

スピーキング技能の系統的伸長のための中高一貫シラバスの開発(2)

—評価規準に照らした形成的評価と指導のサイクル—

深澤 清治 松浦 伸和 池岡 慎 大野 誠
千菊 基司 多賀 徹哉 高森 理絵 山岡 大基
山田佳代子 幸 建志

1. はじめに

2005年度広島大学学部・附属学校共同研究(三浦ほか2006)において、附属福山中・高等学校英語科では、スピーキング・テストにおける生徒のパフォーマンス・データを分析することにより、生徒が目標とすべきスピーキング技能の細目および評価の規準を示すルーブリックの開発を行った。これは、最終的には、生徒のスピーキング技能を、中学校から高校への一貫した指導の中で系統的に伸長させるためのシラバスを開発することを目的とするものであった。同研究において得られた成果としては、生徒のパフォーマンスを英語科の全教員が評価することにより、生徒に育成すべきスピーキング技能について教員間で共通認識が得られたことが挙げられる。また、スピーキング技能のうち「対話をする」という言語活動に関するルーブリックが開発され、以後の指導および評価の規準が設定されたことも大きな成果であった。

2. 課題の設定と研究の目的

しかし、同研究において開発されたルーブリックは、質・量ともに限定的なデータに基づくものであるため、未だ暫定的な性格を持つものである。したがって、同ルーブリックを実際の指導・評価において運用する中で、より妥当なものへと改訂していく余地がある。また、同ルーブリックは、「オーラル・コミュニケーションI」における「対話をする」という内容のまとまりに関わるものとして開発されたが、他の内容のまとまりに関わるルーブリックを開発するうえでの参考資料ともなるものである。今後、スピーキング技能に関する、より包括的なシラバスを構築していくにあたっては、他の内容のまとまりにおける指導と評価のあり方も重要な検討課題となってくる。このことを考慮に入

れるならば、同ルーブリックを他の内容のまとまりの指導・評価へと応用する方途を探ることも必要である。

そこで2006年度においては、次の2点を目的としてスピーキング技能の指導および評価を行った。

- (1) 同ルーブリックを実際の指導・評価において運用することで評価と指導のサイクルを試行的に構築し、そこでの成果と問題点を明らかにしたうえで、より適切な指導と評価のあり方について示唆を得ること。
- (2) 同ルーブリックを他の内容のまとまりに関わる言語活動の指導および評価に試行的に適用することにより、同ルーブリックが適用できる要素と、当該の内容のまとまりに固有の要素を明らかにし、より包括的なシラバスを開発するための示唆を得ること。

3. 研究の概要

2005年度の研究においてルーブリック開発の際に分析の対象となったのは高校1年生によるパフォーマンスであった。このことを踏まえ、本研究においても高校1年生のパフォーマンスを分析の対象とした。本研究の最終目的が中学校から高校へと一貫した体系的なシラバスの開発であることを考慮すると、他学年のデータも併せて対象としていくのが理想的ではあるが、まずは1つの学年における指導と評価のサイクルを確立させることが、後に他学年を含めた系統的なシラバスの開発にとって有益であると判断し、高校1年生における取り組みを優先させることとした。

本研究における指導と評価は、科目「オーラル・コミュニケーションI」(2単位)において行われた。対象生徒数は5クラス214名である。

3. 1. 生徒の現状把握

1学期(4月~7月)の授業におけるスピーキング技能に関わる指導としては、教員(JTE, ALT)と交わすごく平易な日常会話、および読み物教材の内容に関する教員からの質問に回答するという活動があった。しかし、指導の重点がリスニング技能にあったため、スピーキング技能に関しては、体系的な指導はなされていなかった。

そこで、まず生徒のスピーキング技能の現状を把握すべく、2006年7月に、次の2種類のテストを実施した。

- ・テスト1: モノログによるストーリー描写
与えられた4コマ漫画に描かれたストーリーを英語で説明する。発言は全てコンピュータで録音する。事前の準備はなく、発言の途中で言い直しをしたり沈黙が続いたりしても、そのまま録音を続ける。
- ・テスト2: インタビュー
「テレビゲーム」を話題として、教員からの質問に回答する。教員からの質問は原則として固定されたものであるが、生徒の回答の内容に応じて、会話の流れを調整する発言は適宜行われる。また、テストの様子はビデオ録画される。

テスト1・2はそれぞれ、内容のまとまりで言う「話すこと」および「対話すること」に関わるものである。

また、テストの内容は、生徒のパフォーマンスを評価することが容易になるように構造化されている。テスト1では、4コマ漫画の各コマに複数の情報が描写されており、生徒が言及する情報の量をもって、習熟度を測定する基準とすることができる。テスト2では、教員からの質問は、回答の自由度が低く定型的な表現で処理できるものから、自由度が高く創造的に回答しなければならないものまで質問に変化をもたせ、生徒がどの種類の質問にどのようなパフォーマンスを見せるかを分析することができるようになっている。

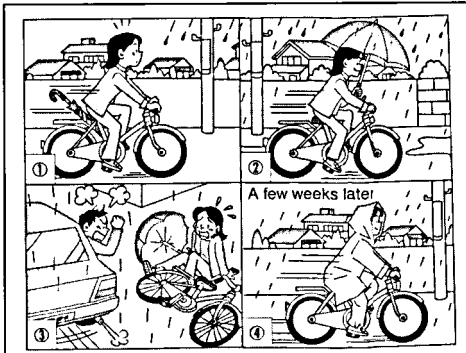


図1 7月テスト1

7月テスト2

Q: Video games are very popular these days.
Do you like video games? -①

→ A: Yes.

→ Q: Why? -②

What kind of game do you like? -③
Some scientists say video games are bad for your brain. What do you think about that? -④

→ A: No.

→ Q: Why not? -②

What do you do when you're free? -③
Some scientists say video games are bad for your brain. What do you think about that? -④

①Yes-No: 回答内容の自由度はきわめて低い。定型的な表現で回答できる。

②Wh-: 回答内容の自由度は、潜在的には高いが、ある程度定型的な表現で回答することができる。また、回答内容が自分自身に関する事なので、idea generationは比較的容易であると想定される。

③Wh-: 回答内容の自由度は比較的低い。定型表現ではないが、ジャンルを表す語彙や平易な表現だけで回答することができる。

④Wh-: 回答内容の自由度は高く、定型表現では回答できない。また、回答内容も、インタビュアーに問われて初めて考えなければならないであろうと想定されるので、idea generationから始めなければならない。

3. 2. パフォーマンス・データの評価(事前)

前項に示したテストにおける生徒のパフォーマンス・データについて、英語科の全教員が評価を行った。評価に当たっては、以下の観点について、それぞれ1~4点の4段階で得点化することとした。

- テスト1: 「表現の能力」, 「知識・理解」, 「情報量」
- テスト2: 「関心・意欲・態度」, 「理解の能力」, 「表現の能力」, 「知識・理解」, 「情報量」

2005年度の研究において作成されたルーブリックは全体的(holistic)なものであるが、その作成過程においては、各グレードを分けるパフォーマンスの特徴を観点別に分析しており、同ルーブリックを改訂するためには、再度観点別に分析を行なうべきであると判断し、本研究においても観点別の評価を採用することとした。

なお、「情報量」は、国立教育政策研究所によって示された評価規準には含まれない観点であるが、2005年度の研究において、各グレードを分ける重要な特徴であることが明らかになっており、また「関心・意欲・態度」「表現の能力」「知識・理解」のいずれか単一の観点到に吸収させて捉えることが困難であるという意見が評価者から出されたため、試みに独立の観点として設定することとした。

評価に当たっては、まず本研究における分析の対象として50名の生徒を抽出した。抽出に当たっては、生徒の習熟度によって、以後の指導によるスピーキング技能の変容に何らかの違いが見られる可能性も考えられたため、1学期の実力テストの成績をもとに「上位」「中上位」「中位」「中下位」「下位」の5グループを設定し、それぞれ10名ずつ抽出することとした。

抽出された各グループの生徒のパフォーマンス・データはCD（テスト1）およびDVD（テスト2）に編集し、評価のための材料とした。

次に、1グループにつき教員2名ずつを割り当て、パフォーマンスの評価を行った。2名の評価者間での相談はしないものとし、各々が独自に評価を行なうこととした。また、2名の評価者の評点を合計したもの（満点8）をその生徒の評点とした。

評価の基準としては、2005年度の研究において明らかになった、観点別に記述された4段階の評価基準を用いた。各グループの平均評点は以下の通りであった。

表1 7月テスト平均評点

		上位	中上位	中位	中下位	下位	平均
テスト1	表現	6.5	5.2	4.6	4.8	4.4	5.1
	知識	6.8	5.3	4.9	4.7	4.4	5.2
	情報	7.5	6.6	5.0	5.1	4.9	5.8
テスト2	関心	7.0	7.2	7.5	5.5	4.4	6.3
	理解	7.2	8.0	6.2	5.4	4.3	6.2
	表現	6.6	5.5	4.5	4.3	3.5	4.9
	知識	6.5	5.5	4.2	4.2	3.5	4.8
	情報	6.1	6.0	4.4	4.1	3.7	4.9
平均		54.2	49.3	41.3	38.1	33.1	43.2

また、評価者は、上述の評価基準では捉えきれないパフォーマンスの特徴が見られた場合や、新たに基準を追加して評価するべきだと思われることがらが観察された場合は、自由記述によって説明を補足した。これは、ルーブリックを改訂する際の手がかりを得ることを意図してのことである。

3. 3. 評価結果の分析（事前）

表2の結果を見ると、テスト2の「表現の能力」「知識・理解」「情報量」の得点が他の観点と比べて低いことがわかった。また、評価者による自由記述においても、「質問の内容及かみあわない回答をしている」「発言と発言の間の沈黙が長い」「単純な文法ミスが多い」などの指摘が見られた。

これらのことから、対象の生徒たちがスピーキング技能に関して抱えている課題は、次の2点であると推測された。

- (1) 相手の発話に即応する流暢さ
- (2) 自分が発話する内容を考えながら、それを文法的に表現する正確さ

3. 4. 分析結果を受けた指導

前項で得られた分析を受け、2学期（9月～12月）の授業において、上述の2点の課題を克服するための指導を授業に組み込むこととした。通常の授業においては、映画を利用しながらリスニング技能の育成を図る活動も行っていたため、全ての授業時間をスピーキング技能の指導に割いたわけではないが、次のような活動を毎回の授業の中でルーティン的に実施した。

(1) 「スラスラ英会話」

平易な日常会話で用いられる表現や、身近な日本文化を英語で説明する際に用いられる表現を会話形式にしたプリントを生徒に配布しておき、毎回の授業の冒頭で、それらの表現を暗唱する練習を行った。また、何らかの話題を与えて、それについて生徒同士で会話を続ける練習も継続的に行った。文法的に正確であることよりも、一定時間沈黙せずに会話を続けることを要求した。

(2) Discourse Completion

100語程度の英語の「笑い話」を教員の読み上げにより1～2度聴く。ただし、「オチ」の部分は読み上げられない。また、メモも取らない。聴き取った情報から「オチ」を自由に考え、口頭で発表する。この活動を、ALTとのティーム・ティーチング授業の中で、生徒とALTのインタラクションを織り交ぜながら行った。

(3) Grammar Dictation

4～6文程度から成る平易な文章を2～3度だけ聴き、重要な情報だけを書き取る。そのうえで、書き取った情報から、読み上げられた元の文章を再構成する活動。その際、原文とまったく同じ文章を作る必要はなく、原文の意図を損なわない限り、別の表現を使ってもよい。ただし、文法的に正確な文章であることを条件とした。生徒は口頭で、再構成した文章を発表した。

(1)は流暢さの向上を意図した活動である。(2)は流暢さの向上と同時に、言いたい内容を正確に相手に伝える力の向上を意図した活動である。(3)は文法的な正確さの向上を意図した活動である。

これらの活動は、評価に用いたテスト1・2とはいずれも形式が異なる。テスト1・2のような活動において好ましいパフォーマンスを発揮することが指導の目標であるとするならば、評価時と同じ種類の活動を指導時にも行うことが望ましいとする考え方もあるであろう。

しかしながら、本研究においては、評価時と同じ活動を指導時に行うと、単なる活動への慣れからパフォーマンスが向上する可能性があると考え、あえて評価に用いるのとは異なる形式の活動で指導を行うこととした。

3. 5. パフォーマンス・データの評価 (事後)

2学期の指導の成果を確かめるため、2006年12月に、7月時と同様の2種類のテストを実施し、7月時と同じ手順で評価を行った。テストの内容は以下のとおりである。

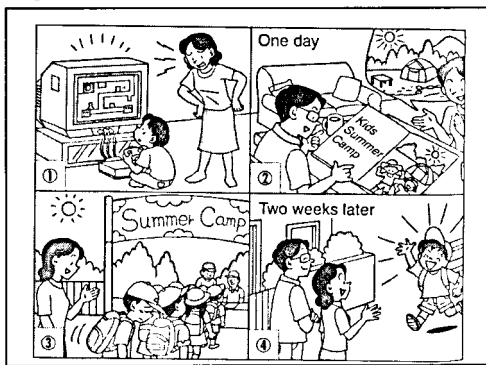


図2 12月テスト1

12月テスト2

Q: There are many kinds of comic books these days.

Do you like comic books? -①

→ A: Yes.

→ Q: Why? -②

What kind of comic books do you like? -③

Some people say high school students should read more books than comic books.

What do you think about that? -④

→ A: No.

→ Q: Why not? -②

What kind of books do you like? -③

Some people say high school students should read more books than comic books.

What do you think about that? -④

また、平均評点は次表の通りである。なお、評価者は各グループとも、7月と12月で同一である。

表2 12月テスト平均評点

		上位	中上位	中位	中下位	下位	平均
テスト1	表現	6.7	5.6	5.7	4.5	4.1	5.5
	知識	7.0	5.9	5.9	4.3	4.3	5.3
	情報	7.8	6.3	6.3	5.5	4.4	6.1
テスト2	関心	7.0	6.7	7.9	5.6	4.4	6.3
	理解	6.6	6.4	6.8	5.8	4.8	6.1
	表現	6.2	5.0	5.1	3.9	3.5	4.8
	知識	6.3	5.5	5.1	3.8	3.7	4.9
	情報	6.4	4.8	5.1	4.2	3.8	4.9
平均		54.0	46.2	47.9	37.6	33.0	43.7

3. 6. 評価結果の分析 (事後)

3. 6. 1. 全体の傾向

7月時と12月時の平均評点を比較すると、テスト2の「理解の能力」および「表現の能力」で評点が下降しており、それ以外の観点では、同点か、やや上昇していることがわかる。しかし、総平均点を見ると、43.2→43.7という変化であり、全体を平均すると、大きな上下動はなかったと判断される。

3. 6. 2. 生徒グループ別の傾向

生徒グループごとに評点の変動を見ると、中位グループがすべての観点において評点が上昇していることがわかる。また、上位グループと下位グループでは、あまり大きな評点の変動が見られない。

後述のように、テスト内容や評価者といった要因によって評点に誤差が生じることは避けがたく、この傾向の解釈にあたっては慎重を期さねばならないが、注意を払うべき点ではある。

3. 6. 3. 観点別の傾向

評価結果を評価の観点ごとに見てみると、テスト1の「知識・理解」において、中位および中上位グループの評点が比較的大きく上昇している。テスト1の「表現の能力」「情報量」およびテスト2の「表現の能力」「情報量」においても、中位グループに同様の傾向が見られる。

また、テスト1の「情報量」における上位および中下位グループ、テスト2の「情報量」における上位グループにおいても評点の上昇が見られた。

いっぽう、テスト1の「情報量」における中上位および下位グループでは評点が下降している。また、テスト2の「理解の能力」においては、上位・中上位グ

ループでは評点が著しく下降しているのに対し、中位・中下位・下位グループでは逆に評点が上昇している。この現象をどのように解釈すべきかについては不明であるが、推測される原因については5で考察する。

3. 6. 4. グループ・観点別の評点変動人数

評点の平均で捉えた全体の傾向は前項に示したようなものであったが、個々の生徒の評点がどのように変動したかを明らかにするため、さらに各グループおよび各観点において、7月時と12月時を比較して、評点が上昇した生徒と下降した生徒の人数をそれぞれ次表のようにまとめた。

表3 評点の変動人数(単位:人)

グループ	評点の上下動	テスト1			テスト2				計	
		表現	知識	情報	関心	理解	表現	知識		情報
上	上	2	5	4	3	2	1	2	5	6
	下	0	2	1	3	4	4	3	2	4
中上	上	6	7	1	2	0	0	3	1	3
	下	2	2	4	6	10	4	2	6	6
中	上	5	7	6	4	5	6	7	4	8
	下	0	1	1	0	0	2	0	1	2
中下	上	3	4	5	3	4	1	1	4	4
	下	3	6	1	4	3	4	4	4	4
下	上	3	3	3	4	4	4	4	4	5
	下	6	4	6	5	1	4	4	2	5
計	上	19	26	19	16	15	12	17	18	26
	下	11	15	13	18	18	18	13	15	21

「表現の能力」においては、テスト1では、中上位・中位グループにおいて評点の上昇した人数が下降した人数を上回ったが、中下位では上昇・下降が同数、下位では下降した人数の方が多かった。

いっぽう、テスト2「表現の能力」においては、評点が上昇した人数が多いのは中位グループのみであり、上位・中上位・中下位グループは下降した人数の方が多かった。

「知識・理解」においては、テスト1では上位・中上位・中位グループにおいて評点の上昇した人数が多かったが、逆に中下位・下位では下降した人数の方が多かった。全グループを通じて見ると、この観点に関しては、評点が上昇した生徒が多かったことがわかる。

いっぽう、テスト2の「知識・理解」においては、テスト1ほど顕著な傾向は見られない。

「情報量」においては、テスト1では上位・中位・中下位グループで、評点が上昇した人数が多かったが、

中上位・下位グループでは下降した人数が多く、全グループを通じて顕著な傾向は見られない。

いっぽう、テスト2の「情報量」では、中上位グループで下降した人数が多かったが、上位・中位・下位グループで上昇した人数が多かった。

4. 自由記述に見られたコメント

4. 1. 評価基準について

3. 2において述べたように、生徒のパフォーマンスを評価するにあたって各評価者は、2005年度に作成したルーブリックにおいては記述されていないが、評価基準として重要であると考えられる項目について、自由記述により見解を述べた。そこにおいて提出された項目を以下に集約する。

「関心・意欲・態度」

- ・適切な声量で話すこと
- ・アイ・コンタクトを取ること
- ・相手の発言が理解できなかったときに聞き返すこと

「表現の能力」

- ・発言したい内容を要領よく簡潔にまとめること
- ・相手の発言に対してタイミングよく応ずること

評価基準の運用にあたって留意すべきことがらとしては、次のような点が指摘された。

- ・1つのことがらについて話すのに多くの語数を費やして述べる場合、「情報量」の観点からは高い評価を与えるべきなのかもしれないが、「表現の能力」の観点からは、情報を整理して発言することができていないとみなされ、評価が下がる。
- ・多くの情報を盛り込んで話す場合、「情報量」や「表現の能力」の観点からは高い評価を与えることができる。しかし、発話が長くなればなるほど文法・語法面での誤りを犯す危険性も高くなる。その場合、「知識・理解」の評価は下げなければならないのか。
- ・情報量が多いが流暢でない話し方をしている場合や、逆に流暢ではあるが情報量が少ない場合があるので、「情報量」を評価の観点として独立させるのは便宜上良いことではないか。
- ・インタビュー・テストの場合、英語で対話すること自体への慣れや個人の性格もパフォーマンスに強い影響を及ぼす。評価の際に、それをどの程度考慮に入れていけばよいか難しい。

4. 2. 生徒のパフォーマンスについて

また、分析の対象となった生徒のパフォーマンスを

観察し、どのような指導・学習を行うことが彼らのスピーキング技能の伸長に寄与するか、についても、各評価者が自由記述により見解を述べた。

「関心・意欲・態度」

- ・日ごろから多様な話題に関心を持っておくことが、英語による会話の助けにもなる。
- ・まずは、相手が何を言っているのかを正確に理解しようとする姿勢が大事。

「表現の能力」

- ・つなぎ言葉を活用できるようにする。
- ・相手にわかりやすい発音を心がけること。
- ・焦らずに、既習の表現を使おうとすること。
- ・自分の言いたいことをすばやくまとめる練習が必要。

「知識・理解」

- ・定型的な構文や lexical phrase のような表現の材料を多く身につけると、よりスムーズに発話することができる。

5. 考察

以上の取り組みに関して、2で述べた2つの目標に照らして、その成果と課題を考察する。

5. 1. 指導と評価のサイクルに関して

本研究においては、2005年度に開発したルーブリックを実際の指導・評価場面において運用することで、生徒のスピーキング技能の伸長を図るサイクルを構築することを試みた。その過程においては、同ルーブリックは、生徒の現状を把握し、以後の指導の指針を得るために用いられた。つまり、診断的あるいは形成的な評価という位置づけにおいて同ルーブリックが運用された。

その結果、3. 3において述べたように、(1)相手の発話に即応する流暢さ、および(2)自分が発話する内容を考えながら、それを文法的に表現する正確さ、の2点に生徒の課題があるのではないかと推測された。

この判断をもとに、2学期の授業において継続的な指導が行なわれた。その結果は3. 5および3. 6に示したとおりである。4段階評価の評点という数値としては、あまり顕著な成果が挙げたとはいえないが、「情報量」の観点において一定の改善が見られた。また、自由記述において、中下位および下位の生徒に関して、「7月時点よりも、英語で話すことに慣れたような印象を受ける」との指摘も見られた。これらを踏まえると、(1)流暢さという課題に関しては、技能向上の兆候が見られたといえるかもしれない。

いっぽう、(2)文法的正確さという課題に関してはそ

のような傾向は見られず、自由記述においても改善を求める指摘が多く見られた。このことから、この課題に関しては、指導の成果が見られなかったといえる。

指導の成果が顕著に表れなかったことに関しては、複数の要因が考えられるが、一つには、事後評価のテストにおける練習効果を避けるために指導においてはテストと異なる種類の活動を用いたことが影響を及ぼしているかもしれない。活動の選定に問題があり、育成したい技能を適切に育成することができていなかった可能性があるということである。

ただし、事前評価を行ってから事後評価までの指導期間が3ヶ月ほどであったことを考慮に入れると、指導の成果を検証するには時期尚早であるともいえる。

しかしながら、そのように課題を発見し、指導の目標を設定する過程において、ルーブリックが大きな役割を果たしたことは確かである。7月時点、12月時点のいずれにおいても、ルーブリックを参照することで生徒のパフォーマンスを評価する視点が明確に定まり、以後の課題の設定も容易に行なうことができた。このことは、本研究において行ったような実践の成果といつてよいであろう。

5. 2. シラバス開発に関して

評価の機会は2回あったが、上述のとおり、そのいずれにおいても、2005年度に作成したルーブリックを基準として運用することに大きな支障は生じなかった。概ね同ルーブリックの記述に従って評価を行なうことが可能であった。

しかし、4. 1において評価者の指摘を挙げたように、同ルーブリックの記述には、まだ評価の実際に適応していない部分も見られた。

そもそも同ルーブリックを作成する際に用いられたパフォーマンス・データはインタビューの音声のみであり、グレード記述の対象となったのは生徒のパフォーマンスの言語的特徴に限定されていた。

しかし、本研究においては、それとは異なる種類のパフォーマンス・データ、つまりモノログの音声に加えてインタビューの映像も分析対象とした。そのため、同ルーブリックには記述のない「アイ・コンタクト」や「ジェスチャー」に評価者の意識が向けられ、同ルーブリックの記述を補足する必要性が明らかになった。

また、同ルーブリック作成の基準となったデータにおける生徒の発話は、「主張する」あるいは「説得する」という機能を担うものであった。本研究におけるテスト2でも、生徒は「主張する」という機能を実現させ

る発話を行なっており、この点においては、同ルーブリックの記述が、テスト2の評価に対しては直接的に適合するものであったといえる。

いっぽう、テスト1において生徒が実現させた言語の機能は、「描写する」であり、この点が、同ルーブリックが直接的には適合しない部分であったといえる。しかしながら、本研究における評価においては、この不適合は実際的な問題とはならなかった。というのは、「描写する」という言語の機能においては、必要な情報を不足のないように、かつ聞き手がわかりやすいように整理して伝える必要があるが、このことは「主張する」や「説得する」とも重複することであり、テスト1では、それに固有の側面が見出されなかったからである。逆に言えば、テスト1のような形式の活動であれば、新たにルーブリックを開発する必要はなく、本研究において用いられたものを適用することができるということである。

いずれにせよ、以上のようなことから、より多面的に生徒のスピーキング技能を評価することが可能になり、また、より幅広い言語の機能を評価の対象とすることができるようになったといえるであろう。

5. 3. その他の問題点

本研究においては、試みに実力テストの成績によって生徒を5グループに分けてデータの分析を行なったが、評点の変動に関しては、そのグループごとに特に顕著な特徴は見られなかった。

しいて挙げるならば3. 6. 2において述べたように、中位グループの評点の上昇傾向が強かったが、隣接する中上位・中下位グループには類似の変動が見られなかった。このことは、本研究における取り組みが、中位層の生徒にとって特に効果のあるものであったというよりは、評価者要因や偶然の結果である可能性を、より強く示唆するであろう。

また、同様に可能性が高く思われるのは、グループ分けの基準として用いた実力テスト実施時から12月時までの間に生徒の習熟度が変動していることによる影響である。12月時には、当初のグループ分けの意味が既に失われていた可能性がある。

以上のように、本研究が扱った範囲に限っては、実力テストにおける習熟度とスピーキング技能の伸長のありようには、有意な関係性は見出すことができなかった。

また、評価者要因に関しては、評価者による自由記述の中で重要な指摘がなされた。すなわち、評価者内で、7月時と12月時で評価基準の解釈が変化してしまっている可能性である。

上述のように、本研究においては既定のルーブリックの記述に従って評価を行なったが、当該の生徒のパフォーマンスが、グレード記述のどのレベルに当たるかの判断は、最終的には評価者の主観による。その際、評価者がこのような評価法に不慣れであったり、生徒のパフォーマンスがグレード間の境界に位置するようなものであった場合は、必ずしも評点が安定するとは限らない。

したがって、このように時期の離れたパフォーマンスを評価し、その結果を比較する場合には、最初の基準点となるパフォーマンス(本研究では7月のテスト)と、評価対象となるパフォーマンス(12月のテスト)を比較しながら評価することで、主観による誤差を最小化することが必要かもしれない。特に、個々の生徒の変容を評価によって捉えようとする場合、そのような方法を採用することで、個人内での技能の伸長を、より正確に測定することができるであろう。

本研究においては、そのような観点からの配慮がなされておらず、評価者内で評価基準にゆれが生じた可能性が否定できない。今後の検討課題である。

6. おわりに

本研究においては、指導と評価のサイクルを構築し、そこからスピーキング技能育成のためのシラバスを開発するために、2005年度に開発されたルーブリックを実際に運用した。既に述べたように、改善すべき点はいくつか明らかになったものの、概ね、2005年度の研究成果を基点に実践と構想の修正を図ってゆけばよいであろうという見通しが立った。今後は、これまでの成果を踏まえ、より幅広い言語の機能を包括したルーブリックの開発や、あるいは、他の学年との連続性を持った系統的シラバスの開発を徐々に進めていくべき段階に入るといえるであろう。

参考文献

- 1) 旺文社(2005)『英検準1級教本』旺文社
- 2) 国立教育政策研究所(2004)『評価規準の作成、評価方法の工夫改善のための参考資料(高等学校)』国立教育政策研究所
- 3) 三浦省五ほか(2006)「スピーキング技能の系統的伸長のための中高一貫シラバスの開発—評価規準・基準の設定と具体的な指導計画の策定のために必要なこと」『広島大学学部・附属学校共同研究紀要』第34号 pp. 131-135 広島大学学部・附属学校共同研究機構
- 4) 文部省(1999)『高等学校学習指導要領解説 外国語編英語編』文部省

Grade	Oral Performance
4	<p>The student answered questions with few grammatical mistakes which will not interfere the listener's understanding. His or her responses to questions were naturally quick enough so that no patience was required from the listener to wait. His or her voice was comfortably loud to hear. He or she maintained natural eye contacts at appropriate times. His or her statement had enough supportive facts which made his or her opinions persuasive, or they had enough information to be conveyed to the listener so that the listener could grasp the situation described. His or her statement was not redundant. He or she had no problems understanding the spoken English given by their interlocutors, most of which were spoken at a normal speed. The student also effectively used fillers that showed he or she was then trying to plan his or her next utterance, or needed very short pauses to do so. His or her pronunciation, intonation, or rhythm was very good at sentence level.</p>
3	<p>The student answered questions with few grammatical mistakes which will not interfere the listener's understanding. His or her voice was sometimes too small or unclear to hear, but not often. He or she tried to maintain eye contacts. His or her statement was persuasive and had supportive facts, or they had almost enough information to be conveyed to the listener so that the listener could grasp the situation described. His or her statement was sometimes redundant, but not to the degree of excess. He or she had little problems understanding the spoken English given by their interlocutors, but failed to answer promptly, and needed fairly long pauses during the conversation to think of suitable language forms to express what he or she wanted to say. But in such cases, he or she used fillers effectively and managed the time. His or her pronunciation, intonation, or rhythm was good at sentence level, but sometimes unnaturalness could be felt.</p>
2	<p>Though the student tried to answer in sentences correctly, his or her response often contained grammatical mistakes so that the listener had to infer his or her intention from the word-level information. His or her response often delayed so long that patience was required even from a sympathetic listener. His or her voice was often difficult to hear. He or she made few eye contacts. His or her statement had limited facts or information. He or she seemed to have serious problems understanding the spoken English given by their interlocutors, which needed to be repeated at times. He or she showed only incomplete sentences or word-level utterances after fairly long pauses rarely led by fillers, which often resulted in the time-management failure. He or she demonstrated problems in the phonological aspects of their utterances: pronunciation highly influenced by their L1 or pauses made at the end of almost each word.</p>
1	<p>The student tried to express himself or herself, but problems in grammar or vocabulary hindered them doing so. What utterance made by him or her could not be understood by their interlocutors easily because of its lack of information, grammatical mistakes, or pronunciation problems. His or her voice was difficult to hear. Particularly in the prompt reaction, he or she showed no significant supportive facts or reasoning, or important information to be conveyed to the listeners. His or her statement had little information. It was very difficult to judge whether he or she understood the questions. Such students had difficulty in controlling the conversation in the task.</p>