

英語能力の測定技法としてのクローズ法 に関する実験的研究

北條 礼子

I クローズ法研究の概観

1. クローズ法

クローズ法 (cloze procedure) は第2外国語としての英語の熟達度測定法として妥当性・信頼性・実用性の高いテストとして近年注目を浴びている。クローズ法とは被験者の能力レベルに合わせて適当と思われる英文を選び出し、規則的に n 番目 (通常 5～10語ごと) の単語を削除し、被験者にその空白部分を補充させる形式のテストである。Oller (1971) はクローズ・テストの長さに関して 5～10語ごとの単語削除法による約 50 項目 (250～500語) で十分に被験者の英語熟達度を測定できると述べている。

テスト作成法には、 n 番目にあたる語を組織的に除く方法以外に機能語 (function words) のみを削除する方法や意味語 (lexical words) のみを削除する方法も考案されている。しかしこのような単語の削除法の試みも、いきすぎてしまえばテスト作成者の主観が入りこむ可能性がある。組織的に n 番目の単語を削除する方法は、客観性が高く、作成も容易であり、実験的にも他の作成法によるクローズ・テストと高い相関を示し、優れた作成法であることが報告されている。(Irvine et al, 1974 ; Stubbs, 1974 ; 伊藤, 1978)。

クローズ・テストの採点法に関しても数種類の方法が工夫されている。まずテスト作成の際に英文から削除された語と完全に一致する解答のみを正解とする “Exact Word Method” がある。次に文脈上意味が通じれば

その語も正答とする Acceptable Word Method がある。これはたとえば …He ran away… という問題があり、下線部の正答 ran のところに went という解答があったときに、仮に went がこの文の前後関係から意味の通じる語であるなら正答として認めるといふ採点法である。また、Darnell(1968)は情報理論の entropy の考え方をとり入れた Clozentropy Procedure を採点法として検討したが、彼自身も採点の手続きが煩雑であり、教育の現場や研究への応用が非常に困難であると述べている。ここでとりあげた3種類の採点法のうち Exact Word Method がその簡便さから最も多く用いられているが、Irvine (1974), Stubbs (1974) の研究によればこの採点法と Acceptable Word Method 間には非常に高い相関が認められている。

クローズ・テストに関する研究の多くがその有効性の根拠として優れたテストの特性である信頼性・妥当性・実用性が高いことをあげている。ことに実用性についてはテスト作成が容易であり、テスト作成に経験の浅い者にも容易に作成できること、実施時間が短かくて済むこと、採点が容易であることなどの利点が注目される。

2. クローズ法に関する研究の概略

クローズ法は、リーダビリティ測定のための新手段として登場したが、徐々にリーダビリティに関連の深い分野ばかりでなく他の分野にも応用され研究が行なわれている。

現在のクローズ法の応用分野を Boyce (1974) の分類をもとに8つの分野にわけ、集めた論文をもとに研究の傾向について述べる。

(1) 読解力を測る手段として。

この分野での研究はかなり多い。クローズ・テストと既存の読解力テストを比較検討し、読解力測定テストとしてのクローズ・テストの適性を調べたものに B. J. Salup (1974) などがあるが研究の約半数がクローズ・テストが読解力テストとして有効であったと報告している。

その他に読解力としてのクローズ・テストの信頼性、妥当性を検討した研究がいくつもあるが、結果はクローズ・テストを支持するもの、否定するものと大きく別れている。

(2) リーダビリティを測る手段として

この分野では、比較的多くの研究がなされ、子供の創作した物語、教科書をはじめタイ語などにも応用され、リーダビリティを予測する尺度としてのクローズ・テストの対象はバラエティに富んでいる。また他のリーダビリティ測定法とクローズ・テストを比較している研究もみられる。

(3) 文脈を理解、あるいは文章中のことばの統語位置を理解する能力の測定手段として

この分野での研究数は少なく、この手段としてのクローズ・テストと他の要因（性別、年齢、IQなど）との関連を調べた研究が注目をひくくらいである。

(4) ある肉体的・性格的障害を有する患者の用いる言語特徴を明らかにし、理解・使用できるようになるための調査手段として。

この分野では知能未発達児や精神分裂症患者などが研究対象になっているが、クローズ・テストは正常者との差を明らかに区別しているものが多く、この分野での研究の発展に多くの示唆を与えている。

(5) 読解力教育のための手段として。

クローズ法による練習問題や教材を用いた授業と他の方法による授業を比較検討している研究が多いが、結果はクローズ法が読解力教育手段として特に優れているとはいえないというものがほとんどである。

(6) 第2外国語としての英語教育の手段として。

まず、クローズ・テストが英語を第2外国語とする学習者のどの技能を測っているのかを検討したものに、J. W. Oller, Jr. & C. A. Conrad, D. C. Johnson (1974), 大友 (1978) があるが、どの研究もクローズ・テストが総合能力を測っているであろうと結論づけている。また

これに関連して、クローズ・テストと聴解力テスト間に高い相関があることも報告されている。(Oller & Conrad, 1974; P. Irvine et al, 1974; J. B. Stubbs, 1974; 伊藤, 1978)。その他にも採点法の比較も行なわれ (J. H. Haskell, 1973; Irvine, 1974; J. B. Stubbs, 1974; 伊藤, 1978), さらにテストとしての信頼性・妥当性が調べられ非常に良い結果が報告されている (K. K. Hisama, 1977)。

(7) 口頭理解力を測る手段として。

研究の数は非常に少なく、これからの研究が期待される。

(8) その他

以上の分野以外にも、異文化間コミュニケーションの正確さを測る手段として用いられたり、以上の7つの分類のいくつかを組み合わせたものなど広い分野での応用が試みられている。

II 研究の目的

1. 目的

本研究の第1の目的は英語を母国語としない人々の英語熟達度測定法としてのクローズ・テストが日本人を対象にした場合に、テストとして適切であるかどうかを確かめ、その開発への第一歩を試みることである。

クローズ・テストが英語を母国語としない人々の英語の熟達度を測定するテスト技法として優れたものであること、またその信頼性・妥当性が高いことは先行研究の報告しているところであるが、クローズ・テストの日本人への適用の例は少なく、あらためてクローズ・テストの信頼性・妥当性を検証することにした。

クローズ・テストの信頼性・妥当性を調べるために、まず項目分析を行ない、項目困難度、弁別力指数を求め、さらに弁別力を検討するため χ^2 検定を行なう。第2に Kuder-Richardson の第20公式により信頼度係数を求め内部的統一性 (internal consistency) を調べる。第3に規準関連妥当性 (criterion-related validity) をみるため外部基準として教師によ

る評価、実力考査、学力検査などの結果を選び、クローズ・テストの相関を求める。第4に、概念的妥当性 (construct validity) を得るため、実験 I で知能偏差値とクローズ・テストとの関連をみる。さらに、いわゆる外国語としての英語熟達度を規定している要因相互間の関係を明らかにするために、得られた測度の関連に対して因子分析をおこなう。このことは、テストの不連続点アプローチと総合的アプローチの問題に関して示唆するものがあると思われる。

これまでの研究をみると、クローズ・テストは聴解力テストと高い相関を示している。Darnell (1968) はクローズ・テストと TOEFL (Test of English as a Foreign Language) の下位テストの中で聴解力と最も高い相関があった ($r = .83$) と報告している。Oller (1974) も両者について相関を求めたが Darnell と同じくやはりクローズ・テストは TOEFL の聴解力と最も高い相関を示した。 ($r = .79$)

現在、TOEFL をはじめ英語熟達度測定テストは大別して聴解力と筆記試験から構成されている。

クローズ・テストが英語の熟達度を測定する技法として優れていることは先行研究によって報告されているが、いくつかの形式が考案されているうちで筆記試験の形式のクローズ・テストが聴解力テストと高い相関を示すことも報告されている。

以上のことから、仮に、英語熟達度測定法としての筆記試験形式のクローズ・テストが聴解力をも同時に測定することが可能であるなら、テストの作成が容易であり実施時間も少ないなどの実用性の高いクローズ・テストの利用が推奨されるであろう。

これまで日本でおこなわれた研究報告をみると、クローズ・テスト、聴解力テスト間には $r = .4$ から $r = .5$ の相関があるということであった。(伊藤, 1978 ; 大友, 1978) しかしこれらの研究は大学生を対象にし、被験者の数も 40~50名という研究もあるので、本研究では中学生・高校生を対象に選び、クローズ・テストと聴解力テストの関係を調べることを第2

の目的としている。

2. 仮説

仮説1：クローズ・テストは日本人の外国語としての英語の熟達度を測定する適切なテストである。

仮説2：クローズ・テストは日本人の外国語としての英語の聴解力テストとして適切なテストである。

上記の仮説の場合、「適切なテスト」とは、妥当性・信頼性・実用性の十分に高いものを意味している。しかし、クローズ・テストの実用性は既述であり、ここでは特に妥当性・信頼性をとりあげる。

3. 予備実験

本研究ではクローズ・テストと聴解力テストについて1979年3月に宮城県内の中学校と、千葉県内の高等学校で予備実験を行なった。

(1) 被験者：

実験Ⅰ 中学2年生 116名（女子57名 男子59名）

実験Ⅱ 高校1年生 87名（女子26名 男子61名）

(2) テスト： 予備実験では次のようにテストを実施した。

実験Ⅰ：a. クローズ・テスト 30項目

（7番目ごとの語を削除した形式）

b. 聴解力テスト 13項目（多肢選択法）

クローズ・テストの題材としては、Intensive Course in English : Elementary Part 2 (English Language Services, INC., 1963)と FRIES American English Series ; Book One (D. C. Heath and Company) を用い聴

解力テストは市販の英語検定4級程度の問題が用いられた。

- 実験Ⅱ：a. クローズ・テスト 50項目
 (7番目ごとの語を削除した形式)
- b. 聴解力テスト 15項目 (多肢選択法)
 クローズテストの題材としては、ここ数年間に日本全国の高等学校入学試験に使用されたものの中からいくつかを選び、聴解力テストについては、クローズ・テスト同様、日本全国の高等学校入学試験に用いられたもの、または英語検定3級程度の問題を用いた。

(3) 実施の方法

実験Ⅰについては授業担当者と筆者がテストを実施し、実施時間はクローズ・テスト30分、聴解力テスト15分であった。

実験Ⅱについては学級担任者と筆者がテストを実施し、実施時間はクローズ・テスト20分、聴解力テスト13分であった。

クローズ・テスト実施の際「空白部分に入る最も適当だと思ふ語を1語だけ記入してください。」という指示が与えられた。

(4) 結果の整理の方法と結果

クローズ・テストの採点法に関しては、Irvine, Atai & Oller (1974)の研究によるとイグザクト・ワード法とアクセプタブル・ワード法間に $r = .94$ の相関があり、また Stubbs (1974) もこれら2つの採点法は1%レベルで高い相関があったと報告している。以上のことから、この研究ではテストの結果は原文の語(つまり、削除された語)と完全に一致した解答のみを正解とするイグザクト・ワード法によって、実験Ⅰ、Ⅱとも採点がおこなわれた。次にこうして得られたクローズ・テストの得点と聴解力テストの得点について、妥当性と信頼性を検証するために項目分析を行ない、項目別の困難度をまず求めた。その結果、実験Ⅰにおいては困難度90%以

上, 10%以下の項目がクローズ・テストに8項目, 聴解力テストに2項目, 実験Ⅱにおいてはクローズ・テストに4項目, 聴解力テストに8項目含まれていた。次に項目ごとの弁別力をみるために実験Ⅰ, Ⅱとも上, 下それぞれ25%の上位群, 下位群を選出し χ^2 検定をおこなった。聴解力テストに関しては, 結果が正規分布を示さなかったので, 中央値で上位群, 下位群に分け χ^2 検定をおこなった。その結果, 実験Ⅰにおいてはクローズ・テスト: 22項目, 聴解力テスト: 13項目に, 実験Ⅱにおいてはクローズ・テスト: 32項目, 聴解力テスト: 10項目が有意であった。

また Kuder-Richardson の第20公式による信頼度係数を求めたが実験Ⅰではクローズ・テスト: $r = .811$; 聴解力テスト: $r = .658$, 実験Ⅱではクローズ・テスト: $r = .784$; 聴解力テスト: $r = .574$ であった。

予備実験の結果から, 困難度が90%以上, または10%以下のものを除外, または修正し本実験に備えた。

Ⅲ 実 験

1. 方 法

(1) 被験者:

実験Ⅰ 宮城県内の中学3年生133名 (女子65名 男子68名)

実験Ⅱ 千葉県内の高校2年生 86名 (男子86名)

全日制の普通高校である。

(2) テスト: 本研究では以下のようにテストを実施した。

実験Ⅰ a. クローズ・テスト: 予備実験の結果, 削除された8項目を除いた22項目

(下位変数 1)

b. 聴解力テスト: 予備実験の結果, 削除し, 調整を加えた13項目 (多肢選択法)

(下位変数 2)

- 実験Ⅱ a. クローズ・テスト：予備実験の結果，削除された
4項目を除いた46項目
- b. 聴解力テスト：予備実験の結果除外し，調整を加
えた15項目（多肢選択法）
（下位変数 2）
- c. 3部からなる市販の学力検査
1. 聞くこと・話すこと 20項目（多肢選択法）
（下位変数 5）
 2. 読むこと 22項目（多肢選択法）
（下位変数 6）
 3. 書くこと 20項目（多肢選択法）
（下位変数 7）
- この学力検査に関しては総点（下位変数 8）も求
める

(3) 実施の方法

実験Ⅰについては，1979年4月に筆者がクローズ・テストと聴解力テストを実施した。実施時間はクローズ・テスト30分，聴解力テスト15分であった。知能偏差値（教研式）（下位変数 3）は1977年6月に，教研式標準英語学力偏差値（下位変数 4）は1978年6月にそれぞれおこなわれた検査の結果であるが，学籍簿の記録から得られた。学年末評定（下位変数 5）は被験者の第2学年次の第3学期末の教師による英語の評定であり，実力考査（下位変数 6）は1979年4月に施行されたものである。クローズ・テスト実施の際，「空所に最も適切と思うことばを1語だけ考えて書いて下さい。」という指示を与えた。

実験Ⅱについては，クローズ・テスト，聴解力テスト，標準学力検査とも筆者が，授業担当者等の協力を得て1979年4月に実施した。テストの実施時間はクローズ・テスト35分，聴解力テスト15分，標準学力検査50分で

あった。リーダー（下位変数 3）、英文法（下位変数 4）は被験者の第 2 学年次末の教師による評定であり、学籍簿から得られた。クローズ・テスト実施の際には、実験 I と同様な指示が与えられた。

(4) 結果の整理の方法

1. 実施したすべてのテストの平均点，標準偏差を求める。なお，クローズ・テストの採点はイグザクト・ワード法でおこなう。
2. 実施したすべてのテストに関して項目分析をおこなう。テストの結果得られた得点から上位，下位，それぞれ 25% ずつの上位群，下位群を抽出し，項目ごとに，困難度，弁別力指数を求め，さらに χ^2 検定をおこない，それぞれのテストの妥当性を検討する。
3. 実施したすべてのテストに関して Kuder-Richardson の第 20 公式による信頼度係数を求める。その際標準誤差も算出する。
4. 実験で得られた得点に基づいて，実験 I では 6 つの下位変数，実験 II では 8 つの下位変数について Pearson プロダクト・モーメント相関係数を求める。
5. 実験 I では 6 つの下位変数，実験 II では 8 つの下位変数から因子分析により共通因子を求める。

IV 実験の結果

1. 平均点，標準偏差

実験 I (N=133) ではクローズ・テストと聴解力テストを，実験 II (N=86) ではクローズ・テスト，聴解力テストと学力検査を実施した。その結果得られた得点から，それぞれのテストについて平均点と標準偏差を求めたのが表 1 である。

表1 実施したテストの満点, 平均点, 標準偏差

	テスト	満点	平均点	標準偏差
実験Ⅰ	クローズ・テスト	22	11.92	5.13
	聴解力テスト	13	4.96	2.60
実験Ⅱ	クローズ・テスト	46	17.40	9.57
	聴解力テスト	15	7.88	2.92
	聞くこと (第1部)	20	10.59	3.72
	話すこと (第2部)	22	9.44	3.79
	書くこと (第3部)	20	7.71	3.70
	総点	62	27.65	9.17

2. 項目分析

クローズ・テストと聴解力テストの項目分析をおこなって, 次のような結果を得た。

a. 実験Ⅰ

クローズ・テストの22項目のうち困難度が90%以上, 10%以下のものはなかったが, 弁別力指数 (Index of Discrimination) が, I.D. = .19 以下のものが2項目 (項目5と18) あった。 χ^2 検定の結果, 0.1%レベルで有意な項目が19項目, 1%レベルで有意な項目が1項目であった。

聴解力テストの13項目をみると, 困難度が90%以上, 10%以下のものはなく, 弁別力指数 I.D. = .19 以下のものが2項目 (項目9と11) あった。しかし χ^2 検定の結果, 13項目全部が有意であり10項目が0.1%レベルで, 1項目が1%レベルで, 2項目が5%レベルで有意であった。

b. 実験Ⅱ

クローズ・テストの46項目のうち, 困難度が10%以下のものが1項目 (項目20), 弁別力指数が, .19以下のものが5項目あった。(項目16, 20, 32, 36, 41) χ^2 検定の結果0.1%レベルで25項目, 1%レベルで8項目, 5%レベルで5項目, 全体で38項目が有意であった。

聴解力テストは15項目であったが、困難度が90%以上の項目が1項目(項目6)あり、弁別力指数がI.D. = .19以下のものはなかった。 χ^2 検定の結果、0.1%レベルで有意な項目が7、1%レベルで有意な項目が3、5%レベルで有意な項目が3あり、全体で13項目が有意であった。

3. 信頼度係数

実施したすべてのテストについて Kuder-Richardson の第20公式による信頼度係数、及び標準誤差を求めた。その結果を表2に示す。

表2 KR20による信頼度係数

	テ ス ト	KR20 : r	標 準 誤 差	
実験 I	クローズ・テスト	0.868	1.864	
	聴解力テスト	0.659	1.518	
実験 II	クローズ・テスト	0.916	2.774	
	聴解力テスト	0.684	1.641	
	学力検査 {	聞くこと	0.735	1.915
		話すこと	0.657	2.220
		読むこと	0.746	1.890
		書くこと	0.854	3.504
総 点				

以上の信頼度係数はそれぞれに高い信頼度であるといえることができる。

4. Pearson プロダクト・モーメント相関係数

a. 実験 I

被験者133名は全変数について全得点が得られ欠測値はなかった。この133名について6下位変数間の Pearson プロダクト・モーメント相関係数を算出した結果を表3に示す。結果は、全変数間の全相関係数が1%レベルで有意であった。

b. 実験 II

被験者86名は全変数について欠測値がなく、被験者全員について8下位

変数間の Pearson プロダクト・モーメント相関係数を求めた。その結果が表 4 であるが、全変数間の全相関係数が 1% レベルで有意であった。

表 3 6 下位変教の相関行列

1	1.00000					
2	0.53393	1.00000				
3	0.62880	0.49959	1.00000			
4	0.82200	0.62050	0.66161	1.00000		
5	0.71023	0.65071	0.63488	0.79153	1.00000	
6	0.70596	0.66791	0.59051	0.75305	0.85999	1.00000
	1	2	3	4	5	6
	クローズ・テスト	聴解力テスト	知能偏差値	英語学力偏差値	学年末評定	実力考査

この表の相関係数はすべて $p < .01$

表 4 8 下位変数の相関行列

1	1.00000							
2	0.58839	1.00000						
3	0.75586	0.60464	1.00000					
4	0.74421	0.55464	0.88685	1.00000				
5	0.67190	0.41281	0.66028	0.59120	1.00000			
6	0.60436	0.41675	0.47065	0.45171	0.55514	1.00000		
7	0.62078	0.54716	0.53371	0.53095	0.54255	0.51259	1.00000	
8	0.75680	0.54925	0.66743	0.62725	0.83312	0.82202	0.82858	1.00000
	1	2	3	4	5	6	7	8
	クローズ・テスト	聴解力テスト	READER 成績	GRAMMAR 成績	聞くこと話すこと (第1部)	読むこと (第2部)	書くこと (第3部)	総点

この表の相関係数はすべて $p < .01$

5. 因子分析

a. 実験 I

VARIABLES..	LABELS..	FACTOR 1
V 1	クロウズ テスト	0.82905
V 2	チヨウカイ リヨク テスト	0.70304
V 3	チノウヘンサチ	0.71370
V 4	エイゴ ガクリヨクヘンサチ	0.90307
V 5	ガクネンマツ ヒヨウテイ	0.90484
V 6	ジツリヨク コウサ	0.88256

表5 実験 I のバリマックス法の結果

クロウズ・テストの得点，聴解力テストの得点知能偏差値，英語学力偏差値，学年末評定，実力考査の得点に基づいてバリマックス法による因子分析をおこなったが，その結果が表5である。結果をみると，第1因子が抽出されたのみであった。

b. 実験 II

VARIABLES..	LABELS..	FACTOR 1
V 1	クロウズ テスト	0.87046
V 2	チヨウカイ リヨク テスト	0.64895
V 3	READER セイセキ	0.83569
V 4	GRAMMAR セイセキ	0.79757
V 5	キクコト ハナスコト	0.77880
V 6	ヨムコト	0.68843
V 7	カクコト	0.74110
V 8	ソウテン	0.94185

表6 実験 II のバリマックス法の結果

実験 I と同様に，クロウズ・テストの得点，聴解力テストの得点，学年末のリーダーの成績，英文法の成績，学力検査の第1部（聞くこと・話すこと），第2部（読むこと），第3部（書くこと），総点に基づいてバリマ

ックス法による因子分析をおこなった。その結果が表6であるが、因子は第1因子のみが抽出された。

6. 仮説の検証

仮説1の検証のため、実験I、IIにおいてクローズ・テストの妥当性、信頼性を検討した。クローズ・テストの結果について項目分析をおこない、 χ^2 検定により弁別力を検定した。またKuder-Richardsonの第20公式による信頼度係数を求め内部的統一性を調べた。 χ^2 検定の結果、実験Iでは22項目中20項目が有意であり、実験IIでは46項目中38項目が有意であった。KR20による信頼度係数は実験Iのクローズ・テストでは $r = .868$ であり実験IIのクローズ・テストでは $r = .916$ であった。以上の結果からクローズ・テストが弁別力に優れ、内部的統一性の高いテストであることが検証された。

さらに、クローズ・テストの妥当性を調べるために、実験Iでは6つの下位変数、実験IIでは8つの下位変数についてPearsonプロダクト・モーメント相関係数を求めた。その結果、クローズ・テストは他の下位変数と $r = .6$ から $r = .8$ の1%レベルで有意な高い相関があり規準関連妥当性、概念的妥当性の高いテストであることが検証された。

以上から「クローズ・テストは日本人の外国語としての英語の熟達度を測定する適切なテストである」という仮説1は支持されたといつてよい。

仮説2の検証のため、実験I、IIにおいてクローズ・テストと聴解力テストの相関係数を求めた。その結果、実験Iでは、 $r = .534$ ($p < .01$)、実験IIでは $r = .588$ ($p < .01$)が求められた。この結果は日本における他の研究結果にほぼ近いものであった。

これら2つの相関係数はOller等の研究に比較して、彼らの実験結果($r = .7$ 以上)ほどには高い相関は得られなかったが、クローズ・テストが聴解力を、かなりの程度予測できるといえる。本研究に関する限り、仮

説2「クローズ・テストは日本人の外国語としての英語の聴解力テストとして適切なテストである」は支持されたと考えられる。

VI 結果の考察

1. 本研究で実施されたクローズ・テストの結果について項目ごとに χ^2 検定をおこなったところ、実験Ⅰでは22項目中20項目が有意であった。(19項目が $p < .001$, 1項目が $p < .01$) 実験Ⅱにおいても46項目中38項目が有意であった。(25項目が $p < .001$, 8項目が $p < .01$, 5項目が $p < .05$)

また Kuder-Richardson の第20公式によるクローズ・テストの信頼度係数も実験Ⅰでは $r = .868$, 実験Ⅱでは $r = .916$ が得られた。実験Ⅰ, Ⅱで用いられたクローズ・テストの項目がそれぞれ22項目, 46項目であったことを考えると、クローズ・テストは極めて高い内部的統一性を有することが検証されたといえよう。

2. 次にクローズ・テストの規準関連妥当性をみるために、実験Ⅰではクローズ・テストと外部基準として、教師による評価、実力考査の結果、学力偏差との相関を求めたが、その結果はそれぞれ $r = .710$ ($p < .01$), $r = .706$ ($p < .01$), $r = .822$ ($p < .01$) と高い相関が得られた。同様に実験Ⅱではクローズ・テストと外部基準として教師による評価(①リーダの成績, ②英文法の成績), 学力検査の結果(③第1部: 聞くこと, 話すこと, ④第2部: 読むこと, ⑤第3部: 書くこと, ⑥総点)との相関を求めたところ ① $r = .756$ ($p < .01$), ② $r = .744$ ($p < .01$), ③ $r = .672$ ($p < .01$), ④ $r = .604$ ($p < .01$), ⑤ $r = .621$ ($p < .01$), ⑥ $r = .757$ ($p < .01$) が得られた。クローズ・テストはこのように外部基準と高い相関を示したが、特に教師による評価と学力検査の総点間に高い相関があった。以上からクローズ・テストは規準関連妥当性の高いことが検証されたといえよう。

3. 実験 I において、クローズ・テストの概念的妥当性を調べるために、知能偏差値（教研式）とクローズ・テストとの相関を求めたが、その結果は $r = .629$ ($p < .01$) というかなり高い相関が得られた。

4. 以上の結果から、クローズ・テストは妥当性信頼性の高いテストであることが検証されたが、この結果は多くの先行研究の報告する結果に一致する。

5. 本研究の第 2 の目的は、クローズ・テストと聴解力テストの関係をみることであったが、この両テスト間の相関を求めたところ実験 I では、 $r = .534$ ($p < .01$)、実験 II では $r = .588$ ($p < .01$) と比較的高い相関が得られた。この結果は $r = .7$ 以上の高い相関を報告した、被験者が日本人ではない先行研究の結果には一致しなかった。しかし、日本人を対象とした 2, 3 の研究結果とほぼ近い値であった。

このように、日本人を対象とした場合、それ程高い相関が得られないのは、聴解力テストが、クローズ・テストと比較して妥当性・信頼性が高くなかったことに関連があろう。実験 I, II では KR 20: $r = .659$, $r = .684$ であり、外部基準との相関も 1% レベルで有意であったものの、他に比べて低い傾向がみられた。仮に、聴解力テストを改良し、より妥当性・信頼性の高いテストを作成できれば、クローズ・テスト、聴解力テスト間の相関もより高いものが得られる可能性があるであろう。

6. 本研究で得られた測度の関連を一括してみるために、実験 I では 6 つの測度、実験 II では 8 つの測度に基づいて、バリマックス法による因子分析をおこなった。その結果、中学校、高等学校ともに抽出された因子はひとつだけであった。

このことから、少なくとも本研究の被験者に限っては、クローズ・テストと聴解力テストが共通のものを測定していることがいえよう。

読む力・書く力を重視する筆記試験の形式のクローズ・テストと聞く力

を重視する聴解力テストに共通するものが何かについて述べると、言語の総合能力と考えることもできるが、Oller (1973) は、期待文法 (the grammar of expectancy) ということばで説明している。この期待文法とは、言語学的順序と言語学的要素以外の文脈とを関係づけることを可能にする文法組織のことである。この文法力が備っている聞き手、読み手は話し手、書き手が次に何を話しあるいは何を書くかを予測し期待しながら自分なりの修正を加えて考えることができるのである。

しかし、本研究の因子分析からは Oller のいう共通因子と同種のものが得られたとは思えないが、本研究の結果からはいわゆる「英語熟達度」という共通因子があることは確かなようである。ただし、因子分析にかけた測度の数が少なかったことに原因があるかもしれないのでさらに様々なテストの測度をふやして因子分析をおこなう必要がある。

7. この研究の結果から、クローズ・テストは英語の熟達度測定法として有効であることがわかったが、今後は次の事項を研究課題として研究をすすめていく必要があると思われる。

- (1) 聴解力テストを改良し、信頼性・妥当性の高いものにしていく。
- (2) 本研究で用いられなかったクローズ・テストの他の採点法であるアクセプタブル・ワード法とイグザクト・ワード法間の相関をみる。また聴解力テストをはじめ、外部基準となる測度との相関を求める。

〔参 考 文 献〕

- Charles, E. "An Investigation of the Use of Cloze Tests to Compare Gain Scores of Students in Science Who Have Used Individualized Science Materials and Those Who Have Used Traditional Science Textbook Materials." *DAI*, 31 (10), 1971 (Apr.), 5026-A.
- Cox, J. A. K. "A Comparison of Two Instructional Methods Utilizing the Cloze Procedure and a More Traditional Method for Improving Reading Comprehension and Vocabulary in Context in a Disadvantaged Fourth-Grade Elementary School Sample." *DAI*, 35 (10), 1975, 6569A.
- Crawford, A. N. "The Cloze Procedure as a Measure of the Reading Comprehension of Elementary Level Mexican-American and Anglo-American Children." *DAI*, 31 (7), 1971 (Jan.), 3162-A.
- Culhane, J. W. "The Use of an Intrative Research Process to Study the Adaptation of Cloze for Improving the Reading Comprehension of Expository Materials." *DAI*, 34 (3), 1973, 997-8A.
- Darnell, D. K. "The Development of an English Language Proficiency Test of Foreign Students, Using a Clozentropy Procedure." Boulder Colorado, 1968. (Final Report to U. S. Department of Health, Education and Welfare.)
- Echols, S. L. "An Investigation of Fourth-, Fifth-, and Sixth-Grade Children's Comprehension of Selected Syntactic Structures Based on Pupil Responses to Systematically Deleted Cloze Tests." *DAI*, 34 (8), 1974, 4982A.
- Edwards, R. D. "The Cloze Procedure as a Measure of the Reading Comprehension of Poetry." *DAI*, 34 (8), 1974, 4982A.
- Fleming, J. B. "Analysis of the Readability of Fifth Grade Social Studies Textbook Using the Cloze Procedure." *DAI*, 36 (1), 1975, 104A.
- Geyer, J. R. & A. R. Carey "Predicting and Improving Comprehensibility of Social Studies Materials: The Role of Cloze Procedure and

- Readability Adjustment." *Reading World*, 12 (2), 1972 (Dec.), 85-93.
- Haskell, J. F. "Defying the Cloze Testing and Scoring Procedures for Use with ESL Students." *DAI*, 34 (6), 1973 (Dec.), 3206-7A.
- Henry, P. E. "The Effect of Interest on Reading Comprehension as Measured by Cloze and Multiple Choice Tests." *DAI*, 30 (9), 1970 (Mar.), 3857-8A.
- Hisama, K. K. "Design and Empirical Validation of the Cloze Procedure for Measuring Language Proficiency of Non-Native Speakers." *DAI*, 37 (9-A), 1977 (Mar.), 5766.
- "A New Direction in Measuring Proficiency in English as a Second Language." 1977 (Apr.) (Paper presented at the Annual Meeting of the American Educational Research Association.) ELEC : ED 150 198.
- Irvine, P., P. Atai & J. W. Oller "Cloze, Dictation, and the Test of English as a Foreign Language." *Language Learning*, 24 (2), 1974 (Dec.), 1974.
- 伊藤弘子 "クローズ・テスト" 『英語教育』大修館, 1978 (5~7月)
- Johnson, D. C. "The Evaluation of Comprehension When English Is a Second Language by Use of the Cloze Procedure." *DAI*, 35 (8-A), 1975 (Feb.), 5121.
- O'Brien, D. "An Investigation of Relationships Existing Between Cloze Form Measures of Reading Behaviors Determined by the Reading Inventory Technique and Standardized Tests." *DAI*, 35 (1), 1974, 324-5.
- Ohtomo, Kenji "Problems and Analysis : Testing Overall English Language Proficiency of Japanese Students." *The Teaching of English in Japan*. Ikuo Koike et al. (Ed.) Tokyo : Eicho-sha Publishing Co., Ltd. 1978, 463-77.
- Oller, J. W. & C. A. Conrad "The Cloze Technique and ESL Proficiency." *Language Learning*, 21 (2), 1971 (Dec.). 183-95.

- Oller, J. W., D. Bowen, T. T. Dien & V. W. Mason "Cloze Tests in English, Thai, and Vietnamese : Native and Non-Native Performance." *Language Learning*, 22 (1), 1972, 1-15.
- Oller, J. W. "Cloze Tests of Second Language Proficiency and What They Measure." *Language Learning*, 23 (1), 1973, 105-18.
- Ramanauskas, S. "Contextual Constraints beyond a Sentence on Cloze Responses of Mentally Retarded Children." *American Journal of Mental Deficiency*. 77 (3), 1972 (Nov.), 338-45.
- Rufener, J. B. "Use of the Cloze Procedure with Thai School Children : an Exploratory Study of Readability and Individual Difference in Reading." *DAI*, 33 (6), 1972 (Dec.), 2774A.
- Rynders, P. "Use of the Cloze Procedure to Develop Comprehension Skill in the Intermediate Grades." *DAI*, 32 (10), 1972 (Apr.), 5676-7A.
- Salup, B. J. "An Investigation of the Cloze Technique for Measuring the Reading Comprehension of College Freshman." *DAI*, 36 (6-A), 1975 (Dec.), 3525-6.
- Sikes, H. C. "A Comparative of Cloze Procedure Scores on Original and Published Materials." *DAI*, 32 (5), 1979 (Nov.), 2320-A.
- Silverman, C. "Psycholinguistics of Schizophrenic Language." *Psychological Medicine*, 2 (3), 1972 (Aug.). 254-59.
- Stubbs, J. B. & G. R. Tucker "The Cloze Test as a Measure of English Proficiency." *Modern Language Journal*, 58 (5-6), 1974 (Sep.-Oct.), 239-41.
- Thompson, K. G. "The Concurrent Validity of a Non-Reinforced of Readability of Selected Fourth Grade Students." *DAI*, 34 (10-A), 1974 (Apr.), 6372-73.

An Experimental Study of the Cloze Procedure as a Testing Device of English

Reiko HOJO

Background :

In Japan little research has been done in the field of foreign language proficiency testing, especially, concerning English proficiency testing.

The cloze procedure or the cloze test has been said to be appropriate as an English proficiency test.

The cloze procedure is a generic term given to a testing technique which deletes the words of English passages according to some mechanical rules (usually every n-th). The subjects are required to fill in blanks.

Since the cloze procedure was first discovered by W. Taylor in 1953, it has been applied in various areas, for example, as a testing device in Teaching English as a Second Language.

Many experiments performed on the cloze showed signs of its validity and reliability. Especially, the cloze test was said to have a high practicality, such as the ease of construction, scoring and administration.

Although it seems necessary to prove the validity and the reliability of the test before it is widely used, experiments of this type have not been performed in Japan.

Since both the cloze test and the listening comprehension test probably measure the internalized grammar of expectancy in English of the subjects who are non-native speakers of English, the correlation between them was found to be high. ($r = .7$) Unfortunately, in Japan, the subjects of the experiments are limited to 40—50 college students. However, in order to reinforce the validity of the experiments, the subjects must be more inclusive in number as well as in grade level. (i. e. the junior high and high school)

Purpose :

The primary purpose of this study was to evaluate the cloze test in terms of three criteria, that is, validity, reliability and practicality.

The secondary purpose of this study was to investigate the relationship between the cloze test and the listening comprehension test in case the subjects were Japanese.

When the correlation between them was significantly high, the cloze test was able to measure the listening comprehension ability of the subjects.

Method :

As pre-tests cloze tests and listening comprehension tests were administered in March 1979. The subjects were 116 junior high-school students and 87 high school student. The Exact-Word scoring method was used in both cases. The item analysis was conducted and based on these results, test formats were improved for the actual experiment in April 1979.

The subjects were 133 junior high school students in Experiment

I and 86 high school students in Experiment II.

Both in Experiment I and II, cloze tests and listening comprehension tests were administered. In Experiment I, teacher's grades, IQ score, the score of the standard test and the score of the English achievement test were gained as the external criteria. In Experiment II teacher's grades and the scores of the standard test were obtained as the external criteria.

The Exact-Word scoring method was used in both Experiments.

In order to evaluate the validity of the tests, 1) difficulty 2) item discriminating index were computed and 3) χ^2 test was employed on each item. For the purpose of examining the internal consistency of all the tests administered in this study, the reliability coefficients were obtained by the Kuder-Richardson formula 20.

In order to evaluate the criterion-related validity and the construct validity, the Pearson Product-Moment coefficient correlations were computed.

For the purpose of investigating the relationship between the cloze test and the listening comprehension test, the Pearson Product-Moment coefficient correlations between them were also obtained.

The factor analysis was employed to examine the relationships of all the variables in this study.

Results and Discussion :

The results of χ^2 test on the cloze tests showed that 20 out of 22 items were significant in Experiment I, while 38 out of 46 items were significant in Experiment II. The reliability coefficients were

obtained by the Kuder-Richardson formula 20 : $r = .868$ (Experiment I) ; $r = .916$ (Experiment II). These results proved that the cloze had a high discrimination power and internal consistency.

The Pearson Product-Moment coefficient correlations among 6 variables gained in Experiment I were all significant at the 1 % level. ($r = .6 - r = .8$) The Pearson Product-Moment coefficient correlations among 8 variables obtained in Experiment II were also significant at the 1 % level. Therefore, the cloze test proved to have high criterion-validity and construct validity.

Although the coefficient correlations between the cloze test and the comprehension test were significant at the 1 % level, they were not as high as those gained in foreign experiments. However, as far as this study is concerned, the cloze test may be appropriate for measuring the listening comprehension ability of the Japanese students. The internal consistency of the listening comprehension was not as high as that of the cloze test. If the listening comprehension test formats are improved, the higher correlation between the cloze test and the listening comprehension test.

In order to examine the relationships of all the variables in this study, the factor analysis was employed. In both Experiment I and II, only one factor was extracted, which could not be named due to insufficient data, but which showed relevancy to both the listening comprehension test and the cloze test.

Cloze tests attract our attention as excellent in quality with a high degree of reliability, validity, and practicality.

It seems that further research should be done :

(1) To increase both the reliability and validity of the listening comprehension tests.

(2) To see the correlation between the Acceptable-Word scoring method and Exact-Word scoring method, both of which are also classified as scoring methods of cloze tests, and at the same time, to compute the correlation between the two methods and those measures which serve as the external criteria.

(3) To increase the number of measures and investigate what items cloze tests and listening comprehension tests have in common.

(4) To explain more clearly the relationship between cloze tests and speaking ability, and then point out the determinants of English proficiency.