

クローズ法による日本語能力の測定

Measurement of Japanese Language Ability by The Clozer Method

秦 喜美恵

I. 研究の背景

1. 外国語としての日本語熟達度テストの現状

1962年12月に日本語教育学会から創刊された『日本語教育』を中心に、これまでの日本語教育に関する研究を調べた。それらは日本語教育の意義から始まり、いかにして日本語を教えていくのがよいのかといった文法事項の説明方法、漢字の教え方、基本語彙を頻度別に扱ったものなど、実際に日本語を教える場面で役立つものが研究の中心であった。日本語のテストに関する研究は、1964年の4, 5号において興水によってはじめて言及されている。興水はその論文の中で、言語テストを R. Lado(1961)^{49 注1}のいう言語要素(lements: 発音, 文法, 語彙等)と言語技能(skills: 聞くこと, 話すこと, 読むこと, 書くこと)を測定する方法にそって、次の9種類に分けて紹介している⁴⁴。1) 言彙テスト, 2) 文法テスト, 3) 音韻テスト, 4) 文字テスト, 表記テスト, 5) 読解テスト, 6) 作文テスト, 7) 書写テスト, 8) 談話テスト, 9) 聴解テスト, の9種類である。これらのテストのうち1)2)3)4)7)は言語要素のテスト, 5)6)8)9)は言語技能のテストといえよう。しかし興水は日本語をテストするという意味において, 4)文字テスト, 表記テスト, および7)書写テストを付け加えた。つまり, 日本語は英語のようにアルファベットによってすべてが書き表せる言語ではなく, その表記方法にはローマ字, カタカナ,

注1. この番号は, 参考文献を示すものである。

ひらがな、漢字の4種類があるからである。4)文字テスト、表記テストでは、漢字、ひらがななどの読み、書き、送りがなの付け方、句読点の打ち方などをテストしようとしたものである。7)書写テストは、日本語の文字が、きちんと読みやすく書けるかどうかをテストするもので、その場合、書く速さや筆順などがテストされる。また、同号には日本語教育の編集委員会によって「実例からみた日本語の試験問題——出題意図と形式による分類——」⁶⁰の結果も発表されているが、これは実際に使われている日本語の試験をまとめ、分析して各項目が測定しようとした日本語の能力は何であったかを報告したもので、これらのテスト分類および分析はいずれも構造主義言語学の流れをふまえた、不連続点テスト (discrete-points test) の範囲にとどまる。

次に、日本語のテストが問題にされるのは、1977年32号の日本語教育「特集：評価と標準テスト」においてである⁹⁷。そこで報告されている研究のほとんどは、実際の試験を紹介しながらその測定方法を解説したものである。その後、東海大学留学生別科(1978)⁸⁰、国際交流基金(1979)⁴³、国際基督教大学(1980)³⁴(1981)³⁵、日本語教育学会(1981)⁶¹、東京外国語大学外国語学部附属日本語学校(1981)⁹⁶、早稲田大学語学教育研究所(1981)¹⁰⁸、などの日本語教育機関から、日本語のテストに関する調査分析結果が報告されている。これらの多くは測定する要素および技能をあげ、それぞれをどんな試験問題で測定したかを報告したものである。その中で、日本語教育学会(1981)⁶¹と石田等(1980)³⁴(1981)³⁵は、実施されたテストの妥当性と信頼性について統計的処理をおこない検討している。

植松(1977)⁹⁷はテストを実施した後の処理方法として、テストの各項目の難易度、問題文の適否、問題の内容の適否などを見直すことが必要だとし、テストの妥当性、信頼性を統計的に処理することの必要性を述べている。

各種の日本語テストの結果報告を検討してみて、もっとも疑問に思ったことは、そのテストの目的はなんであるかということである。テストの多くは、日本語を要素および技能に分けて測定しているが、それが学習者の持つ日本語の知識をみるためだったり(学力テスト)、カリキュラムに基づいておこな

われたテストであるならば、その到達度をみるためのテストとして（到達度テスト）、学習者の日本語の総体的な能力をみるためであれば、それは熟達度または能力を測定するためのテスト（能力テスト）として、検討されるはずであるが、日本語テストの結果報告の多くでは、テストの目的が明確に述べられていない。

テストの目的を明確にすることは、テスト作成においての第1の条件である。また、「ある特定の研究や実務にたずさわるのに必要な学力があるかどうかを調べるため」「語学の能力別クラスを編成するために学習者の言語能力のレベルを調べる」「各生徒個人の弱点特色を診断するため」であれば、それは熟達度テスト (proficiency test) であるし、「各学習者が設定された教育到達目標にどれだけ到達できたかを測定するため」「教師の側から授業がどのくらい効果的におこなわれているかを評価するため」ならば、それらは到達度テスト (achievement test) であるといえる。

これまでみてきた日本語のテストの目的は、主に「留学生のための統一試験」「日本の大学に入学するための試験」「レベル分けのための試験」「日本語能力検定試験」であった。このような目的を有するテストは一般に熟達度テストの要素が強く含まれているといえる。なぜならば、これらの目的で実施されるテストでは、受験者はそれぞれに異った過去の学習経験を持っており、従って、テストの目的はある限定された期間内の習得度を測定するだけでなく、むしろ各学習者の過去の経験を越えた学習活動の成果の総体的な語学力を測定しようとするものだからである。

さまざまな学習経験を持つ外国人の日本語の学習者に対して、総合的な学力が測定できるような標準化されたテストを開発することが、今後の日本語教育の主要な課題である。

2. 日本語におけるクローズ・テストの作成と問題点

内在している言語能力を引き出すための総合学力テスト法として、P. Holtzman(1967)は、1)クローズ・テスト (cloze test), 2)ノイズテスト (noise

test), 3) クローゼントロピープロセジャ (clozentropy procedure) を高く評価している²⁹。また, P. Meara (1974) は作成がかなり簡単で, 個々の項目でなく総体的な言語能力を測定しようとするテストに, 1) クローズ・テスト, 2) ノイズテストをあげ, すぐれたテストとして評価している⁵⁴。本研究では, 外国語熟達度テストとして妥当性, 信頼性, 実用性を兼ね備えた外国語の総合能力の測定法としてクローズ・テストに注目し, これを日本語の熟達度テストとして用いることができるかどうかを調べてみたいと思う。

日本語の熟達度テストとしてのクローズ・テストの作成はこれが初めての試みである。そこでクローズ・テストを作成するにあたり, いくつかの問題点を検討しておかなければならない。

クローズ・テストの作成方法は基本的には, これまでに標準的となっている作成法にのっとっておこなうが, 英語と日本語とでは, 言語の構造上の違いがあるので, その違いを認識したうえで日本語のクローズ・テストを作成しなければならない。基本的には二つの問題が考えられる。ひとつには, 表記法の問題があげられる。日本語には, ローマ字, カタカナ, ひらがな, 漢字の4種類の表記方法があるが, このうちのどの表記方法でおこなうか, またはいくつかの折衷案を用いるかである。もうひとつの問題は, どの語をもって1語とするかという「クローズ単位 (cloze unit)」の設定方法である。英語では一般に語 (word) は互いに分離しているので, これを単位として扱うのは, その操作上からいって大変簡単である。英語のこのような単語に相当するものとして日本語では「語」があげられるが, それは文中では特に分離されてはいない。

芝祐順 (1957) は「読み易さの測り方—クローズ法の日本語への適用—」の中で, クローズ単位 (cloze unit) について次のような報告をしている⁸⁴。

「字単位 (この場合「字」とは活字単位をいう) および語単位 (この場合「語」とはローマ字の分ち書きの基準によって分離する) 別で, 八つの例文についてその読み易さの検定をおこなったところ, 字単位のクローズ法よりも, 語単位のクローズ法の方が識別力が強いことが分かった。またクローズ

法は日本語にも適用できることが分かった。その際クローズ単位としては語を用いるべきである。」

次に表記の問題であるが、書かれたものを読んでおこなうというテストの性質上、日本語の表記をどのように統一するのがもっとも良いのかということが問題になる。ローマ字書き、カタカナ書き、ひらがな書き、漢字、ひらがな、カタカナ書きなどの表記方法が考えられるが、漢字圏の学生と、非漢字圏の学生との条件を等しくするためには全部カタカナ書きかひらがな書きを用いるのが良いと思われる。

しかし本研究では、被験者が3カ月以上日本語を学習したという経験を考慮して、既習の漢字で理解できると思われる漢字を最低限に使用し、また外来語はカタカナ書きにした。従って、漢字、ひらがな、カタカナ書きを用いることにした。

II. 研究の目的

1. 研究の目的

本研究の第1の目的は、英語の熟達度を測定するテスト技法としてのクローズ・テストが日本語を学習している外国人を対象とした場合に適切なテストであるかどうかを明らかにすることである。

第2の目的は、クローズ・テストにより得られた結果と日本語の四つの技能、1)聴解力、2)書く力、3)話す力、4)読解力との関係を明らかにすることである。特に読解力をみるテストとしてのクローズ・テストが英語の聴解力と高い相関を示すことは、これまでの外国でおこなわれた研究で明らかにされている。

2. 研究の仮説

仮説1：クローズ・テストは外国人の外国語としての日本語熟達度テストとして妥当性、および信頼性の高いテストである。

仮説2-1: クローズ・テストは外国人の外国語としての日本語の聴解力と関連

する。

2-2: クローズ・テストは外国人の外国語としての日本語の書く力と関連する。

2-3: クローズ・テストは外国人の外国語としての日本語の話す力と関連する。

2-4: クローズ・テストは外国人の外国語としての日本語の読解力と関連する。

クローズ・テストが日本語の四技能と高い相関があると証明されるならば、日本語におけるクローズ・テストの利用価値は、非常に高いものと思われる。特に聴解力と話す力とを測定する方法上の実際上の困難さを考えるとき、これらとの相関が十分に高ければ、クローズ・テストの利用範囲がさらに拡大する。

III. 研究の方法と結果

1. 実験の方法

クローズ・テストの作成にあたっては、これまでに吟味されてきた方法に従って、

- 1). An Introduction to Modern Japanese, The Japan Times Ltd, (1978)⁵⁶ から 1 節
- 2). 「外国人学生用日本語教材書、初級（改証版）」早稲田大学語学教育研究所編(1977)¹⁰¹ から 1 節
- 3). 「日本語学力検定試験問題集 1」早稲田大学語学研究所編(1975)¹⁰² から 1 節
- 4). 「日本語学力検定試験問題集 2」早稲田大学語学研究所編(1981)¹⁰³ から 1 節

合計 4 節を選び出した。また、表記方法は、既習の漢字およびひらがなを使い、外来語にはカタカナを使用した。分ち書きは、国際基督教大学語学科日本語の集中日本語コース I で使用している教科書、Modern Japanese for

University Students Part I (1976)⁴²のローマ字分ち書きに従った。

合計4節から成るクローズ・テストのうち、基本的には7語ごとの語を削除したが、最初の1節(1)は、半数以上が時間や場所を表す名詞に当り、どんな語が入っても意味が通じてしまうので、最初の1節に限り、8語ごとの語を削除することにした。

- a. 被験者：国際基督教大学で開設されている日本語プログラム、集中日本語 I, 日本語 I～V, および上級日本語のうち、任意に集中日本語 I をとっている学生30名が抽出された。表1はその国籍別、性別のうちわけである。

表1 被験者の国籍別、性別分類

国 籍	男	女	
米 国	7	5	12
中 国 系	1	3	4
日 本	1	2	3
ド イ ツ	0	2	2
デ ン マ ー ク	1	1	2
フ ラ ン ス	1	0	1
韓 国	0	1	1
フ ィ リ ピ ン	1	0	1
イ ン ド ネ シ ア	0	1	1
バ ン グ ラ デ シ ュ	1	0	1
カ ナ ダ	0	1	1
タ イ	0	1	1
合 計	13	17	30

- b. テスト：実施されたテストは、(1)クローズ・テストと、通常おこなわれている学期末テストとしての、(2)聴解力テスト、(3)漢字の読み書きテスト、(4)構文テスト、(5)実力テスト、(6)話す力のテストである。これらのテストは表2に示す通り、さらに13の変数に分類される。但し、(3)漢字

の読み書きテストは、漢字圏、非漢字圏の学習者が混合のため変数から除外する。また、(4)構文テストの中に含まれている作文テストを、ここでは書く力のテストとする。

表2 実施されたテスト

	テ ス ト	項目数	分類前に属していたテストNo.	テストのタイプ	正答に与える点	誤答に与える点
1	ク ロ ー ズ テ ス ト	30	(1)	記述(1語)	1	0
2	聴解力(文型理解)テスト	49	(2)	5肢選択及び○×式	1	0
3	聴解力(長文理解)テスト	15	(2)	○×式	1	0
4	聴解力(合計)テスト	64	(2)			
5	構 文 (助 詞) テスト	19	(4)	5 肢 選 択	2	0
6	構 文 (文 法) テスト	21	(4)	5 肢 選 択	2	0
7	構文(動詞・形容詞の活用)テスト	20	(4)	記 述	3	0
8	構文(自動詞・他動詞)テスト	10	(4)	記 述	3	0
9	構 文 (合 計) テスト	70	(4)			
10	語 彙 テ ス ト	25	(4)と(5)	○×式 (4) 記 述 (5)	項目1~10=2 項11~25=1	0
11	読 解 力 テ ス ト	20	(4)と(5)	○×式	1	0
12	書く力のテスト(作文テスト)	5	(4)	記 述		
13	話 す 力 の テ ス ト		(6)	* 1分のSpeech		

* Speechの課題は「私の将来について」

- c. 実験日時、場所：1981年11月27日午前8時半から午後12時半まで(1)~(5)のテストが期末テストとしておこなわれた。クローズ・テストおよび聴解力テストはラボでおこなわれ、その他のテストはその後で、別の教室に移動しておこなわれた。1981年11月24日午前8時半から午前10時半の間に(6)話す力のテストを実施した。30人の被験者を一人ずつ呼び、約1分の即興スピーチをおこない、それを録音した。この日は通常の授業のテストが同時に実施されたため、スピーチのテストは被験者が通常の課のテストを受けている間におこなわれたので、スピーチの課題が前もっ

て知られてしまうことはない。

- d. 実験手続き：問題用紙（書き込み式）を配布し，(1)クローズ・テストが25分実施された。実施に際し，空所にもっとも適当だと思うことばを1語だけ入れるように指示がなされた。その後，前もって録音された指示に従って(2)聴解力テストが約45分実施された。別の教室へ移動してから(3)漢字の読み書きテスト，(4)構文テスト，(5)実力テストの順で実施された。この時学生は自己のペースでテストを受け，12時30分までに全員提出した。

e. 結果の処理方法：

1. クローズ・テストの平均点，標準偏差を求める。クローズ・テストの採点は時間的効率を重視して Exact word 法でおこなう。各テストについての点の与え方は表2に示した通りである。
2. 実施したテストについて（作文テスト，話す力のテストは除く）項目分析をおこなう。困難度，全得点と個別得点との相関係数を求める。テストの結果，得られた得点から中央値を求め，中央値で上位群，下位群に分け， χ^2 検定をおこなう。
3. 項目分析の結果から，困難度10%以上90%未満かつ全体得点と個別項目の相関係数が10%，5%，1%で有意であった項目のみを採用し，各テストの平均点，標準偏差を求める。
4. 実施したテスト（作文テスト，話す力のテストは除く）について，3の手続きを通過した項目のみを対象として，クロンバックの α 係数を求め，信頼度を検討する。
5. 実施されたテスト全部についてのピアソン (Pearson) のプロダクト・モーメント相関係数を求める。
6. すべてのテストのバリマックス法による因子分析をおこなう。
7. 基準変数をクローズ・テストとし，他の実施されたテストを予測変数とし，クローズ・テストに回帰させ，他のテストのクローズ・テストへの寄与率を求める。

8. 話す力のテストについては、3人の日本語の先生によって評価してもらう。5項目 ((1)プロナンスーションおよびイントネーション, (2)語彙 (適切度), (3)文法 (正確度), (4)流暢さ, (5)総合評価), 7段階評価を評価基準とする。3人の先生の評価の結果について一致係数(w)を求め有意であれば、この3人の先生の評価得点を加算し、話す力の得点とする。

2. 実験の結果

a. 平均点, 標準偏差

クローズ・テスト実施の結果得られた得点から、平均点, 標準偏差を求めた。また、その他のテストについては、項目分析の結果、困難度10%以上90%未満かつ全体得点と個別項目の相関係数が10%, 5%, 1%レベルで有意な項目のみを得点とし、その平均点, 標準偏差を求めたのが表3である。

表3 実施したテストの満点, 平均点, 標準偏差

		満 点	平 均 点	標準偏差
1	ク ロ ー ズ テ ス ト * 1	30(1x30)	18.90	4.604
2	ク ロ ー ズ テ ス ト 2	11(1x11)	6.97	2.189
3	ク ロ ー ズ テ ス ト 3	15(1x15)	9.60	3.729
4	聴 解 力 (文 型 理 解) テ ス ト	19(1x19)	13.40	4.264
5	聴 解 力 (長 文 理 解) テ ス ト	4(1x4)	6.10	2.057
6	聴 解 力 (合 計) テ ス ト	23	19.50	5.734
7	構 文 (助 詞) テ ス ト	8(2x4)	5.93	2.318
8	構 文 (文 法) テ ス ト	12(2x6)	9.00	3.227
9	構文(動詞・形容詞の活用)テスト	51(17x3)	33.80	10.685
10	構 文 (自 ・ 他 動 詞) テ ス ト	24(3x8)	18.10	5.921
11	構 文 (合 計) テ ス ト	93	66.83	18.488
12	語 彙 テ ス ト	(2x5) 23(1x13)	16.00	5.414
13	読 解 力 テ ス ト	8(1x8)	5.97	1.903
14	書く力のテスト(作文テスト)	20	16.10	3.263
15	話 す 力 の テ ス ト **	105	85.80	15.661

* クローズ・テストの項目分析の結果、クローズ・テストを1,2,3と三つに分類した。分類基準については、III—2—C「クローズ・テストの信頼性」参照

** 3人の先生の評価の一致係数($W=0.96$, $P<.01$)が有意であったので3人の評価得点を加算し話す力の得点とした。

b. クローズ・テストの妥当性および信頼性

クローズ・テストの項目分析：クローズ・テストの項目分析の結果が表 4 である。困難度が90%以上10%未満の項目は 7 項目（項目 1,4,5,7,10,26,28）あった。また、全得点と個別項目との相関係数が10%で有意な項目は 3 項目（項目2,20,25）、5 %で有意な項目は 5 項目（項目 6,11,12,15,21）、1 %で有意な項目は 7 項目（項目3,14,19,23,27,29,30）であった。北條(1979)²⁸においては、 x^2 検定で項目分析をおこなっていた。それと比較参照のため x^2 検定の結果も合せて表 4 に示した。

表 4 . クローズ・テストの項目分析

N = 30

	困 難 度	全体得点と個別 項目との相関係数	X^2	
1	100.00			
2	60.00	0.323 *	0.833	
3	63.33	0.476 ***	10.000	$P < 0.01$
4	100.00			
5	3.33	0.254	1.429	
6	73.33	0.424 **	5.568	$P < 0.05$
7	100.00			
8	70.00	0.194	0.000	
9	30.00	0.091	0.625	
10	100.00			
11	70.00	0.453 **	7.640	$P < 0.01$
12	40.00	0.395 **	4.537	$P < 0.05$
13	50.00	0.181	5.000	$P < 0.05$
14	86.67	0.499 ***	6.923	$P < 0.01$
15	33.33	0.398 **	4.471	$P < 0.05$
16	13.33	0.264	4.000	$P < 0.05$
17	76.67	0.195	2.301	
18	36.67	-0.013	0.023	
19	86.67	0.738 ***	6.923	$P < 0.01$
20	70.00	0.331 *	1.296	
21	40.00	0.444 **	8.356	$P < 0.01$
22	86.67	0.246	2.356	
23	63.33	0.519 ***	4.043	$P < 0.05$
24	56.67	0.235	1.833	
25	86.67	0.314 *	6.923	$P < 0.01$
26	100.00			
27	83.33	0.546 ***	2.222	
28	6.67	0.069	2.222	
29	36.67	0.471 ***	11.579	$P < 0.001$
30	66.67	0.608 ***	15.625	$P < 0.001$

* $P < .10$ ** $P < .05$ *** $P < .01$

クローズ・テストは四つの節からできているが、以上の結果から、第4節に弁別力の高い項目が集中していることがわかった。従って、第4節部分は他の部分よりも妥当性が高いように思われる。

c. クローズ・テストの信頼性

クローズ・テストの信頼性については、項目分析の結果を参考にして次の3種類のクローズ・テストの信頼性を求めた。

クローズ・テスト1：全項目の得点。

クローズ・テスト2：既述のように弁別力の高い項目が集中している節4節部分の得点。

クローズ・テスト3：項目分析の結果から困難度10%以上90%未満かつ

表5 クロンバックの α 係数による信頼度係数

N = 30

		クロンバックの α 係数	標準誤差
クローズテスト1	1	0.807	0.841
クローズテスト2	2	0.678	0.400
クローズテスト3	3	0.835	0.681
聴解力(文型理解)テスト	4	0.840 (0.837)*	0.778
聴解力(長文理解)テスト	5	0.753 (0.703)	0.376
聴解力(合計)テスト	6	0.875	1.047
構文(助詞)テスト	7	0.586 (0.582)	0.423
構文(文法)テスト	8	0.705 (0.687)	0.639
構文(動詞・形容詞の活用)テスト	9	0.845 (0.852)	1.951
構文(自・他動詞)テスト	10	0.703 (0.725)	1.081
構文(合計)テスト	11	0.896	3.375
語彙テスト	12	0.848 (0.834)	0.988
読解力テスト	13	0.694 (0.736)	0.347
書く力のテスト(作文テスト)	14	—————	0.596
話す力のテスト	15	—————	2.859

* () 内は、全項目数での α 係数

全体得点と個別項目との相関が10%， 5%， 1%
レベルで統計的に有意な項目の得点。

クローズ・テスト 1， クローズ・テスト 2， クローズ・テスト 3それぞれの信頼度をみるためにクロンバックの α 係数を求めたのが表 5 である。

結果をみると，クローズ・テスト 1 は $\alpha = 0.807$ ， クローズ・テスト 2 は $\alpha = 0.678$ ， クローズ・テスト 3 は $\alpha = 0.835$ であった。以上の結果から，クローズ・テストの信頼性は全体的に高く，内部的統一性が高いといえる。クローズ・テスト 2 の α 係数が他の値より低かったのは，クローズ・テスト 2 の項目数（11項目）が少なかったためと思われる。

d. クローズ・テストの規準関連妥当性

クローズ・テスト 1, 2, 3 と実施されたすべてのテスト 15 変数のピアソン (Pearson) プロダクト・モーメント係数を算出した結果が表 6 である。この結果から，クローズ・テストは他の実施されたテストと 1%あるいは 0.1%レベルの有意な正の相関があることがわかった。特に，聴解力(合計)テスト，構文（文法）テスト，語彙テスト，作文テスト，話す力のテストと 0.1%レベルの有意に高い相関が認められた。以上のことから，クローズ・テストは，規準関連妥当性の高いテストであるといえる。

e. バリマックス法による因子分析の結果

バリマックス法による因子分析をおこなったところ，表 7 のような結果が得られた。クローズ・テスト 1, 2, 3 それぞれについておこなった結果次の 2 因子を得た。

第1因子：助詞（構文）、動詞、形容詞の活用（構文）、自・他動詞（構文）、語彙テスト、読解力テスト、作文テスト。

第2因子：クローズ・テスト1,2,3、聴解力テスト、文法（構文）、話す力のテスト。

以上のことから、実施されたすべてのテストは第1因子として読む力、書く力の因子、第2因子として話す力、聴く力の因子に分けることができる。しかし、第2因子として得られたクローズ・テストは第1因子としても高い値を示している。

f. 重回帰分析の結果

クローズ・テスト1,2,3をそれぞれ基準変数とし、その他実施されたテストを予測変数としてクローズ・テスト1,2,3に回帰させ、各テストのクローズ・テストへの寄与率を求めた結果が表8である。

		第1因子	第2因子
クローズ・テスト 2		0.50329	0.57054
	1	0.56843	0.64389
	2	0.24751	0.68812
	3	0.79169	0.15343
	4	0.47360	0.57920
	5	0.86299	0.19743
	6	0.57030	0.38426
	7	0.79747	0.42718
	8	0.51893	0.41102
	9	0.73449	0.42934
	10	0.14573	0.89856
		第1因子	第2因子
クローズ・テスト 3		0.43065	0.74246
	1	0.57292	0.65256
	2	0.23979	0.68235
	3	0.78939	0.17279
	4	0.46802	0.59149
	5	0.81177	0.18157
	6	0.57827	0.37004
	7	0.79031	0.43630
	8	0.52313	0.41369
	9	0.71554	0.43476
	10	0.14743	0.89322

表 8 重回帰分析

従属変数 (基準変数) クローズ・テスト 1

独立変数 (予測変数)	重相関係数(R)	決定係数(R^2)	決定係数の変化の増分	単純相関係数
書く力のテスト (作文テスト)	0.74906	0.56109	0.56109	0.74906
話す力のテスト	0.83241	0.69291	0.13181	0.65777
構文(文法)テスト	0.84773	0.71864	0.02573	0.68932
語彙テスト	0.86390	0.74632	0.02768	0.74706
聴解力(文型理解) テスト	0.87877	0.77224	0.02592	0.65206
聴解力(長文理解) テスト	0.88747	0.78761	0.01537	0.61282
構文(助詞・形容 詞の活用)テスト	0.89490	0.80084	0.01323	0.53582
構文(助詞)テスト	0.89995	0.80992	0.00908	0.55503
読解力テスト	0.90355	0.81641	0.00649	0.50745
構文(自・他動詞) テスト	0.90396	0.81714	0.00073	0.50509

従属変数 (基準変数) クローズ・テスト 2

独立変数 (予測変数)	重相関係数(R)	決定係数(R^2)	決定係数の変化の増分	単純相関係数
書く力のテスト (作文テスト)	0.75849	0.57531	0.57531	0.75849
聴解力(長文理解) テスト	0.80064	0.64103	0.06571	0.62107
構文(文法)テスト	0.82290	0.67716	0.03614	0.59065
聴解力(文型理解) テスト	0.86134	0.74191	0.06475	0.52979
語彙テスト	0.87152	0.75956	0.01764	0.68958
読解力テスト	0.88288	0.77947	0.01992	0.43855
構文(動詞・形容 詞の活用)テスト	0.89054	0.79306	0.01369	0.50096
構文(自・他動詞) テスト	0.89390	0.79906	0.00600	0.45520
話す力のテスト	0.89526	0.80149	0.00243	0.54898
構文(助詞)テスト	0.89720	0.80497	0.00348	0.44799

従属変数(基準変数) クローズ・テスト3

独立変数 (予測変数)	重相関係数(R)	決定係数(R^2)	決定係数の変化の増分	単純相関係数
語彙テスト	0.71060	0.50495	0.50495	0.71060
話す力のテスト	0.81546	0.66498	0.16003	0.69951
書く力のテスト (作文テスト)	0.83863	0.70330	0.03833	0.70918
構文(動詞・形容 詞の活用)テスト	0.85741	0.73514	0.03184	0.43154
構文(文法)テスト	0.87400	0.76387	0.02872	0.67631
聴解力(文型理解) テスト	0.88692	0.78662	0.02275	0.66327
聴解力(長文理解) テスト	0.90086	0.81156	0.02493	0.66180
構文(助詞)テスト	0.90559	0.82009	0.00853	0.49144
読解力テスト	0.90655	0.82183	0.00174	0.50845
構文(自・他動詞) テスト	0.90678	0.82224	0.00041	0.45169

クローズ・テスト1には、作文テスト、話す力のテストが、クローズ・テスト2には、作文テスト、聴解力テスト、文法テストが、クローズ・テスト3には、語彙テスト、話す力のテストが大きく寄与している。

クローズ・テスト1,2,3それぞれについて各テストの寄与率が異っている。この結果から、クローズ・テストの得点とすべき項目によって寄与するテストが異っていることがわかる。このことは、測定しているものがちがっていると考えるよりも、おそらくは、材料と学習者との関係、つまり、クローズ・テストの材料の難易による差がでてきたものと思われる。しかし、いずれにしても、クローズ・テストの材料の選択はクローズ・テストの効果を規定する大きな要因のように思われる。以上のことを考えると、今後、クローズ・テストそのものをクローゼントロピー法などにより検討していく必要があると思われる。

3. 実験結果の考察

仮説1の検証のため、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、全体得点と個別項目と

の相関係数を求めた。また、クロンバックの α 係数を求め、内部的統一性を調べた。クローズ・テストと他のテストについてピアソンのプロダクトモーメント係数を算出した。

全体得点と個別項目との相関係数が有意な項目は、30項目中15項目（10％レベル3項目，5％レベル5項目，1％レベル7項目）であった。クロンバックの α 係数は， $\alpha = .807$ （クローズ・テスト1）， $\alpha = .678$ （クローズ・テスト2）， $\alpha = .835$ （クローズ・テスト3）で，内部統一性の高いテストであることが検討された。さらに，クローズ・テスト1, 2, 3と他のテストとのピアソンプロダクトモーメント係数を求め，クローズ・テストの妥当性を調べた。その結果，クローズ・テストと他のテストとは，0.1％あるいは1％レベルの有意な正の相関があった。ゆえに，クローズ・テストは規準関連妥当性の高いテストであるといえよう。以上のことから「仮説1：クローズ・テストは外国人の外国語としての日本語の熟達度テストとして妥当性および信頼性の高いテストである。」は支持されたといえる。

仮説2-1：クローズ・テストは外国人の外国語としての日本語の聴解力と関連する。

仮説2-1の検証のため，クローズ・テスト1, 2, 3と聴解力テストとの相関係数を求めた。その結果は表9の通りである。

表9 クローズ・テストと聴解力テストとの相関係数

N = 30

	聴解力(文型)テスト	聴解力(長文)テスト	聴解力(合計)テスト
クローズ・テスト1	$r = .65 \quad P < .001$	$r = .61 \quad P < .001$	$r = .70 \quad P < .001$
クローズ・テスト2	$r = .53 \quad P < .001$	$r = .62 \quad P < .001$	$r = .62 \quad P < .001$
クローズ・テスト3	$r = .66 \quad P < .001$	$r = .66 \quad P < .001$	$r = .73 \quad P < .001$

表10 クローズ・テストと書く力(作文テスト)との相関係数

	話す力のテスト	N = 30
クローズ・テスト 1	$r = .75$ $P < .001$	
クローズ・テスト 2	$r = .76$ $P < .001$	
クローズ・テスト 3	$r = .71$ $P < .001$	

表11 クローズ・テストと話す力との相関係数

	話す力のテスト	N = 30
クローズ・テスト 1	$r = .66$ $P < .001$	
クローズ・テスト 2	$r = .55$ $P < .001$	
クローズ・テスト 3	$r = .70$ $P < .001$	

この結果、クローズ・テスト1, 2, 3と聴解力テストとの相関は、 $r = .53$ から $r = .73$ で、0.1%レベルの有意な高い相関であった。よって仮説2-1は支持されたといえよう。

仮説2-2: クローズ・テストは外国人の外国語としての日本語の書く力と関連する。

仮説2-2の検証のため、クローズ・テスト1, 2, 3と書く力（作文テストを書く力のテストとする）との相関係数を求め、表10にまとめた。

この結果から、クローズ・テスト1, 2, 3と書く力のテストとの間に $r = .71$ から $r = .76$ で0.1%レベルの有意な高い相関が認められた。ここで、作文テストが書く力を測定し得るという前提で、クローズ・テストは書く力と有意な高い正の相関があるといえる。よって、仮説2-2は支持されたといえるであろう。

仮説2-3: クローズ・テストは外国人の外国語としての日本語の話す力と関連する。

仮説2-3を検証するために、クローズ・テスト1, 2, 3と話す力のテストとの相関係数を求め、表11に示した。

この結果から、クローズ・テストは話す力のテストと $r = .55$ から $r = .70$ で0.1%レベルの有意な高い相関があった。よって仮説2-3は支持されたといえよう。

仮説2-4: クローズ・テストは外国人の外国語としての日本語の読解力と関連する。

仮説2-4を検証するため、クローズ・テスト1, 2, 3と読解力テストとの相関係数を求めた。その結果が表12である。

この結果から、クローズ・テストと読解力テストとの間に $r = .44$ から $r = .61$ で1%の有意な正の相関が認められた。よって、仮説2-4は支持されたといえよう。

<仮説1><仮説2-1><仮説2-2><仮説2-3><仮説2-4>のうち、すべての仮説が支持された。この結果からみて、クローズ・テストは、本実験に限

表12 クローズ・テストと読解力テストとの相関係数

	読解力テスト	N = 30
クローズ・テスト 1	$r = .51$ $P < .01$	
クローズ・テスト 2	$r = .44$ $P < .01$	
クローズ・テスト 3	$r = .51$ $P < .01$	

り、日本語の熟達度テストとして妥当性および信頼性の高いテストであるといえる。また、日本語の総体的な能力を測定するテスト技法として、日本語教育へ適用できるのではないかとと思われる。

しかし、日本語のクローズ・テスト実施については本実験が初めての試みなので、今後も追実験をおこなう必要がある。また、今後の実験については、1)被験者を増やす（今回はN=30）、2)クローズ・テストの作成法の問題点をさらに検討する、3)クローズ・テストをもう少し長い文章を使う等の点に留意しておこないたいと思う。

IV. 考察と今後の課題

外国語の総体的な能力を測定するテストに、熟達度テストがあげられる。外国語の熟達度テストとしてのクローズ・テストは1)妥当性、2)信頼性、3)効率性、4)実用性の高いテストとして高く評価されている。英語の熟達度としてのクローズ・テストは、妥当性、信頼性、効率性、実用性が高く、特に外国（非英語圏）での研究によれば、英語の聴解力と有意な高い相関があると報告されている。これらの参考研究を参考として、クローズ・テストを日本語へ応用してみた。その結果、日本語のクローズ・テストも、妥当性、信頼性ともに高く、本実験においては、日本語の総体的な能力を測定する熟達

度テストとして適用できるのではないかという結果が得られた。

しかし本研究では、日本語熟達度テストとしてのクローズ・テスト作成は初めての試みであったので、今後クローズ・テストの作成方法について、基礎的な研究をおこなう必要がある。クローズ・テストは、規則的に n 語ごとの語を抜いていくという点でテスト全体が一つのまとまりであって、クローズ・テスト自体の妥当性、信頼性を左右するのはクローズ・テストの内容、特に文章の難易度にあるといえよう。これらのことを考慮したうえで、日本語熟達度テストとしてのクローズ・テストを今後再検討していきたいと思う。

さらに、クローズ・テストの今後の研究課題を列記すると、次のような研究領域または課題を挙げることができよう。

1. 人間の情報処理能力の視点からのクローズ・テストの理論的枠組みの検討。
2. 母国語としての国語能力の測定法としてのクローズ・テストの活用。
3. クローズ・テストにおける誤り（特に、エグザクト法）の意味の検討。
4. クローズ・テスト（総合テストとしての）と、不連続点テストとの関係の吟味。特に、クローズ・テストによる結果の適用範囲、または限界の検討。

参 考 文 献

1. Abu-Sayf, F.K., Herbolich, J.B. and Spurling, S. "The Identification of the Major Components for Testing English as a Foreign Language." *TESOL Quarterly*, 1979, 13,1, 117-120.
2. Aitken, K.G. "Using Cloze Procedure as an Overall Language Proficiency Test." *TESOL Quarterly*, 1977, 11,1, 59-67.
3. Alderson, J.C. "The Cloze Procedure and Proficiency in English as a Foreign Language." *TESOL Quarterly*, 1979, 13,2, 218-219.
4. Banathy, B.H. *Instructional Systems*, Fearon Publishers, 1968.
5. Bloom, B.S., Hastings, J.T. and Madaus, G.F. *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill, 1971.
6. Briere, E.J. "Are We Really Measuring Proficiency with Our Foreign Language Test?" *Foreign Language Annals*, 4, 385-391. Reprinted in Allen and Campbell, 1977, 321-330.
7. Carroll, J.B. 『英語の評価と教授』 大学英語教育学会編訳注 大修館 1972.
8. Carroll, J.B. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students." in Allen and Campbell, 1972.
9. Carroll, J.B. and Freedle, R.O. *Language Comprehension and the Acquisition of Knowledge*. (Washington, V.H.), Halsted Press Division of Wiley, New York, 1972.
10. Carroll, J.B. *Testing Communicative Performance: An Interim Study*. Oxford, Pergamon Press, 1980.

11. Chihara, T., Oller, J., Weaver, K. and Chavez-Oller, M.A. "Are Cloze Items Sensitive to Constraints across Sentences?" *Language Learning*, 1977, 27,1, 63-73.
12. Chomsky, N. *Aspects of the Theory of Syntax*. MIT Press, 1965.
13. Clark, J.L.D. "Theory and Practice." *Foreign Language Testing*, The Center for Curriculum Development Inc., 1972.
14. Darnell, D.K. "The Development of an English Language Proficiency Test of Foreign Students, Using a Clozen-Tropy Procedure." U.S.D Department of Health, Education and Welfare, 1968.
15. Darnell, D.K. "Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students." *Speech Monographs*, 1970, 37, 36-46.
16. Dieterich, T.G., Freeman, C. and Crandall, J.A. "A Linguistic Analysis of Some English Proficiency Tests." *TESOL Quarterly*, 1979, 13,4, 535-550.
17. Evans, G.H.K. "Experiments with Cloze Procedure." *English Language Teaching*, 1976, 31,1.
18. Farhady, H. "The Distinctive Fallacy between Discrete-Point and Integrative Tests." *TESOL Quarterly*, 1979, 13,3, 347-358.
19. Finocchiaro, M. *English as a Second Language: From Theory to Practice*. New York, Regents Publishing Co., Inc., 1964.
20. Fries, C.C. *Teaching and Learning English as a Foreign Language*. University of Michigan Press, 1945.
21. Gronlund, N.E. (ed.), *Reading in Measurement and Evaluation; Education and Psychology*. New York, MacMillan, 1968.
22. Gronlund, N.E. *Measurement and Evaluation in Teaching*, (3rd ed.) New York, MacMillan, 1976.
23. Gronlund, N.E. *Measurement and Evaluation in Teaching*, (4th

- ed.) New York, MacMillan, 1981.
24. Harris, D.P. Testing English as a Second Language. McGraw-Hill, 1969.
 25. Haskell, J.F. "Refining the Cloze Testing and Scoring Procedures for Use With ESL Students." *DAI*, 1973, 34,6, 3206-A.
 26. Hirst, W, Spelke, E.S., Reaves, C.C., Caharack, C. and Neisser, U. "Dividing Attention without Alternation or Automaticity." *Journal of Experimental Psychology*, 1980, General 109, 98-117.
 27. Hisama, K.K. "Design and Empirical Validation of the Cloze Procedure for Measuring Language Proficiency of Non-Native Speakers." *DAI*, 1977, 37,9-A, 5766-A.
 28. 北條 礼子 『英語のテスト技法としてのクローズ法に関する実証的研究』, 国際基督教大学大学院教育学研究科提出教育学修士論文, 1979.
 29. Holtzman, P. "English Language Proficiency Testing and the Individual." Selected Conference Papers of the Association of Teachers of English as a Second Language, 1967.
 30. 池田 央 『新しいテスト問題作成法』 第一法規出版, 1980.
 31. 池田 央 『テストで能力がわかるか』 日本経済新聞社, 1978.
 32. 井上 健治編 『テストの話』 中央公論社, 1978.
 33. Irvine, P., Atai, P. and Oller, J.W.Jr. "Cloze Dictation and the Test of English as a Foreign Language." *Language Learning*, 1974, 24,2, 245-252.
 34. 石田 敏子, 稲垣 滋子, 中村 妙子, 仲野 桂子, 堀口 純子, 「日本語教育用テストの標準化 I, 国際基督教大学における予備テストの結果」, 『日本語教育』日本語教育学会 4 号, 1980.
 35. 石田 敏子, 稲垣 滋子, 中村 妙子, 仲野 桂子, 宮崎 寿子, 「日本語教育用テストの標準化 II — 国内における結果 —」

- Annual Reports*, 6, ICU The Division of Language, 1981, 31-47.
36. 石垣貴千代 「初級日本語の文型テストについて — 書き試験 —」『日本語教育 特集：評価と標準テスト』 1977, No32, 47-61.
 37. 伊藤 弘子 “Cloze Test.”『英語教育』 1978 (5月～8月)
 38. Jespersen, O. *How to Teach a Foreign Language*. London, George Allen and Unwin Ltd., 1904.
 39. Johnson, D.C. “The Evaluation of Comprehension When English is a Second Language by Use of the Cloze Procedure.” *DAI*, 1975, 35,8-A, 5121-A.
 40. 垣田 直己編集 『英語教育学研究ハンドブック』, 大修館, 1979.
 41. 垣田 直己他編 『言語教育学叢書 第1期6』, 文化評論出版, 1967.
 42. 小出 詞子編集 『Modern Japanese for University Students, Part I.』 国際基督教大学, 1976.
 43. 国際交流基金 『日本語試験問題分析報告 昭和51年度海外日本語講座成績優秀者研修会第一次選考試験』, 1979.
 44. 輿水 実 「言語テストの種類と作り方」『日本語教育』 1964, No4-5, 2-12.
 45. Lado, R. *Linguistic Science and Language Tests*. Tokyo, Taishukan, 1950.
 46. Lado, R. “Survey of Tests in English as a Foreign Language.” *Language Learning*, 1950, 3,1-2, 51-66.
 47. Lado, R. “Testing Control of the Structure of a Foreign Language.” *Language Learning*, 1951, 4,1-2, 17-41.
 48. Lado, R. 『言語科学と言語テスト』, 西岡淑雄訳注 英語教育シリーズ 16, 大修館, 1959.
 49. Lado, R. *Language Testing: The Construction and Use of Foreign Language Tests*. McGraw-Hill, 1961.
 50. Lado, R. *Linguistics Across Cultures, Applied Linguistics for Language Teachers*. McGraw-Hill, 1961.
 51. Lado, R. 『言語テスト：外国語テストの作成とその利用』, 門司 勝

他訳, 大修館, 1959.

52. MacGinitie, W.H. "Contextual Constraint in English Prose Paragraphs." *Journal of Psychology*, 1961, 51, 1121—1130.
53. Mackey, W.F. 『言語教育分析 (Language Teaching Analysis)』, 伊藤 健三他共訳, 大修館, 1979.
54. Meara, P. "Two Tests of Overall Ability in a Foreign Language." *Audio-Visual Language Journal*, 1974—75, 12, 3.
55. Miller, G.A. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." *The Psychological Review*, 1956, 63, 81—97.
56. Mizutani, O. and Mizutani, N. An Introduction to Modern Japanese. The Japan Times, Ltd., 1978.
57. Mussen, K.A. "An Alternative to the Cloze Test of TESOL '79." *The Learner in Focus TESOL*, 187—192.
58. 中野 照海 「評価と統計学の基礎的知識」『英語教育』1968 (7月), 7—9, 96.
59. 中野 照海 「メディアの効果をめぐって——視聴覚教育の評価に関する覚書1——」『視聴覚教育』1981, 35, 1, 24—27.
60. 日本語教育編集委員会 「実例からみた日本語の試験問題——出題意図と形式による分類」『日本語教育』1964, No4, 5, 13—59.
61. 日本語教育学会 『外国人のための日本語能力検定試験に関する調査研究の経過報告, 1980年2—3月日本語能力検定予備試験』, 1981.
62. 沼野 一男 『教育工学』NHK 市民大学叢書 日本放送出版協会, 1973.
63. 大友 賢二 「外国語テストの最近の動向」『英語教育』1973, 22 (7月増刊号)
64. 大友 賢二 「テストと評価の問題点」『英語教育』創刊25周年記念増刊号, 1977(9月), 98—100.

65. Ohtomo, K. "Problems and Analysis: Testing Overall English Language Proficiency of Japanese Students." *The Teaching of English in Japan*, Koike, I. et. Al. (ed.) Tokyo, Eichosha, 1978, 463-477.
66. 岡田 明 『言語教育の心理』, 新光閣書店, 1971.
67. Oller, J.W.Jr. and Nevin, I. "A Cloze Test of English Prepositions." *TESOL Quarterly*, 1971, 5,3, 315-326.
68. Oller, J.W.Jr. and Conrad, C.A. "The Cloze Technique and ESL Proficiency." *Language Learning*, 1971, 21,2, 183-196.
69. Oller, J.W.Jr., Bowen, J.D., Dien, T.T. and Mason, V.W. "Cloze Tests in English, Thai, and Vietnamese: Native and Non-Native Performance." *Language Learning*, 1972, 22,1, 1-16.
70. Oller, J.W.Jr. "Assessing Competence in ESL: Reading." *TESOL Quarterly*, 1972, 6,4, 313-323.
71. Oller, J.W.Jr. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language." *The Modern Language Journal*, 1972, 56,3, 151-158.
72. Oller, J.W.Jr. "Dictation as a Test of ESL Proficiency." *Focus on the Learner*, Massachusetts, Newbury House, 1973.
73. Oller, J.W.Jr. "Discrete-Point Tests Versus Tests of Integrative Skills." *Focus on the Learner*, Massachusetts, Newbury House, 1973.
74. Oller, J.W.Jr. "Cloze Tests of Second Language Proficiency and What They Measure." *Language Learning*, 1973, 23,1, 105-118.
75. Oller, J.W.Jr. "Pragmatic Language Testing." *Language Science*, 1973, 28, 7-12.
76. Oller, J.W.Jr. and Nagata, N. "The Long-Term Effect of FLES: An Experiment." *The Modern Language Journal*,

- 1974, 58,1-2, 15-19.
77. Oller, J.W.Jr. *Language Tests at School: A Pragmatic Approach*.
London, Longman, 1979.
 78. Oller, J.W.Jr. and Perkins, K. *Research in Language Testing*.
Rowley, Mass., Newbury House, 1980.
 79. 大内 茂男編 『講座 英語教育工学 第5巻 研究と評価』, 研究社,
1973.
 80. 尾崎 明人, 小西 久也 「電子計算機を利用した初級段階における格
助詞習得度テスト」『東海大学紀要』留学生別科, 1978,
Vol.2, 15-54
 81. Perren, G.E. "Testing Ability in English as a Second Language: 1,
Problems." *ELT*, 1967, 21,2, 99-106.
 82. Rivers, W.M. "Rules, Patterns, and Creativity in Language
Learning." *ETF*, Croft, Kenneth, (ed.) 1972, 8,6.
 83. 芝 祐順 『因子分析法 東大出版会』, 1972.
 84. 芝 祐順 「読み易さの測り方 — クローズ法の日本語への適用 —
」『心理学研究』 1957, 28,2, 67-73
 85. 下瀬川慧子, 若松 久恵, 川幡愛恵美, 宮城 幸枝 「中級日本語の語
法例集 — いくつかの日本語教科書から — 」『東海大学
紀要』留学生別科, 1978 (2月)
 86. Spolsky, B. "Language Testing — The Problem of Validation."
TESOL Quarterly, 1968, 2,2, 88-94.
 87. Spolsky, B., Sigurd, B., Sato, M., Walker, E. and Anterburn,
C. "Preliminary Studies in the Development
Techniques for Testing Overall Second Language
Proficiency." *Language Learning*, 1968, Special Issue,
3, 79-102.
 88. Spolsky, B. *Reduced Redundancy as a Language Testing Tool*
Perren, George E., (ed.) 1971.
 89. Spolsky, B. "What Does It Mean to Know a Language; or How Do
You Get Someone to Perform His Competence?"

Focus on the Learner, Massachusetts, Newbury House, 1973, 164-176.

90. Sweet, H. *The Practical Study of Language: A Guide for Teachers and Learners*. London, Oxford University Press, 1964.
91. 高見沢 孟 「米商務省日本語研修所の場合」『日本語教育 特集：評価と標準テスト』 1977, No32, 35-46.
92. 武井 一美 「私費外国人留学生統一試験について」『前掲』 1-6.
93. Taylor, W.L. "Cloze Procedure: A New Tool for Measuring Readability." *Journalism Quarterly*, 1953, 415-453.
94. Taylor, W.L. "Recent Developments in Use of 'Cloze Procedure'." *Journalism Quarterly*, 1956, 33, 42-48.
95. Taylor, W.L. "'Cloze' Readability Scores as Indices of Individual Differences comprehension and Aptitude." *Journal of Applied Psychology*, 1957, 41, 1, 19-26.
96. 東京外国語大学外国語学部附属日本語学校 『外国人留学生の日本語能力の標準と測定——外国人留学生に対する大学入学のための日本語教育の内容(5)—— =測定=』 1981.
97. 植松 清 「テストを「テスト」する」『日本語教育 特集：評価と標準テスト』 1977, No32, 15-35.
98. Upshur, J.A. "Objective Evaluation of Oral Proficiency in the ESOL Classroom." *TESOL Quarterly*, 1971, 5, 1, 47-59.
99. Valdman, J.A. 『言語教育の基本問題 (Trends in Language Teaching.)』 鳥居 次好他訳注, 大修館, 1974.
100. 早稲田大学語学教育研究所編 『外国学生用日本語教科書初級(改訂版)』 1977.
101. 早稲田大学語学教育研究所編 『外国人留学生の日本語能力の標準と測定——日本語学力検定試験問題の分析を中心に——』 1981.
102. 早稲田大学語学研究所編 「日本語学力検定試験問題集 1」 1975.
103. 早稲田大学研究所編 「日本語学力検定試験問題集 2」 1981.
104. 山田 純 「日本語クローズ法の基礎的研究」『読書科学』 1979, X

X II, 1-2, 10-18.

105. 四方 実一 「心理テストの妥当性 その1,2,3」『教育心理』1962, 10月, 11月, 12月号.
106. 吉川 武時 「国費留学生現地採用試験（日本語）について」『日本語教育 特集：評価と標準テスト』1977, No.32, 7-14.

Measurement of Japanese Language Ability by the Cloze Method

Kimie Shin

I. BACKGROUND

A proficiency test which can measure the overall Japanese language ability has not yet been developed with regard to Japanese language education as a foreign language.

The cloze test (or cloze procedure), a highly regarded procedure developed by W. Taylor (1953), originally measured sentence readability. In this test, words are systematically (usually every n th word) deleted, and subjects are required to fill in the blanks with the words which they think are most appropriate.

In contrast to the discrete-point test which measures elements of language individually, the cloze test as a foreign proficiency test can be said to be an integrative test which attempts to measure the subject's integrated language ability. Validity, reliability, efficiency and practicality can be regarded as conditions for a good test and from many supportive researches, the English cloze test is proven to be a test with high validity, reliability, and practicality.

As a step in Japanese language education toward the development of a Japanese language proficiency test for foreigners, the author would like to investigate the suitability of the Japanese language cloze test as a proficiency test for Japanese as a foreign language. The investigation will be achieved through the application of the cloze test in the Japanese

language.

II. PURPOSE

The primary purpose of this research is to see whether or not the Japanese language cloze test, with high validity and reliability as a testing device for measuring English language proficiency, is a suitable test which can measure the Japanese language proficiency of people whose native language is not Japanese.

The secondary purpose is to see the relationship between the results obtained from the cloze test and the subject's aural comprehension, writing ability, speaking ability, and reading comprehension in the Japanese language.

HYPOTHESIS 1 : The cloze test is a test which has high validity as well as reliability as a Japanese language proficiency test for foreigners.

HYPOTHESIS 2-1: The cloze test relates to the aural comprehension of foreigners in the Japanese language.

HYPOTHESIS 2-2: The cloze test relates to the writing ability of foreigners in the Japanese language.

HYPOTHESIS 2-3: The cloze test relates to the speaking ability of foreigners in the Japanese language.

HYPOTHESIS 2-4: The cloze test relates to the reading comprehension of foreigners in the Japanese language.

III. METHOD

In November, 1980, 1) cloze test (also as the external criteria), 2) aural comprehension test, 3) sentence construction test, 4) vocabulary test, 5) reading comprehension test, 6) composition test, and 7) speaking ability

test were carried out. The Exact Word Scoring Method was used in scoring the cloze test.

In order to evaluate the reliability of the cloze test, item analysis of the cloze test was used, and 1) difficulty, 2) critical values of the correlation coefficient between the total score and each item were determined, and 3) χ^2 -test was employed.

Also, in order to examine the internal consistency of the cloze test, Cronbach's α was determined. After eliminating the items with difficulty above 90% and below 10% as well as items with low-discriminating ability (critical values of the correlation coefficient of $r < .306$) from all of the other test that were carried out, the Cronbach's α was computed. In order to evaluate the criterion-related validity, the Pearson Product-Moment Coefficient Correlations was also computed.

Furthermore, in order to evaluate the relationship between the cloze tests and 1) aural comprehension, 2) writing ability, 3) speaking ability, and 4) reading comprehension, the Pearson Product-Moment Coefficient Correlations were determined. Multiple Regression and Factor Analysis were carried out as well.

IV. RESULTS AND DISCUSSION

From the results of the cloze test's item analysis, the items in which the critical values of the correlation coefficient between the total score and individual items were significant were fifteen items out of thirty. The cloze test is made up of four clauses, but in the fourth clause concentration is on the significant items of the critical values of the correlation coefficient between the total score and individual items.

Thus, the fourth clause is thought of as having higher reliability than the other clauses. The reliability of the three cloze tests was determined

by referring to the results of the item analysis.

CLOZE TEST 1 : The score of all items had an $\alpha = .807$.

CLOZE TEST 2 : The score of the fourth clause section had an $\alpha = .678$.

CLOZE TEST 3 : From the results of item analysis of cloze test 1, the score of the statistically significant item had an $\alpha = .835$.

Moreover, in order to investigate the validity of cloze tests 1, 2, and 3, the Pearson Product-Moment Correlations were determined from the fifteen variables. As a result, cloze test 1, 2, and 3 had significantly high correlation with the other tests which were carried out, thus indicating that this is a test with high criterion-related validity.

Therefore, it can be said that Hypothesis 1 has been supported. In order to verify Hypothesis 2-1, by determining the correlation coefficient between cloze test 1, 2, and 3 and aural comprehension, a definitely significant correlation from $r = .53$ to $r = .73$ ($p < .001$) was found. Therefore, it can be said that Hypothesis 2-1 has been supported. In order to verify Hypothesis 2-2, by determining the correlation coefficient between cloze test 1, 2, and 3 and writing ability, a definitely significant correlation from $r = .71$ to $r = .76$ ($p < .001$) was found. Therefore, it can be said that Hypothesis 2-2 has been supported.

In order to verify Hypothesis 2-3, by determining the correlation coefficient between cloze test 1, 2, and 3 and speaking ability, a definitely significant correlation from $r = .55$ to $r = .70$ ($p < .001$) was found. Therefore, it can be said that Hypothesis 2-3 has been supported. In order to verify Hypothesis 2-4, by determining the correlation coefficient between cloze test 1, 2, and 3 and writing ability, a definitely significant correlation from $r = .44$ to $r = .51$ ($p < .01$) was found. Therefore, it can be said that Hypothesis 2-4 have been supported.

Hypothesis 1,2-1,2-2,2-3, and 2-4 have been supported. Thus, according to this research, the Japanese language cloze test can be said to be a suitable integrative test which can measure the proficiency in Japanese language as a foreign language.