

## 採点基準の違いが読解のリコールに及ぼす影響 —大学院生の場合—

平 野 絹 枝

(平成7年10月31日受理)

### 要 旨

リコールテストは読解力を測定するのによく使用される方法の一つである。読解におけるリコール (recall) を採点する時採点基準を厳密にすべきかどうかという評価の問題がある。もとの文章の内容を忠実に復元させる厳密な採点基準と大体の情報が再生されていれば正解とするゆるやかな採点基準がある。本研究では、リコール・プロトコル (recall protocol) の異なった採点基準による採点 (厳密な採点基準とゆるやかな採点基準) が、(1)テキスト全体、(2)テキスト中の重要な情報(主題文全体)、(3)各パラグラフ全体、(4)各パラグラフの主題、の再生量にどのような影響を及ぼすかを分析、検討した。その結果、(1)テキスト全体のリコールでは、2種類の採点基準の相関は、極めて高かった ( $r=.96$ )。しかし(2)重要な情報量(主題文全体)のリコールでは、相関は、.46であった。また(3)各パラグラフ全体と(4)各パラグラフの主題間の再生率のパターンが、採点基準の違いによって、異なることが明らかになった。

### KEY WORDS

recall	リコール	recall protocols	リコール・プロトコル
assessment	評価	strict criterion	厳密な採点基準
loose criterion	ゆるやかな採点基準		

### 1. は じ め に

再生テストであるリコール・タスク (recall task) は読解力を測定するテストの一つとして、L1 (第1言語) だけでなく L2 (第2言語) でも使用されている。読解のリコールテストとは、学習者にテキストを読ませた後、それを見ないで、おぼえているテキストの内容を想起、再生させるテストである。読み手は理解している内容は、再生できるのであり (Bransford 1979)、テキストの理解と再生は密接な関係があると言われている。リコール・プロトコル (recall protocol) によって、読み手がどのように、情報を組織化し、どのような想起方略を使用しているか、どのように情報を再構築するかなどを知ることが出来る (Bernhardt 1983: 31)。Bernhardt

(1991: 200) は、リコール・プロトコルは、読解力をより純粋に測定すると主張している。特に L2 テキストのプロトコルが読み手の母語で書かれている時は、そのリコール・プロトコルによって、テキストの意味の予測、体系化、類推といった読み手の論理的操作がどのようにテキストの語彙や統語についての理解と相互作用しあっているかが明らかになり、どの程度 L2 学生が理解しているかがよりよくわかる (Swaffar et al. 1991: 164)。

リコールに影響を及ぼす要因としては、テキスト構造、リコールする時の言語 (L1 か L2 か)、英文を読む前に行なうリコールの指示の影響 (後でリコールを要求することを英文を読む前に知らせておくかどうか)、英語学力のレベル等が、先行研究で報告されている。

Carrel (1984) は、ESL (English as a Second Language) の大学生を被験者にして実験を行った結果、テキストの構成の型 (比較、原因と結果、問題解決、記述の型) の違いによって、読解のリコールの成績が異なることを明らかにしている。すなわち比較、原因と結果、問題解決型の構成の説明文のほうが、記述型の説明文よりリコールの成績が良かった。

Swaffar et al. (1991: 164) が指摘しているようにリコール・プロトコルの評価の問題がある。マクロの命題かミクロの命題かによって配点を変えるのか、またリコール・プロトコルを L1 で書くのか、L2 で書くのか、といった問題点がある。L1 でリコール・プロトコルを書くのは、多くの ESL 環境にはあわないし、一方 L2 で書く場合は、英語の読解と表現の両方を要求するので、読み手に負担がかかる。Lee (1986) は、L1 と L2 のリコールを比較した結果、L1 でリコールしたほうが L2 でリコールするよりテキストの復元の量が多かったことを報告している。

また、Lee (1986) はリコールをさせる時の指示の影響も調べた。実験の結果、スペイン語を外国語として学ぶ大学 1 年生のレベル 2 の学生では、リコール・テストを受ける時、英文を読んだあとにリコール・タスクが与えられることを前もって知らされた学生のほうが、リコールが後で要求されるという指示を受けなかった学生より多くリコールした。しかし、2 年生はその指示があろうと、なかろうとリコールの量は変らなかった。従って能力の低い学生にはリコールを後で課すということをテストの前に言うておくほうが、再生が容易になるであろう、と Lee は指摘している。

英語学力がリコールに及ぼす要因の一つであることも報告されている。Takahashi (1994) の研究では、日本人大学生は、重要部分のほうが詳細部分より正確に再生され、また学力の高い学生は低い学生より重要部分の再生の成績が良かった。

重要な情報 (主題) の再生率と発達レベルの関係を調べた研究に Lee and Ballman (1987) の FL 研究がある。スペイン語を学習する FL 大学生によるテキスト全体のリコール量は彼等の学期のレベルによって有意に差があったが、重要なアイデア・ユニットのリコール量と、それが重要なアイデア・ユニットかを識別する力は、学期レベルの影響を受けなかったことを明らかにしている。

リコール・プロトコルの分析の際、採点基準の違いが再生量の評価にどのような影響を及ぼすかは、興味深い問題点である。再生されたアイデア・ユニットを採点する時、完全に意味内容が再生されていれば正解とする厳密な採点基準 (strict criterion) と、情報が大体か一部再生されていれば正しいとみなすゆるやかな採点基準 (loose criterion) がある。情報がどの程度再生されていれば正解とすべきか、また厳密に採点しても、ゆるやかに採点しても読解力を測定する妥当性として差がないのかは、採点の際、悩むところであるが、これまで再生量の評価における採点基準の及ぼす影響を調べた研究は見当たらない。

本研究の目的は、異なった採点基準が、文章全体・主題全体・各パラグラフ全体・各パラグラフの主題についてのリコールの成績にどのような影響を及ぼすかを調べることである。ただし、リコール・テストで読解力を測定する時、採点基準の厳しさによって妥当性が異なるかどうかは、今回の課題として、本研究では考察しない。

本研究では具体的には、次の3点を明らかにする。

- (1)厳密な採点基準とややゆるやかな採点基準を使用した場合、採点基準の違いによって、リコール（文章全体と主題全体）の得点間に差があるか。
- (2)①文章全体のリコールの量、つまりトータル・リコール（total recall）と、②テキスト中の重要な情報、つまり主題（main idea）全体のリコールにおいて、厳密な採点基準による採点法とややゆるやかな採点基準による採点法の相関がどの程度か。
- (3)厳密な採点基準とややゆるやかな採点基準を使用した場合、採点基準の違いによって、各パラグラフ全体のリコール率と各主題のリコール率の順位（高低）のパターンに違いが生ずるか。

## 2. 方 法

### 2.1. 被験者

国立大学大学院生、計27人。リコールの量はテキストの親密さ（familiarity）の影響を受けるので、既知の内容かどうか調べたが彼等は全員リコール・テストの英文と同じ（様な）内容の英文を以前読んだことがないと報告した。また4ヶ月以上の英語圏滞在経験者は1名（5ヶ月）を除いていなかった。英語学習経験年数は平均約17年（最低10年～最高27年）。平均年齢約29歳（22歳～39歳）。なお27人中9人はproficiency testとしてのCELT Form A（listening, vocabulary, structure. 計300点満点）を受けており、平均点は207.8点（SD=26.45）で、得点範囲は174点～259点であった。

### 2.2. 実施時期——1993年9月。1994年6月。1995年6月。

2.3. 材料——リコール・テストは *Reading power* (Mickulecky and Jeffries, 1986:10) の英文 (Appendix 参照) をそのまま使用した。リコール・タスクは被検者には慣れていないので、語彙や統語が容易な243語の英文（4つのパラグラフから構成）を用意した。リコール・テストの際、はじめに英文を20分よく黙読して、英文に書いてあることを理解し、その後英文を見ないで、英文に書いてあったことをできるだけ多く、詳しく、そして正確に思い出して、解答用紙に日本語で15分間で書くように指示した。

### 2.4 採点法

#### 2.4.1 文章全体

Carrel (1985) に従って、筆者が説明文 (expository text) を66個のアイデア・ユニット (Appendix) に分けて採点した。また中学校英語教員1名にアイデア・ユニットの数の設定と被検者9人 (全体の30%) のリコール・プロトコルの採点を依頼した。inter-rater reliability は、リコール・プロトコルの採点では、strict criterion で.85であった。

#### 2.4.2 主題の採点

各パラグラフの主題は書き手の中心的見解を示し、最も重要な情報を伝えるものである。各パラグラフの主題を決定する時、11人の英語母語話者と6人の日本人英語教員に依頼したが、完全な一致をみるのは難しいことがわかったので、彼等の回答を参考にして、Mickulecky and Jeffries (1986: 264) をほぼ採用した。4つの主題文は合計10個のアイデア・ユニットからなる。主題文の満点は10点。各パラグラフの主題は以下のとおりである。

- (1) Main Idea 1 = Many people moved/out of American cities/after World War II.  
(3 idea units)
- (2) Main Idea 2 = American businesses (also) left the city/during the 1950s/and 60s.  
(3 idea units)
- (3) Main Idea 3 = Cities began/to have money problems.(or Cities began/to have many serious problems.) (2 idea units)
- (4) Main Idea 4 = Many people believed/that American cities were dying.  
(2 idea units)

#### 2.4.3 採点基準

厳密な採点基準では、もとの文章の内容と完全にあっていれば、1点とし、誤りは0点とした。ただし、もとの文章にないような別の文や節が付加されている場合はその情報は採点しないで除外した。Carrel (1983: 189) の定義では、ゆるやかな採点基準では、もとの文章の意味の大部分か一部が正しく再生（または言い直し）されている場合は正解とし、多少の意味のゆがみ(distortions)は許容する。本研究のゆるやかな採点基準では、もとの文章の一つのアイデア・ユニットの大部分（約3分の2の情報）が正しく再生されていると考えられる場合、正解として1点とした。それ以外は0点。どちらの採点基準も、言い直し(paraphrase)は許容して採点した。

例えば、American cities changed (この情報は一つのアイデア・ユニットからなり3語による下位情報から構成)の再生が、(1)アメリカが変わった、都市が変わった、(2)アメリカ人が変わった、(3)アメリカの多くの都市が変わった、の場合、(1)は情報の一部(cities, American)が削除、(2)ではcitiesが「人々」で置換、(3)は、「多くの」という情報が一つのアイデア・ユニット内に付加されている例である。厳密な採点基準では、これらの(1)(2)(3)のアイデア・ユニットは誤りとされ0点であるが、ゆるやかな採点基準では、アイデア・ユニットの情報の大体3文の2以上再生されているので、それぞれ正解とみなし、1点とした。一般に、再生する場合、原文をそのまま再生または言い直しする場合と、文やアイデア・ユニットを幾つか組合わせて再生したり、言い直したりする場合がある。例として、「都市に住む多くの人が都市の周辺の郊外へと移住した」はアイデア・ユニットのNo.5, 4, 8, 7, 6が正解でどちらの採点基準でも5点となり、「郊外に移った豊かなたいのいの人々や企業は都市のそのような問題を気にしなかった」は、アイデア・ユニットのNo.60, 61, 62, 63, 64, 65, 66の7個が正解となりいずれの採点基準でも7点とした。

Table 1. Means and SDs of all ideas and main idea units recalled

	n	Strict Criterion		Loose Criterion		Difference F
		M	SD	M	SD	
Total Recall <sup>a)</sup>	27	37.78	5.51	45.70	6.32	401.41**
All Main Ideas <sup>b)</sup>	27	5.85	0.97	7.56	1.26	54.15**

\*\* $p < .01$ <sup>a)</sup>Maximum possible score=66<sup>b)</sup>Maximum possible score=10

Table 2. Correlations of recall

	Loose Criterion	
	Total Recall	Total Main Ideas
Strict Criterion		
Total Recall	.95***	—
Total Main Ideas	—	.46*

\* $p < .05$     \*\*\* $p < .001$ 

### 3. 結 果

#### 3.1 トータル・リコールと主題のリコールの平均値・標準偏差・分散分析の結果

表1は厳密な採点法とゆるやかな採点法の2種類の異なった採点法による文章全体のトータル・リコール(66点満点)と4つの主題(計10個のアイデア・ユニットから構成, 10点満点)のリコールの得点の平均値, 標準偏差, 分散分析の結果を示す。

表1から, わかるように, 分散分析の結果, トータル・リコールでもすべての主題文の再生でも得点は, ゆるやかな採点法のほうが, 厳密採点法より有意に高かった(それぞれ  $F(1, 26) = 401.41, p < .01$ ;  $F(1, 26) = 54.15, p < .01$ )。

#### 3.2 相関係数

表2が示すように文章全体のトータル・リコールでは, 厳密な採点基準とゆるやかな採点基準の相関は, .95( $p < .001$ )で極めて高い相関係数であった。しかし主題文を構成するアイデア・ユニットの総数では, .46( $p < .05$ )の相関係数で,それほど高くはなかった。

#### 3.3 パラグラフごとのリコールの平均率・標準偏差・分散分析

各パラグラフ全体の平均再生率と標準偏差を示したものが表3である。

採点基準の違い(厳密な採点法とゆるやかな採点法)が各パラグラフ全体の再生率間の差に影響を及ぼすかどうかを調べるために, 2(採点基準)×2(パラグラフ)の分散分析を行った。その結果, 表4が示すように, 採点基準とパラグラフの主効果は有意であった(それぞれ  $F(1, 26) = 353.12, p < .01$ ;  $F(3, 78) = 34.99, p < 0.1$ )。また, 採点基準(A)とパラグラフ(B)の

Table 3. Mean percentage of the recall of each paragraph

	Strict Criterion	Loose Criterion
	Mean(SD)	Mean(SD)
Paragraph 1	78.1% (0.12)	87.5% (0.10)
Paragraph 2	63.4% (0.11)	72.7% (0.12)
Paragraph 3	54.2% (0.10)	76.4% (0.14)
Paragraph 4	41.2% (0.18)	50.9% (0.22)

Table 4. Results of ANOVA for paragraphs recalled

Source	SS	df	MS	F
Sub	1.3976	26	0.0538	
Criterion(A)	0.8727	1	0.8727	353.12**
SxA	0.0643	26	0.0025	
Paragraph(B)	3.6965	3	1.2322	34.99**
SxB	2.7464	78	0.0352	
AxB	0.1642	3	1.1547	21.64**
SxAxB	0.1973	78	0.0025	
Total	9.1390	215		

\*\* $p < .01$ 

Table 5. Mean percentage of the recall of each paragraph main idea

	Strict Criterion	Loose Criterion
	Mean(SD)	Mean(SD)
Main Idea 1	76.7% (0.22)	90.2% (0.15)
Main Idea 2	77.9% (0.26)	84.0% (0.23)
Main Idea 3	22.2% (0.25)	53.7% (0.33)
Main Idea 4	38.9% (0.31)	63.0% (0.40)

交互作用が有意であった ( $F(3,78) = 21.64$ ,  $p < .01$ )。そこで、各水準ごとに単純効果を分析した結果、どのパラグラフにおいても、ゆるやかな採点基準によるほうが、厳密な採点基準よりリコール率が有意に高かった (パラグラフ1  $F(1,26) = 51.05$ ,  $p < .01$ ; パラグラフ2  $F(1,26) = 73.53$ ,  $p < .01$ ; パラグラフ3  $F(1,26) = 162.29$ ,  $p < .01$ ; パラグラフ4  $F(3,78) = 42.09$ ,  $p < .01$ )。また LSD 法による多重比較の結果、厳密な採点基準の場合、パラグラフ1 > パラグラフ2 > パラグラフ3 > パラグラフ4 の順の再生率となり全ての対間に有意差がみられたが ( $MSe = .02$ ,  $p < .05$ )、ゆるやかな採点基準の場合、パラグラフ1 > パラグラフ3 = パラグラフ2 > パラグラフ4 の順で、パラグラフ2 とパラグラフ3 の間には有意差は認められなかった ( $MSe = .02$ ,  $p < .05$ )。

Table 6. Results of ANOVA for paragraph main ideas recalled

Source	SS	df	MS	F
Sub	2.0215	26	0.0777	
Criterion (A)	1.9097	1	1.9097	58.55**
SxA	0.8481	26	0.0326	
Main Idea (B)	8.1659	3	2.7220	18.13**
SxB	11.7077	78	0.1501	
AxB	0.5091	3	0.1697	6.26**
SxAxB	2.1143	78	0.0271	
Total	27.2763	215		

\*\* $p < .01$ 

### 3.4 パラグラフの主題ごとのリコールの平均率・標準偏差・分散分析の結果

各パラグラフの主題の平均再生率と標準偏差は、表5の示すとおりである。

採点基準の違いが、各パラグラフの主題の復元率間の差に影響をおよぼすかどうかを検討するため、2（採点基準）×2（主題）の分散分析を行った結果が、表6に示されている。すなわち、採点基準と主題の主効果は有意であった（それぞれ  $F(1,26)=58.55, p<.01$ ;  $F(3,78)=18.13, p<.01$ ）。また採点基準と主題の交互作用が有意であった（ $F(3,78)=6.26, p<.01$ ）。各水準ごとに単純効果を分析した結果、いずれのパラグラフの主題においても、ゆるやかな採点基準の再生率のほうが、厳密な採点基準の再生率より、有意に高かった（主題1  $F(1,26)=17.87, p<.01$ ; 主題2  $F(1,26)=5.91, p<.05$ ; 主題3  $F(1,26)=44.20, p<.01$ ; 主題4  $F(1,26)=12.77, p<.01$ ）。またLSD法による多重比較の結果、厳密採点基準では、リコール率が主題2＝主題1＞主題4＞主題3の順であったが（ $MSe=.08, p<.05$ ）、ゆるやかな採点基準では、主題1＝主題2＞主題4＝主題3の順であった（ $MSe=.09, <.05$ ）。

## 4. 考 察

(1) やや採点基準をゆるくすると、テキスト全体のリコールも、すべての主題のリコールも、許容される情報が増えて、両方の得点が厳密採点法より有意に高くなった。

(2) トータル・リコールでは厳密採点と少しゆるやかな採点基準による得点の相関が.95で、きわめて高かった。このことから、トータル・リコールではどちらの採点基準を使っても、得点の相対的順位にほとんど変化がないといえるであろう。従って、文章全体のリコールの採点の時は、完全にもとの文章の内容を復元している情報だけを正解とする厳密な採点基準がよいのか、大体の情報を再生していれば、正解とするのか、どちらの基準を採用するかは、神経質にならなくてもよいであろう。ゆるやかな採点基準では、もとの文章内容の忠実な再生でなくても、もとの文章のアイデア・ユニットの情報を一部削除したり、本文にない別の情報で置換、追加、強調したりすることを許容して大体の情報が再生されていれば可とすることは、より多様な再

生を認め被験者の能力を顕在化させることができる長所がある。しかし、実際採点する場合、原文内容の忠実な復元を要求する厳密な採点基準よりも、かなりの労力と時間がかかり、採点の信頼性を高めるためにいろいろな解答例を記録して、採点の一貫性を保とうとする配慮がより必要になる。この事を考えると、どちらの採点法が、より読解力や総合的英語学力を測れるかという妥当性にこだわったりする場合や、より柔軟な原文内容の復元を認めたい場合を除いては、文章全体のリコールの採点の時は、労力と時間を節約できる厳密な採点基準のほうが実用性が高いと言えるであろう。ただし、テキスト内の重要な情報量（主題文を構成するアイデア・ユニットの総数）については、厳密な採点法とややゆるやかな採点法では、トータル・リコールの採点の場合ほど相関は高くなかった。

(3)各パラグラフ全体のリコール、各パラグラフの主題のリコールの比較では、採点基準×パラグラフ、および採点基準×主題の交互作用が有意であった。つまり採点基準の違いによって各パラグラフ全体のリコール率および各パラグラフの主題のリコール率の順位のパターンが異なることが明らかになった。厳密な採点法では、各パラグラフの再生率の比較をする時、隣接するパターン間すべての対の間に有意差が認められたが、ややゆるやかな採点法では、厳密な再生でなくても、多少の内容の情報を省略、追加、置換しても点がとれるので、甘くなるせいか、第2と第3パラグラフの間の差は有意でなかった。またパラグラフごとの主題文のリコール率の比較でも、厳密な採点基準のほうが、ゆるやかな採点基準より、各主題の再生率間に有意差が認められやすかった。このことから、厳密な採点基準にすると、有意差がでる傾向があると言えよう。再生率の順位のパターンの差異を生じさせた原因としては、第3パラグラフの特性が考えられるが、詳細な検討は今後の研究課題として残される。

以上をまとめると、大学院生の読解において、テキスト全体のリコールの採点の場合を除いて、①主題の総数のリコールの採点、②各パラグラフの主題のリコール率の比較、③各パラグラフ全体のリコール率の比較を行う時は、採点基準を厳密にする場合とやや甘くする場合とでは、①における相関は.4台であり、再生率の順序のパターンに相違が認められるので、どちらの採点基準を使用するか配慮する必要がある。

ただしリコール・テストが総合的英語学力や読解力を示すのにどちらの採点基準が妥当性が高いかについては、今後検討していかなければならない。

(4)パラグラフによって再生率が有意に異なることがわかった。いずれの採点基準を使っても、最初に出てくる第1パラグラフのリコール率のほうが他のパラグラフより有意に高く、最後の第4パラグラフのリコール率が他より有意に低かった。このことから第1パラグラフがより理解され、最後のパラグラフがより理解されなかったということが考えられよう。これはパラグラフの特性や位置が影響していると推測される。第1パラグラフでは、他のパラグラフより語数が最も少なく50語からなりたっている。また統語的複雑度(T-unitの平均の長さ=8.3)も第4パラグラフ(T-unitの平均の長さ=7.5)を除けば、他のパラグラフ(第2パラグラフでは、9.8、第3パラグラフでは、10)より低く、内容の理解も容易である。また開始部なので、再生されやすかったかもしれない。一方第4パラグラフは最も語数が多く長い。内容も読者の背景知識を必要とするからであろう。また、各パラグラフの主題も前半の主題のほうが後半のそれよりも再生率が有意に高かった。テキスト情報の想起は、テキストの型や他の要因等も影響しており、今後検討していく必要がある。



## 5. お わ り に

本研究では、読解力を測定するリコールテストの採点基準の違いが、大学院生のリコール・プロトコルの成績に影響を及ぼすかどうかを検討した。その結果、次の主な点が明らかになった。

- (1) 文章全体のリコールの得点においては厳密な採点法とやや厳密な採点法の相関は.95で非常に高いことから、文章全体のリコールの採点ではどちらの採点基準を使用するかは、それほど神経質にならなくてよいであろう。
- (2) ただし、テキスト内の重要な情報（主題の総数）のリコールの得点では二つの採点法の相関は.4台であり高くなかった。また、採点基準の違いによって、各パラグラフのリコール率と各パラグラフの主題のリコール率の順位のパターンが異なったこと（例えば、厳密な採点法の場合のほうが各パラグラフのすべての対間にリコール率の有意差が認められたこと）から、各パラグラフおよび各主題間でリコール率の比較をする時は、どちらの採点基準を使用するかは考慮する必要があるだろう。
- (3) どちらの採点基準を使用しても第1パラグラフのリコール率が一番高く、最後の第4パラグラフが最も低かった。

本研究ではゆるやかな採点基準でも、やや厳密な方法を使用した方が、アイデア・ユニットの半分以上の情報を伝えている場合は正解とする場合や、採点が1点か0点ではなく、情報に重みづけを行い1点から3点というように採点に幅をもたせた場合も今後検討する必要がある。

本研究の被験者は英語学習経験年数が10年以上ある大学院生でかなり英語学習経験年数に幅がある。従って英語学習年数が類似しているような学部生を被験者として、英語学力および他の形式の読解力テストとリコールの相関、他の論説文および論説文以外の他のテキストの型のリコールの分析も今後の検討課題である。

## 参 考 文 献

- Bernhardt, E. 1983. Three approaches to reading comprehension in immediate German. *Modern Language Journal*, 67, 111-115.
- Bernhardt, E. 1991. *Reading development in a second language*. Norwood: Ablex Publishing.
- Bransford, J. 1979. *Human cognition: Learning, understanding, and remembering*. Belmont, CA:Wadsworth.
- Brown, A., and Smiley, S. 1977. Rating the importance of structural units of prose passages: a problem of metacognitive development. *Child Development*, 48, 1-8.
- Carrel, P. 1983. Three components of background knowledge in reading comprehension. *Language Learning*, 33, 183-207.
- Carrel, P. 1984. The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, 18, 441-469.

- Carrel, P. 1985. Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19, 727-752.
- Johnson, R. 1970. Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 9, 12-20.
- Lee, J. 1986. On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 83-93.
- Lee, J. and Ballman, T. 1987. FL learners' ability to recall and rate the important ideas of an expository text. In B. VanPattern et al.(Eds.), *Foreign language learning* (pp.108-118). Rowley, MA: Newbury House.
- Mikulecky, B. and Jeffries, L. 1986. *Reading power*. Reading, Mass.: Addison-Wesley Publishing Company.
- Swaffar, J. et al. 1991. *Reading for meaning: An integrated approach to language learning*. Englewood: Prentice Hall.
- Takahashi, T. 1994. The role of lexical and syntactic knowledge in understanding a well-organized prose passage. *ARELE* 5, 121-130.
- Wolf, D. 1993. A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal*, 77, 473-489.

#### Appendix: Idea Unit Analysis

##### (Paragraph 1)

1. American cities changed
2. in many ways
3. after World War II.
4. Many people moved
5. out of the city.
6. They moved
7. to the suburbs,
8. the areas around a city.
9. Most of the people...were rich.
10. who moved [9 and 10=Most of the people who moved were rich.]
11. Poor people usually did not have enough money
12. to move.
13. They stayed
14. in the cities.

##### (Paragraph 2)

15. During the 1950s
16. and 60s
17. there was another important change
18. in American cities.
19. Businesses began..., too.

- 20. to leave the city
- 21. They left
- 22. because the people...were poorer.
- 23. in the city
- 24. Poor people did not have much money
- 25. to buy things.
- 26. So, many shops
- 27. and restaurants moved
- 28. to the suburbs.
- 29. People...had more money
- 30. in the suburbs
- 31. to spend.[29, 30 and 31=People in the suburbs had more money to spend.]

(Paragraph 3)

- 32. Cities began
- 33. to have many serious problems.
- 34. The rich people
- 35. and the businesses did not pay city taxes anymore.
- 36. The poor people could not pay much money
- 37. in taxes.
- 38. So cities had less money
- 39. for schools
- 40. and housing.
- 41. Sometimes they could not pay their police officers
- 42. or firefighters.
- 43. And they could not take good care of their streets
- 44. and parks.

(Paragraph 4)

- 45. But money was only part
- 46. of the problem.
- 47. Many people believed
- 48. that American cities were dying.
- 49. They had good reason
- 50. to believe this.
- 51. City streets were sadly empty.
- 52. Many neighborhoods...were dirty
- 53. and parks
- 54. and dangerous.[52, 53, 54=Many neighborhoods and parks were dirty and dangerous.]
- 55. In some places
- 56. buildings were even falling down.
- 57. And nobody seemed
- 58. to care.

- 59. This was the real problem.
- 60. Most of the people
- 61. and businesses...were
- 62. with money
- 63. in the suburbs.[60, 61, 62, and 63=Most of the people and businesses with money  
were in the suburbs.]
- 64. They did not care
- 65. what happened
- 66. to the cities.

## The Effects of Scoring Criteria on Recall Protocols: Focusing on Japanese EFL Graduate Students

Kinue HIRANO

### ABSTRACT

The recall procedure is a frequently used measure of reading comprehension. This study investigates the effects of two different scoring criteria on the written recall of an expository text. Using both a strict and a loose criterion of recall, the written recalls of twenty-seven Japanese graduate students of English were scored for (1) the total number of idea units recalled correctly, (2) the total number of main ideas recalled correctly, (3) the percentage of correct recall of each paragraph, and (4) the percentage of each paragraph main idea recalled correctly. The strict criterion disallowed distortions and additions, though it allowed paraphrases. The loose criterion allowed some distortions of the original meaning, i. e, cases in which most of the meaning was correctly recalled. The results revealed that the correlation between the strict and the loose criterion was extremely high for total recall of the expository text ( $r = .95$ ,  $p < .001$ ), while for total main ideas, the correlation was not so high ( $r = .46$ ,  $p < .01$ ). In addition, the criterion of scoring recall protocols applied proved to be a significant factor affecting the recall of each paragraph and of each main idea. That is, there were significant differences between the strict and the loose criterion in terms of the percentage of correct recall across paragraphs and paragraph main ideas. For paragraph recall, there were significant differences between all four adjacent paragraphs recalled in the strict criterion: Paragraph 1 > Paragraph 2 > Paragraph 3 > Paragraph 4, but in the loose criterion, significant differences were not found for all paragraphs: Paragraph 1 > Paragraph 2 = Paragraph 3 > Paragraph 4. For the recall of each main idea, the pattern that emerged was different for each criterion: for the strict criterion, Main Idea 1 = Main Idea 2 > Main Idea 4 > Main Idea 3; for the loose criterion, Main Idea 1 = Main Idea 2 > Main Idea 3 = Main Idea 4.