

採点基準の違いが読解のリコールに及ぼす影響

——高校生（上位者）の場合——

平野 絹枝*

(平成9年11月14日受理)

要 旨

読解力を測定するテストの一つにリコール (recall) テストがある。読解におけるリコールには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、高校生(特に上位校に在籍する高校生)を対象に、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(厳密な採点基準とやや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検証した。その結果、(1)2種類の採点基準の相関は極めて高かった。すなわちテキスト全体のリコールでは、2種類の採点基準の相関係数は .96で、重要な情報量(主題文全体)のリコールでは、.81であった。また(2)各パラグラフ全体及び各パラグラフの主題の再生率の順位(高低)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

読解力を測定するテストの一つにリコールテスト(再生テスト)がある。これは学習者にテキストを読ませた後、テキストの内容を想起、再生させるテストである。リコール・タスクはテスト項目によって読解が干渉を受けないことがないので、Bernhardt(1991: 200)は、理解を測定するより純粋な方法であると主張している。Bernhardt(1983: 28)は、テキストと読み手のコミュニケーションに焦点をあてる、外国語読解力のテスト法として、文字によるリコール・タスクの使用を奨励している。リコール・プロトコル (recall protocol)は、読み手がどのように情報を再構築するかを示し (Bernhardt 1983: 31)、どの程度学習者がテキストを理解しているかを明らかにしてくれる (Swaffner et al. 1991: 164)。従って読解におけるリコール・タスクでは再生が正しく多ければ、その学習者の読解力が高いということになる。Barnett(1989: 158)はリコール・プロトコルによって学生がどこで解釈を誤っているか、彼らのスキーマがテキスト

* 言語系教育講座

とどのような相互作用をしているかを知ることが出来る、と述べている。さらに Barnett (1989: 161) は自分で再構成したものとクラスで論じた標準的な再構成を学生に比較させることは読解の育成に役にたつとして、異なったりリコールが読み手の理解について何を明らかにしてくれるかを学生と議論することを提案している。

第二言語学習者のリコールに関する実験研究では、リコールに影響を及ぼす要因として、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 Lee (1986) や平野 (1996) で指摘されている。

この他採点基準の違いもリコールの得点に影響を及ぼす要因の一つであることが平野 (1996) で報告されている。平野は大学院生27名を対象にして、採点基準の厳しさの違い(厳密な採点基準とそれよりややゆるやかな採点基準)がリコールの成績にどのような影響をおよぼすかどうかについて検証した。その結果、テキスト全体のリコールでは、2種類の採点基準の相関係数は .95 で極めて高かったことから、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ことを明らかにした。さらに、テキスト内の重要な情報量(各パラグラフの主題文の総計)のリコールでは、相関は .46 であまり高くなかったこと、および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を示した。従って、テキスト全体のリコール量を採点する場合は、採点基準を厳しくすべきかどうかはあまり問題にならないが、テキスト内の重要な情報(主題の総数)のリコールの採点をする時、及び各パラグラフと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、と指摘している。

リコール・プロトコルを分析する際、再生量の評価に及ぼす採点基準の影響は、学習者の英語学習経験年数や年齢の異なる中学生、高校生、大学生、大学院生によって違いがみられるのかは、読解力の評価の観点から大変興味深い。平野 (1996) の研究は日本人の大学院生を対象にしており、日本人 EFL 高校生を対象にして、リコール量の測定に異なった採点基準がどのような影響を及ぼすかを調べた研究は見当たらない。

従って、本研究の目的は、学習者が高校生(ただし英語学力上位校に在籍する高校生に限定)の場合、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、英語学力の低い高校生における採点基準の違いによる影響は、今回の課題として本研究では考察しない。

本研究では、学力上位校に在籍する高校生を対象にして、具体的には、次の3点を明らかにする。

- (1) 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、高校生の場合どの程度か。
- (2) 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、高校生の文章全体のリコール及び主題全体のリコールの得点に差があるか。
- (3) 高校生において、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、各パラグラフ全体及び各パラグラフの主題のリコール率の順位(高低)の

ターンに違いが生ずるか。

2. 方 法

2.1 被験者

都立高校2年生2クラス。都立高校のなかでトップ校の一つに在籍する高校生である。本研究の目的が学力上位の高校生に限定して検証することなので、2次募集で入学した高校生は分析から除いた。さらに日本人EFL学習者を対象とするので、半年以上の英語圏滞在経験者も除いた。その結果被験者は53人となった。テスト後のアンケートで被験者全員リコール・タスクはこれまで一度も受けたことがなく、またリコール・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。平均年齢約16歳(15~16歳)。

2.2 実施時期

1995年6月

2.3 材料

大学院生と高校生を比較したいので、大学院生を対象にした平野(1996)と同様、リコール・テストは *Reading Power* (Mikulecky and Jeffries, 1986:10)の英文(Appendix 参照)を使用した。リコール・タスクは、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。10年以上の教師経験のある中学校英語教員1名にアイデア・ユニットの数の設定と被験者13人(全体の約25%)のリコール・プロトコルの採点を依頼した。inter-rater reliabilityは、idea unitの設定では1.0、リコール・プロトコルの採点では、strict criterionで、.97 ($p < .01$)、loose criterionでは、.95 ($p < .01$)であった。

2.4.2 採点基準

採点基準は平野(1996)で詳述したので、本稿では簡単に言及する。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。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点を与え、それ以外は0点とした。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3語による下位情報から構成された一つのidea unit)の情報の一部が削除、置換、付加された場合厳密な採点基準では、誤りとなり0点であるが、ゆるやかな採点基準では、アイデア・

Table 1 Differences in point between strict criteria and loose criteria

		strict criterion (point)	loose criterion (point)
examples			
reduction	アメリカが変わった, 都市が変わった	0	1
replacement	アメリカ人が変わった,	0	1
addition	アメリカのいくつかの都市が変わった。	0	1

ユニットの情報の大体3文の2以上再生されている場合は、それぞれ正解とみなし1点とした (Table 1参照)。

2.4.3 主題

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。どのように各パラグラフの主題を決定したかは平野(1996)に述べてあるので省略する。

4つのパラグラフの各主題文は4つあり、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. (2 idea units)
- (4) Main Idea 4=Many people believed/that American cities were dying.
(2 idea units)

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

表2では、厳密な採点法とやや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リコール、すまわちトータル・リコール (total recall, 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, 分散分析の結果が示されている。

表2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、当然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 $F(1, 52) = 415.98, p < .01$; $F(1, 26) = 85.08, p < .01$)。

3.2 相関係数

表3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 .96 ($p < .01$) で極めて高い相関係数であった。一方主題文を構成するアイデア・ユニットの総数でも .81 ($p < .01$) という高い相関係数を示した。

Table 2 Means and SDs of all ideas and main idea units recalled

	n	Strict Criterion		Loose Criterion		Difference
		M	SD	M	SD	F
Total Recall ^{a)}	53	35.00	8.11	41.66	8.74	415.98**
All Main ideas ^{b)}	53	6.02	1.76	7.43	1.84	85.08**

**p<.01

a) Maximum possible score=66

b) Maximum possible score=10

Table 3 Correlations of recall

	Loose Criterion	
	Total Recall	Total Main Ideas
Strict Criterion		
Total Recall	.96**	—
Total Main Ideas	—	.81**

**p<.01

Table 4 Mean percentage of the recall of each paragraph

Paragraph	Strict Criterion	Loose Criterion
	Mean (SD)	Mean (SD)
Paragraph 1	76.3%(12.24)	83.8%(11.03)
Paragraph 2	58.5%(18.19)	68.1%(18.96)
Paragraph 3	53.0%(18.34)	70.0%(19.21)
Paragraph 4	34.0%(18.56)	42.0%(21.09)

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は表4の示すとおりである。

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5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 $F(1, 52) = 431.09$, $p < .01$; $F(3, 156) = 80.47$, $p < .01$ ）。また、採点基準(A)とパラグラフ(B)の交互作用が有意であった ($F(3, 156) = 27.28$, $p < .01$)。そこで、各水準ごとに単純主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 $F(1, 52) = 114.72$, $p < .01$; パラグラフ2 $F(1, 52) = 160.78$, $p < .01$; パラグラフ3 $F(1, 52) = 195.49$, $p < .01$; パラグラフ4 $F(1, 52) = 106.31$, $p < .01$ ）。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 ($MSe = 195.704$, $p < .05$)、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

Table 5 Results of ANOVA for paragraphs recalled

Source	SS	df	MS	F
Sub	64721.87	52	1244.65	
Criterion(A)	11777.19	1	11777.19	431.09**
S X A	1420.63	52	27.32	
Paragraph(B)	94942.02	3	31647.34	80.47**
S X B	61353.60	156	393.29	
A X B	1532.09	3	510.70	27.28**
S X A X B	2920.42	156	18.72	
Total	238667.81	423		

** p<.01

Table 6 Mean percentage of the recall of each paragraph main idea

	Strict Criterion	Loose Criterion
	Mean(SD)	Mean(SD)
Main Idea 1	72.21%(23.91)	83.0%(23.90)
Main Idea 2	71.07%(32.40)	78.6%(29.74)
Main Idea 3	51.89%(34.97)	66.0%(34.71)
Main Idea 4	31.13%(36.63)	64.2%(45.95)

Table 7 Results of ANOVA for paragraph main ideas recalled

Source	SS	df	MS	F
Sub	131902.88	52	2536.59	
Criterion(A)	26731.28	1	26731.28	80.37**
S X A	17295.33	52	332.60	
Main Idea(B)	65727.64	3	21909.21	12.26**
S X B	278724.36	156	1786.69	
A X B	11030.31	3	3676.77	12.23**
S X A X B	46885.62	156	300.55	
Total	27.2763	215		

** p<.01

られなかった (MSe = 216.31)

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

各パラグラフの主題の平均再生率と標準偏差は、表6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7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 $F(1, 52) = 80.37, p < .01$; $F(3, 156) = 12.26, p < .01$)。

26, $p < .01$)。また採点基準と主題の交互作用が有意であった ($F(3, 156) = 12.23$, $p < .01$)。各水準ごとに単純主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 $F(1, 52) = 18.67$, $p < .01$; 主題2 $F(1, 52) = 15.22$, $p < .05$; 主題3 $F(1, 52) = 20.53$, $p < .01$; 主題4 $F(1, 52) = 37.69$, $p < .01$)。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1 = 主題2 > 主題3 > 主題4の順であったが($MSe = 976.752$, $p < .05$)、ゆるやかな採点基準では、主題1は主題3・主題4よりも有意に高く、また主題2は主題4よりも有意に高かった。ただし、主題1と主題2の差は有意でなく、主題2と主題3、及び主題3と主題4の差も有意でなかった ($MSe = 1110.49$, $p < .05$)。

4. 考 察

4.1 2つの採点基準の相関

4.1.1 トータル・リコール

高校生のトータル・リコールでは厳密採点とややゆるやかな採点規準による得点の相関が.96で、非常に高かった。日本人EFL大学院生を対象にした平野(1996)でも、同様にきわめて高い相関係数 ($r = .95$)が検出されている。このことから、被験者が大学院生でも、高校生(ただし学力の高い高校生の場合)でもトータル・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、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のではないだろうか。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を問題にする場合や、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では、厳密な採点基準のほうが採点の労力と時間を節約できるので実用性が高いと言えるであろう(平野1996: 462)。

4.1.2 主題全体(主題の総数)のリコール

また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)においては、平野(1996)における大学院生の場合、厳密な採点法とややゆるやかな採点法の相関係数($r = .4$ 台)はあまり高くなかったが、本稿の高校生の場合、 $r = .81$ でかなり高い相関係数が検出された。リコールに使用する英文の難易度が学習者のレベルに適切かどうか、また英語学習経験年数や年齢が異なると重要な情報(主題)のリコールの採点の場合に採点基準の厳しさの程度が有意な影響を及ぼすのかどうかについては、今後さらに被験者を増やし別の英文を用意して継続研究を行う必要がある。

4.2 採点基準の違いの影響

4.2.1 テキスト全体及び主題全体のリコール

平野(1996)の大学院生と同様に、ややゆるやかな採点基準では、テキスト全体のリコールも、すべての主題のリコールも、両方の得点が厳密採点法より有意に高くなった。採点基準をややゆるくすれば許容される情報が増えるので、当然の結果としてうなずける。

4.2.2 パラグラフ及びパラグラフの主題のリコールの比較

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、および採点規準 × 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、隣接するパラグラフ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、多少厳密な再生でなくても点がとれ甘くなるせいか、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。

大学院生（平野 1996）でも、本研究での英語力のある高校生の読解の場合でも、①各パラグラフ全体のリコール率の比較、および ②各パラグラフの主題のリコール率の比較を行う際は、採点基準を厳密にする場合とやや甘くする場合とでは、再生率の順序のパターンに違いがあるので、どちらの採点基準を使用するか配慮する必要があるだろう。なおリコール・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において妥当性が高いかについては、今後の検討課題である（平野 1996）。

テキスト情報の想起は、読解力、学力、テキストの型、背景的知識、情報の興味、重要性の度合い等の要因も関係していて、諸要因の影響を今後検討していかなければならない。読解力を測定する上で、パラグラフの位置も関係しているかどうか、パラグラフの再生率を検討するのは興味深い。本研究では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。大学院生の場合（平野 1996）と同様に、いずれの採点基準を使っても、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。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（平野 1996）。テキストの後半に係するリコールが低いのは、読解・再生の疲労度とも関係があるかもしれない。

5. おわりに

本研究では、比較的英語学力の高い高校生の場合、読解力を測定するリコールテストの採点基準の違いが、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ゆるやかな採点法の相関は.96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) 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も二つの採点法の相関は.81でかなり高く、(1)と同様のことが言えよう。
- (3) 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際は、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。

(4) 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。なお、(1)(3)(4)は平野 (1996)を支持する結果となった。

本研究では英語学習経験年数が類似している、英語学力のかなり高い高校生を被験者としている。英語学力の低い高校生を対象にした場合、類似の結果がみられるかどうかは、他のテキストの型の英文のリコールの分析も加えて、今後検討していく必要がある。

参 考 文 献

- Barnett, M. 1989. *More than meets the eye*. Englewood Cliffs, New Jersey: Prentice Hall Regents.
- Bernhardt, E. 1983. Syntactic and lexical/semantic skill in foreign language reading comprehension: the immediate recall protocol. *Die Unterrichtspraxis*, 16, 27-33.
- 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.
- 平野 綱枝. 1996. 「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——」『上越教育大学研究紀要』第15巻, 455-467.
- 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.
- Riley, G. L. and Lee, J. F. A comparison of recall and summary protocols as measures of second language reading comprehension. 1996. *Language Testing*, 13, 173-189.
- 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 Effect of Scoring Criteria on Recall Protocols: Focusing on Japanese EFL High School Students with High Proficiency

Kinue HIRANO*

ABSTRACT

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 53 Japanese EFL high school students (focusing on students with high proficiency) were scored. The strict criterion disallowed distortions and additions, though it allowed paraphrases. The loose criterion allowed some distortions for 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 the total number of idea units recalled correctly ($r = .96, p < .01$). This finding for Japanese high school students supported the very high correlation found with Japanese graduate students in Hirano (1996). For the total number of main ideas recalled correctly, the correlation was also high ($r = .81, p < .01$). In addition, the criterion of scoring recall protocols applied proved to be a significant factor which affected the comparisons of the recall of each paragraph and of each paragraph main idea. This finding reaffirmed the results of Hirano (1996). That is, significant differences were found 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. In the loose criterion, however, there were no significant differences for all paragraphs: Paragraph 1 > Paragraph 2 = Paragraph 3 > Paragraph 4. For the recall of each paragraph main idea, the pattern that emerged was different for each criterion: for the strict criterion, Main Idea 1 = Main Idea 2 > Main Idea 3 > Main Idea 4; for the loose criterion, Main Idea 1 > Main Idea 3 = Main Idea 4, Main Idea 2 > Main Idea 4. (Main Idea 1 = Main Idea 2, Main Idea 2 = Main Idea 3)

* Division of Languages: Department of Foreign Languages