合は、文頭から数えて8番目の単語が消去されている。)

これは、一見単純な作業のようであるが、被験者は空欄に記入すべき単語を予想する時、文法・語法・語彙等の知識の助けを借り、前後の意味を勘案したうえで、瞬時のうちに判断を下さなければならない。つまり、正答が機能語の場合は、まず文法的意味を類推し、品詞を決定し、その中から適切と思われる単語を選択する能力を問われる。また、内容語が正答の場合は、機能語を選択する際に求められる能力の他に、場合によっては、かなり広範囲のコンテキスト全体の意味の把握が要求される。(消去された語が文章の後半に位置する場合は、該当する文を含む文章全体の意味を把握していなければ正答を導き出せない場合もある。)通常、消去される単語の数「50」と言う数字は、被験者の潜在能力を引き出すのに必要な量であると同時に、統計上の処理をし易くする為と考えられている。

●総合能力テストとしてのクローズ・テストの実証データ

クローズ・テストは当初、Native Speakers (以下 N. S.) を対象に、英語の散文の文章の難易度 (readability) を測定する手段として用いられた。テイラー (Taylor, 1953) は任意に選んだ三つの文章からクローズ・テストを作成し、N. S. に解答させた。その結果、クローズ・テストの正解数により判明した文章の難易度と、それまで文章の難易度を測る尺度として用いられていたフレッシュ (Flesh) 及びデール・チャール (Dale-Chall) らによる文章の難易度決定方式 (Readability Formula) を用いて算出した難易度と、ほぼ一致するという事を発見した。

テイラーはまた、自動的に何番目かの単語を消去するクローズ・テストの方が、機能語または内容語のみを消去する方法より、統計的に優れ

たデータ（読書力を測るうえで、クローズ・テストによる方が、はるかに能力の差が明確になる）を産み出すことを指摘している。

当初、クローズ・テストは N. S. を対象に文章の難易度、及び N. S. の読書能力を測定する為に用いられたが、その後、エストラダ (Estrada, 1969) やクロフォード (Crawford, 1970) 等によって、英語を第二外国語として使用する子供を対象に、英語の読解能力の測定にも用いられるようになった。これらの実験が行なわれた時期とほぼ平行して、ダーネル (Darnell, 1968) はクローズ・テストが外国人留学生の為の能力テスト (TOEFL) との間で高い相関関係 (+0.82) があり、とりわけ TOEFL 中の聴覚力テストと高い相関関係 (+0.73) がある事をつきとめた。

その後、幾つかの実験により、クローズ・テストが N. S. の読解力を測定する手段として有効な事が検証され、次いで、クローズ・テストのデータが語彙力・聴解能力等と相関関係の高い事が判明した。

一つの興味深い実験がテイラー (Taylor, 1956) によって為された。それは、同じ被験者に同じ文章から成るクローズ・テストを一定期間経過後に与えた場合、一回目の平均点よりも二回目の平均点が高いと言うものであった。この事から、テイラーはクローズ・テストが単なる個人の言語能力のみを測定するものでなく、知識や記憶の長さ (クローズ・テストの単一の施行の場合も影響を及ぼすと考えられる)、及び一般的知識等から成る総合力を測定するものであろうと指摘している。

転換期を示唆するとも言えるダーネルによる実験に続いて、オラーとコンラッド (Oller & Conrad, 1971) はクローズ・テストを米国在住の留学生とアメリカ人大学生 (大学院生も含む) から成る幾つかのグループに与え、それらグループ間の弁別力を調べる事によってクローズ・テストが総合能力を測定し得るかと言う疑問に答えようとした。(Table I)

Table I の中で興味深いデータは IV グループ (Advanced Composition) と VI グループ (Freshman Composition) の間では平均点の差が殆んどなく、同じ N. S. グループでも VII グループ (Graduates in TESL)

TABLE I
MEAN SCORES OF SUBJECTS ON CLOSE TEST¹⁾

	Group	N	Mean Score	Standard Deviation	Range
Non-native Subjects	I Beginning ESL	18	7.00	3.88	1-15
	II Intermediate ESL	26	11.42	4.27	6-18
	III Advanced ESL	26	18.38	4.54	14-30
	IV Advanced Composition	23	21.00	4.11	14-28
	V Graduates in TESL	9	20.92	4.68	12-29
	VI Freshman Composition	26	21.88	4.04	16-28
	VII Graduates in TESL	14	32.50	3.71	27-38

の方がVIグループ (Freshman Composition) よりはるかに高い平均点を示していると言う事であろう。このデータから推察される事はクローズ・テストが学習者の側に、単なる読解力のみならず、長く維持できる記憶力 (longer memory span)、演繹力、帰納力、推理力、分析力、総合力等を要求するかもしれないと言う事である。

クローズ・テストの妥当性 (validity) を調べる為に、オラーとコンラッド (Oller & Conrad, 1971) はクローズ・テストとカリフォルニア大学での留学生の為の英語能力テスト (UCLA ESL Placement Examination) との間の相関係数を算出した。その結果は、Table IIの通りである。

クローズ・テストと読書力との高い相関係数 (+0.80) は十分予想されるものであるが、書き取り (dictation) との間の最も高い相関係数 (+0.82) は、両者の言語の伝達手段の相違 (クローズ・テストが文字

TABLE II
 MULTIPLE REGRESSION ANALYSIS OF CLOZE
 SCORES WITH THE VARIOUS PARTS
 OF THE UCLA ESLPE²⁾

COEFFICIENT OF DETERMINATION							.77
MULTIPLE CORRELATION COEFFICIENT							.88
Subtest :	Vocab.	Rdg.	Gram.	Art.	Dict.	Cloze	
Vocabulary	1.00	.66	.69	.20	.59	.59	
Reading		1.00	.68	.10	.80	.80	
Grammar			1.00	.21	.60	.58	
Article				1.00	.17	.33	
Dictation					1.00	.82	
CLOZE TEST						1.00	

言語であるのに対して、書き取りは音声言語)を考慮に入れる時、一見、意外な感がある。しかし、書き取りで高得点を挙げる為には、単なる聴解能力のみでは不可能で、既に音声で与えられた情報をもとに先を予想する能力 (expectancy) を必要とする事は、ほぼ間違いないであろう。そして、適確な予想をする為には、クローズ・テストで要求されるような、かなり高い次元の思考力を必要とするのではないだろうか。

ここで重要と思われる事は、今後クローズ・テスト以外の総合能力テストを開発する場合、被験者に提示される言語の伝達手段が音声であるか文字言語であるかと言う事は、それ程問題ではなく、被験者に要求する能力が単なる言語能力のみでなく、より複雑で高度な思考力を必要とするものを盛り込む事が、必要条件となるのではないかと言う事であろう。この事は、クローズ・テストとカリフォルニア大学の英語能力テス

トの各部門テスト (Subtest) との相関係数が+0.82 (対書き取り), +0.80 (対読解力), +0.59 (対語彙力), +0.58 (対文法), +0.33 (対冠詞) となっているように, 解答を引き出すに至る思考過程が単純な程, 相関係数が低い事からもある程度説明が付くかもしれない。

●クローズ・テストの採点法

N. S.を対象にクローズ・テストを施行する場合は, オリジナルの文章に使われた単語のみを正解とする方法 (the Exact-word Scoring Method)の方が, 文脈から判断して妥当と思われる単語をも正解とする (the Contextually-acceptable-word Scoring Method) 方法よりも能力差をより鮮明に表わす点で, より優れている事がテイラー (Taylor, 1953) 等によって報告されている。

しかし, クローズ・テストを英語を母国語としない人を対象に使用する場合, 文脈上, 適切と考えられる答えも正解とする採点法の方がクローズ・テストの得点と英語能力テストの総合点との間で, より高い相関係数を生み出す事が, スタップズとタッカー (Stubbs & Tucker, 1974) 等によって報告されている。

英語を外国語として教える教師が, N. S.の助力なしに採点する場合, 文脈上適切と思われる複数の解答を用意する事は極めて困難な作業と言わざるを得ない。この問題に答えるかのように, スタップズとタッカー (Stubbs & Tucker, 1974) 等は, 両採点法の間には+0.97と言う非常に高い相関関係がある事を発見し, クローズ・テストの正解をオリジナルの文章に表われた単語のみを正解とする方法も, 文脈上可能な複数の答を正解とする方法同様に有効である事を指摘している。この事に関しては, オラー (Oller, 1974) も同様な内容の報告をしている。以上の実験に基づく報告は, 日本人教師がクローズ・テストを使用する際に生ずる最も大きな問題を解決したと言う意味で, クローズ・テストの実用性を, 更に一層高めたと言えるであろう。

●クローズ・テストを英語能力の向上を測定する尺度として使用する事の可能性

これまではクローズ・テストが総合能力テストとして使用し得るかと言う点を主眼に、クローズ・テストを様々な角度からみて来た。クローズ・テストが、もし学習者の総合能力を測定する為の有効な手段であるとするれば、学習者が外国語を一定期間学んだ後に、クローズ・テストを与えた場合、クローズ・テストの得点が増加すると予想される。この点に関して、ブリール他 (Briere, et al., 1978) は、ドイツ語、日本語、ロシア語、スペイン語を学ぶアメリカ人に各々の言語で書かれた文章 (同じ程度の難易度を持つ内容) をもとにクローズ・テストを作成し、各学期末に与えた結果、クローズ・テストの得点が学期の始めに比べて増えている事を明らかにしている。(Table III)

TABLE III
THE MEAN NUMBER OF CORRECT RESPONSES IN THE
CLOZE TESTS FOR EACH LANGUAGE³⁾

Language	Semester		
	I	II	III
German	5.80	15.55	17.17
Japanese	2.35	15.25	25.00
Russian	1.45	8.45	14.71
Spanish	6.44	17.04	22.33

この実験は、三つの各々別の語族を代表とする言語 (ゲルマン、スラブ、ロマンス) と、これら三つの語族と全く異なる言語 (日本語) に対してもクローズ・テストが十分使用し得ると言う事を証明するうえで、極めて大きな意義を持つものと思われる。

結 論

理想的には、大学入試に於いて英語を入試科目として志願者に課す場合、英語の潜在能力を測る為の技能別テストを施行する事が望ましい。しかし、現時点では、四技能のうち、とりわけ信頼性・妥当性・経済性を兼ね備えた「話す能力テスト」は存在しないと言っていいであろう。そして、将来、「話す能力テスト」が開発される可能性はないとは言えないまでも、今後も恐らく最も困難な課題として残るであろう経済性及び客観性の解決には、なお、かなりの時間を要すると思われる。しかし、大学入試(高校入試も同様)の場で、音声面の能力の査定を全くせず、現状のままの測定方法を続行して行くとすれば、中学校・高等学校の英語教育が音声面の能力を含めた四技能をバランスよく向上させるよう、根底から変わっていく事は期待しにくい。

そこで、聞く能力と話す能力を何らかの形で、大学入試の合否判定基準に組み入れる事ができるようにすれば、完全とは言えないまでも、少なくとも以前よりは、中学校・高等学校に於ける英語教育の内容が、知識としてでなく技能としての聞く能力、話す能力の向上を目指す方向へ変わって行くのではないかと言う期待が持たれる。

その為の具体案の一つとして、推薦入試の合否判定の資料として活用できるように、調査書の中に音声面の能力の評点を記入させると言う提案をしてみた。この提案の根拠の一つとして、音声面の能力のうち、特に話す能力の測定に関しては、教師が通常期末テストとして与える単一施行のテストで評価を下すよりは、主観テストを学期中何回かに分けて、話す能力を様々な角度から測定しておき、それらの評価の平均値を最終評価として出した方が妥当性の面で優れていると言うデータが報告されている事が挙げられる。

今後、話す能力テストを開発するに当っては、話す能力を構成する要素を何に求めたらよいか(この事は同時に、評価基準の設定に結びつく)と言う問題、即ち話す能力テストに妥当性を盛り込ませる為の最も基本

的な要素に関して、更に研究し、実証して行く必要がある。例えば、従来は評価の中心が被験者の発話のうち、音声面（発音・ストレス・イントネーション）の正確さ、文法的正しさ、適切な語彙の選択の度合い等に置かれていた感がある。だが、近年になって、被験者に要求する能力レベルに関係なく、相手の発話や質問に対する発話あるいは返答の適切さや、発話の文法的正しさよりも語法上の適否等を中心に、測定・評価すべきとの声が高まっている。

この事と平行して、被験者の潜在能力を最も効率よく、効果的に測定する為には、どのようなテスト形式（指示を与えて答えさせる形式、絵を提示して課題を与える方式、会話方式、スピーチをさせた後に質疑応答をする方式、言語の余剰性を利用して、雑音の中で会話を行なう方式等）が望ましいか、また被験者が内在させている発話の多様性を引き出す為には、どのようなテスト形式がふさわしいか等についても、今後研究の余地がある。

学習者の潜在能力を最大限引き出すべく作成される能力テストは、入試や各種検定試験（文部省認定実用英語検定試験や TOEFL がこれに属する）に於いて、その機能が発揮される。しかし、能力テストは同時に、学習者の四技能の熟達度を測定するうえで、重要な役割を担っていると言える。基礎的な文法、語彙、音韻に関する知識なくしては、各技能の能力の育成は不可能だが、学習者がこれらの知識をどの程度、技能として活用し得るかと言う側面をも教師は測定する必要がある。即ち、特定の教師のもとで、特定の言語材料、題材を用いて学習した内容が真の潜在能力（既習の知識を未知の場面で技能として活用し得る度合）に結び付く事の保証はないからである。

能力テストは、このような二つの目的に対して効力を発揮する訳であるが、現状では、既に述べたように、妥当性・信頼性・経済性を兼ね備えた技能別テスト（特に音声面の能力）を被験者に同じ条件のもとに施行できる状態にはない。このような状況の中から、クローズ・テストが学習者の総合能力を測定する手段として脚光を浴び、およそ15年の間に、

その有効性が数多くの実験で実証されてきた。

クローズ・テストの最も大きな特徴は、総合能力テストとしての妥当性・信頼性に優れている事他に、特に経済性に優れている事であろう。即ち、問題作成とその評価に要する時間が他の能力テストに比べ、極めて短くて良いと言う点にある。もちろん、クローズ・テストを作成する際、文章の内容が特定の知識に片寄ったり、被験者の英語能力とかけ離れていると、総合能力テストとしての妥当性に欠ける事になるので、文章の選択の際には慎重な検討が必要であろう。

いずれにしても、クローズ・テストを大学入試の英語の問題の一環として採り入れる事によって、志願者の英語の潜在能力をより効果的に測定する事が可能と思われる。また、中等教育の現場に於いて、一定期間内に学習者がどの程度潜在能力を身に付けてきているかを測る手段として用いる事によって、従来より、一歩進んだ評価を下す事も可能になるであろう。クローズ・テストが総合能力テストとして最も有効な測定手段の一つであるとすれば、生徒の熟達度を測る尺度として貴重な資料を提供してくれるのみならず、教える側からみると、教師が学年初めに掲げる指導目標とそれを達する為の教授法、教材を含めた英語教授プログラム全体が妥当であったかどうかを判断する為の材料にもなり得ると言う事を指摘して、この稿を終りたい。

Notes

- 1) John W. Oller, Jr. and Christine A. Conrad, "The Cloze Technique and ESL Proficiency," *Language Learning*, 21 (December, 1971), 190.
- 2) *Ibid.*, p. 190.
- 3) Eugene J. Briere, et al., "A Look at Cloze Testing across Languages and Levels," *Modern Language Journal*, 62(March, 1978), 25.

References

- Alderson, J. Charles and Hughes, Arthur. ed. *Issues in Language Testing*. London : The British Council, 1981.
- Darnell, D. K. "The Development of an English Language Proficiency Test of

- Foreign Students Using a Clozentropy Procedure." *ERIC ED* 024-039, 1968.
- Kakita, Naomi. ed. *Eigoka no Kenkyu*. Tokyo : Kyodo Shuppan Publishing Company, 1982.
- Levenston, E. A. "Aspects of Testing the Oral Proficiency of Adult Immigrants to Canada." *Papers on Language Testing*. Edited by Leslie Palmer and Bernard Spolsky. Washington D. C. : Teachers of English to Speakers of Other Languages, 1975.
- Oller, John W., Jr. ed. *Issues in Language Testing Research*. Massachusetts : Newbury House Publishers, Inc., 1983.
- Oller, John W., Jr. ed. *Language Tests at School*. London : Longman Group Ltd., 1979.
- Oller, John W. and Conrad, Christine A. "The Cloze Technique and ESL Proficiency," *Language Learning*, 21 (December, 1971) 183-195.
- Oller, John W. Jr. ; Irvine, Patricia ; and Atai, Parvin. "Cloze, Dictation, and the Test of English as a Foreign Language." *Language Learning*, 24 (1974) 245-252.
- Rye, James. *Cloze Procedure and the Teaching of Reading*. London and Exeter (N H) : Heinemann Educational Books Ltd., 1982.
- Stubbs, Joseph Barton and Tucker, G. Richard. "The Cloze Test as a Measure of English Proficiency ." *Modern Language Journal*, 58 (September-October, 1974) 239-241.
- Taylor, Wilson C. "Cloze Procedure : A New Tool for Measuring Readability." *Journalism Quarterly*, 30 (Fall, 1953) 415-433.
- Taylor, Wilson C. "Recent Developments in the Use of Cloze Procedure," *Journalism Quarterly*, 33 (Winter, 1956) 42-48.
- Upshur, J. A. "Objective Evaluation of Oral Proficiency in the ESOL Classroom." *TESOL Quarterly*, 5 (March, 1971) 47-51.