

英検型スピーキングテストの量的分析からみた  
指導法の提言のための研究

公益財団法人日本英語検定協会 英語教育研究センター  
委託研究 2013 年度報告書

2014 年 3 月 31 日

簗田 由己子 (清泉女学院短期大学)  
飯野 厚 (法政大学)  
中村 洋一 (清泉女学院短期大学)

## 1. はじめに

本研究は 2011 年 4 月から 3 年間の予定で開始されたが、東日本大震災の影響もあり、実質的には 2012 年、2013 年の 2 年間の研究となった。

本研究は、異なる形式を組み合わせて作られている英検スピーキングテストにおいて、その形式間にはどのような関係があるのか、それらは総合得点にどのように関係しているのか、という疑問から出発した。標準化された多くのスピーキングテストが、口頭エッセイ (verbal essay)、口頭プレゼンテーション (oral presentation)、統制インタビュー、情報転移、インタラクション・タスク、ロールプレイなど、複数の形式を用いてスピーキング力を測定している。「スピーキング力」の定義は未だ議論の余地のあるところであるが、スピーキング力が複雑な要素を含んでいることは確かであり、その評価には複数の形式を組み合わせてテストを構成することが多い。しかしながら、これらの形式の間にどのような関係があるのか、ある形式がスピーキング力全体の評価にどのように寄与しているのかについては未だ研究が少ない。

また、普通の授業でもスピーキングの練習は行われているものの、テストの場面においては不安感を覚え十分に力が出せない学習者も多い。スピーキングのパフォーマンスを高めるためにはどのような指導が必要なのか、ひいては効果的に英語でコミュニケーションをとるにはどうしたらよいのかを考える必要があると思われる。

そこで、本研究では「英検型スピーキングテストの量的分析からみた指導法の提言のための研究」を研究課題とし、2012 年の研究開始時には、以下の 4 点に焦点をしばった。

- 1) 音読・モノログ (絵描写課題)・ダイアログ (応答問題) 型スピーキングテストの間にはそれぞれどのような関係があるのか、またスピーキング力全体とはどのような関係にあるのか
- 2) スピーキングテストを受けた受験者のアンケートからみられる、受験者のテストに対する意識はどのようなものなのか
- 3) 英検型スピーキングテストの結果をその他の基準 (CEFR 等) と比較した時に言えることは何か
- 4) 調査結果に基づいて、スピーキングテストに効果的な指導法の提言

## 2. 2012 年度研究の概要

2012 年度は、上記研究課題の 1) と 2) を主に、データ収集と評価、評価結果の量的分析を行った。データ収集は、2012 年 4 月～5 月に行われ、スピーキングテストの評価結果、テスト直後のアンケート、英語力の指標である CASEC の 3 種類のデータを収集した。対象の学生は、非英語専攻の大学 3 年生 8 名、2 年生 13 名 計 21 名、英語・国際文化専攻の短期大学 2 年生 10 名、1 年生 10 名 計 20 名、全体で 41 名である。CASEC の平均点は

555 点 (1,000 点満点、 $SD=82.0$ ) であり、英検目安 (CASEC を開発した教育測定研究所による) ではおおよそ準 2 級～2 級の間に該当することがわかった。

## 2.1 スピーキングテストの分析と結果

スピーキングテストは、英検スピーキングテスト 2 級の問題を使用した。内容は、音読 (RA)、音読のパッセージに関する質問 (Q1)、3 コマの絵描写 (Q2)、受験者自身の意見を問う応答問題 (Q3, Q4) である。テスト中の発話はすべて録音され、3 人の評価者によって、英語検定協会が示している評価基準に基づいて 5 段階で評価された。その結果、以下の 4 つのことが確認された。①スピーキングテストの総合点と下位項目の間には相関がある。②スピーキングテストの構成では、Q3、Q4 の応答問題が約半分の説明率を示し、総合点の中で重要な役割を担っている。③スピーキングテスト総合点と下位項目との関係においては、音読とダイアログ (応答問題) との間に強めの相関がある。このことから音読とダイアログはスピーキング力の重要な要素と考えることができる。④音読とダイアログは Q2 における文法・語法(VGU)や CASEC 総点とも強めの相関がある。このことから、語彙・文法・語法といった言語知識と聴解能力はスピーキング力の前提として重要な要素であることも判明した。

## 2.2 アンケート分析と結果

テスト直後に、受験者自身が自分のパフォーマンスを振り返ってアンケートを記入した。「質問は理解できたか」、「十分に語ることはできたか」に焦点をあて、4 段階で自己評価をしてもらうとともに、絵描写と応答問題に関しては「言いたかったが言えなかったこと」についての自由記述欄も設けた。結果の分析からは、ダイアログにおいて受験者が「質問を理解できた」ことと教員評価との間にある程度の相関がみられたことから、スピーキング力の前提として、リスニング力も必要であることがわかった。また、受験者の自己評価と教員評価の関係では、2 者の間にある程度の相関を観察できたことから、受験者の手ごたえは教員評価とそれほどかけ離れていないということも判明した。絵描写課題のみが他の項目にはない負の相関を見せたこと、自己評価と教員評価の相関が弱かったことから、Q2 の絵描写課題は他の構成要素とはやや異質であることも分かった。

## 3. 2013 年度研究の概要

2013 年度は研究課題の 3) 英検型スピーキングテストの結果をその他の基準(CEFR 等)と比較した時に言えることは何か および 4) 調査結果に基づいて、スピーキングテストに効果的な指導法の提言の 2 点が研究ポイントであった。しかし 2012 年度の研究結果を検討した際に、発話内容をもう少し詳しく分析し検討をしてみたいという意見があり、本年度

は、3)の研究課題を、「受験者の発話内容を詳細に検討し、テスト結果と合わせて分析した時に言えることは何か」と変更し、研究を行うこととした。4)は予定通り行った。

#### 4. 分析方法

スピーキングデータをより詳しく分析するために、41人分全てのデータを書き起こして文字化した。文字化は次のルールに基づいて行われた。①聞こえたとおりにアルファベットで記述する。②1つの意味のかたまりをエクセルシートの1行に記述する。③”uhmm”, “etto”などのフィラーは[f]として、聞こえた回数分記述する。④おおよそ5秒以上の沈黙をポーズとして、[s]で表す。データの書き起こしはフィリピン人の英語教員2名が行い、本研究の研究者が内容を確認した。

書き起こされたデータは、言語分析指標である、流暢さ、語彙の複雑さの観点で分析され、Q2のみ正確さについても検討した。言語分析指標（もしくは談話分析的指標（Ellis, 2003））は、談話の中で起こる現象を、特定の言語的特性に焦点を当てて算出する指標である。流暢さとは、発話の中でフィラーやポーズなど、非流暢さを示す語を使わずに話せるかであり、本研究では1分あたりの語数（フィラーを除いた述べ語数÷話した秒数×60）、いわゆるWord Per Minute (WPM)を用いた。複雑さには、統語的複雑さと語彙的複雑さがある。統語的複雑さは、1ユニット当たりの節や語数を測る方法があり、語彙的複雑さには総語数に対する異なり語数を算出する方法があるが、本研究では語彙的複雑さのみを扱うこととした。語彙的複雑さを表す指数としては、延べ語数を異なり語数で割った異なり語率を使用するケースが多いが、異なり語率は発話語数に影響される難点がある。その解消法として、Giraud Index（ギロー指数）が用いられるケースが多い。Giraud Indexは延べ語数の平方根を用いるもので、異なり語率よりも比較的発話語数に影響を受けないといわれるので（小泉, 2009）、本研究ではこちらを採用した。正確さにおいては、1つの節や1つのASユニットあたりの誤りのない節やASユニット数を検討するケースが多い。本研究では、和泉・内元・伊佐原（2004）で取り上げられている、日本人学習者の英語スピーキングコースで使われている13の頻出エラーと書き起こしたデータを検討した結果、述語動詞の使用のみに限定して正確さを分析することとした。そして、Q2での発話の平均が5文程度であったことから、言い始めから5つ目までの文において、動詞が正しく使えているかどうかに着目することとした。

流暢さ、複雑さ、正確さそれぞれの指標を数値化した後、Q2からQ4について、スピーキングスコアと流暢さ、複雑さ、正確さの3項目間の分析を行った。さらに、その結果とともに、書き起こしたデータと合わせたときにそこから読み取れるものについて検討を行った。

## 5. 結果と考察

### 5.1 Q2（絵描写課題）について

#### 5.1.1. 記述統計

絵描写課題 Q2 について、受験者の産出言語における流暢さ (Fluency)、正確さ (Accuracy)、語彙の複雑さ (Complexity) と、教師による描写内容の評価 (Content Score) と言語的評価 (VGU Score : Vocabulary, Grammar and Usage Score) の 5 つの変数の記述統計と変数間の相関係数を算出した (表 1)。記述統計から、絵描写の内容スコア平均は 3.44 であった。評価基準では、3 コマの絵の中にある 5 つのポイントのうち、各コマ 1 つずつポイントを述べれば 3 点、残りの 2 点を述べられているかどうかを 4 点、5 点として採点するとあり、今回の結果では、3 コマの絵の中の 5 か所の言及ポイントのうち平均 3.44 箇所が話されたことになる。また、言語的評価 (VGU) は語彙、文法、語法の使用を総合的に判断するスコアで、平均値が 2.92 であった。1 分間当たりの発話語数 (WPM) は 48.2 語とゆっくりとしたものであった。絵の内容を言語に変換しながら発話する課題では、繰り返しやポーズが多発したことを示している。正確さ (accuracy) は、産出言語の中の最初の 5 つの述語動詞に着目し、その時制が過去形で使用されているかを百分率で示したものである。その平均値は 63.1% であった。おおむね半数以上の述語動詞が正しい形で現れたことになる。語彙の複雑さ (Complexity) は、Giraud Index を利用して算出した。これに関しては数値が高いほど使用語彙が豊富であることを示す。大学生のライティング分析では平均 6 ポイント台であるので (杉浦, 2008)、今回はそれよりはやや低い結果となった。

表 1. Q2 内容スコア、VGU スコア、流暢さ、正確さ、語彙の複雑さ間の相関と記述統計 (N= 41)

Q2	Content Score	VGU Score	Fluency	Accuracy	Complexity
平均	3.44	2.92	48.18 wpm	63.1%	4.94 (63.3%)
標準偏差	0.79	0.59	14.46 wpm	31.1%	0.63 (9.5%)
VGU	.667**	-			
Fluency	.529**	.750**	-		
Accuracy	.572**	.335*	.313*	-	
Complexity	.499**	.216	.340*	.439**	-

\*\*  $p < .01$ 、\*  $p < .05$

#### 5.1.2 相関分析と重回帰分析

相関係数の算出にあたり、はずれ値の検定によって除外されたデータはなく、正規分布をなすデータとしてみなされたので、ピアソンの積率相関係数を用いた (表 1)。まず Content Score と VGU の相関は .667 であり、中程度の関係が観察された。(図 1)。この関係は 1 つの課題に対して Content と VGU の 2 つの評価観点を設けているため、一定の重

なりが生まれてしまったものと思われる。さらに言えば、VGU を独立して評価することの難しさが課題として示されている。実際の英検スピーキングテストの場面では、録音をしていない為、発話を再度聞くことができず、その場で評価を下さなければならない。そのため、発話言及個所数に注意を払いながら、語彙・文法・語用の3点を総合してVGUを評価することは、評価者が妥当な判定をくださるのかも検討を要するのではないかと考える。

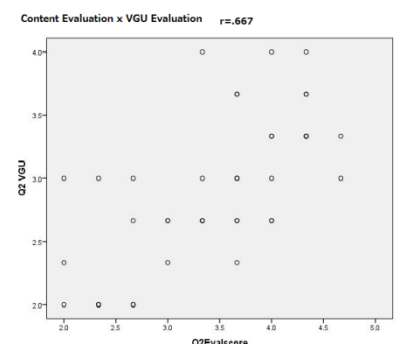


図1 Q2 Content Score と VGU の関係

3つの言語分析指標と Content Score の関係は、いずれも相関係数.50 前後と中程度の関係を示した。微量な差ではあるが Accuracy (.572,  $p < .01$ ) > Fluency (.529,  $p < .01$ ) > Complexity (.499,  $p < .01$ )という順であった(図2~4)。

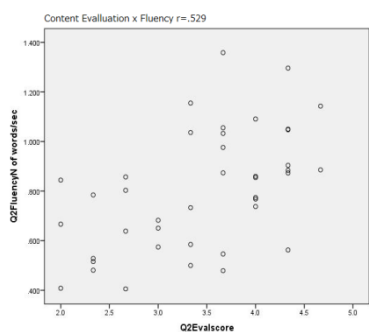


図2 Content Score と Fluency

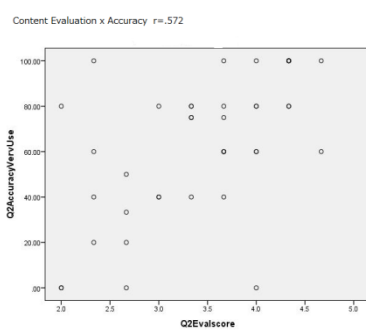


図3 Content Score と Accuracy

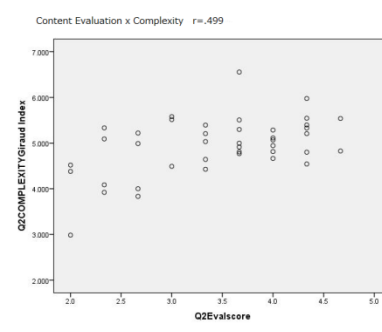


図4 Content Score と Complexity

3つの指標から Content Score を予測するために、ステップワイズ法による重回帰分析を行った。その結果、Complexity は除外され、Accuracy と Fluency の数値が予測に有意で、この2つの指標で46.3%を説明していることがわかった(表2)。十分な説明率が得られたとは言いが、とりわけ Accuracy が約33%を説明しており、正確さが絵描写の内容評価において重要であることが示された。

表 2 Content Score を独立変数とするステップワイズ法による重回帰分析の結果 (N= 41)

	B	標準誤差	標準化係数 B	t	有意確率	R <sup>2</sup> (調整済み R <sup>2</sup> )
Step 1						.327(.310)
Q2Accuracy	.015	.003	.572	4.357	.000	
Step 2						.463(.435)
Q2Accuracy	.011	.003	.451	3.601	.001	
Q2Fluency	1.268	.410	.387	3.095	.004	

VGU と言語分析指標との関係では、Fluency が.750 と最も強い関係を示した (図 5)。Accuracy とは.35 で弱い関係 (図 6)、また Complexity とは.216 でほとんど関係は見られなかった (図 7)。VGU と語彙の豊富さの指標である Complexity に相関がみられないという結果であった。

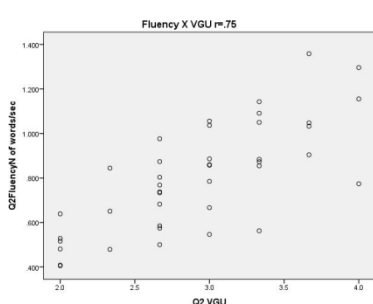


図 5 VGU と Fluency

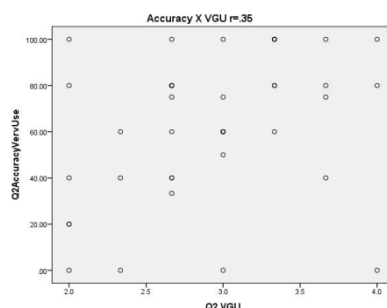


図 6 VGU と Accuracy

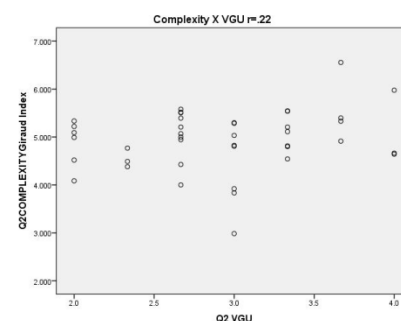


図 7 VGU と Complexity

ステップワイズ法による重回帰分析を行ったところ、Accuracy と Complexity は変数から除外され、Fluency のみによって 56.3%が説明される結果となった (表 3)。本研究では、使用語彙の多様さを指標として Complexity を扱っており、VGU における Vocabulary の評価とは高い相関が期待された。しかし、結果はそのようにならなかった。また、Grammar と共通する文法的正確さ (Accuracy) も従属変数から除外され、人的評価において VGU の観点の運用実態を精査する必要性を示す結果となった。本研究の評価者のコメントとして、その場で面接を行いながら発話の採点を行う場面では、発話メッセージの妥当性や一貫性、発話言語の流暢さ程度までしか判断ができず、録音した音声聞き直すなどしない限り語彙の複雑さや文法の正確さは、正確には評価できないのではないかと意見も聞かれた。

表 3 VGU を独立変数とするステップワイズ法による重回帰分析の結果 (N= 41)

	B	標準誤差	標準化係数 B	t	有意確率	R <sup>2</sup> (調整済み R <sup>2</sup> )
Step 1						.563(.552)
Q2Fluency	1.821	.257	.750	7.085	.000	

言語分析指標間の相関関係においては、全体として弱めの相関がみられた。そのなかでは、Accuracy と Complexity が.439 という中程度の関係を示した。この点は、正確さと複雑さのトレードオフに関する知見 (Yuan and Ellis, 2003 など) に沿わない結果であった。絵描写課題における 3 つの言語指標の独立性とその関係の検討にはさらなるデータやタスクタイプとの比較を要すると考えられる。

### 5.1.3 個別の発話データ精査

ここでは、スピーキングの評価結果と言語分析指標からみて特徴的な受験者の発話を取り上げ、それぞれについて検証を行う。

#### Student A Content Score が低いにもかかわらず VGU で高い評価を受けた受験者

Content Score は 2.0、VGU が 3.0 である。3 コマの中での言及個所は 2 か所のみであるので 2.0 は妥当な評価である。しかし、書き起こしデータを見てみると述語動詞の使用は正確ではないが、評価は 3.0 となっており、妥当とは言えない。原因としては、発話量は少ないが単語の羅列ではなく文で話すことができているため、実際よりは良い印象を与えている可能性があり、また評価者も印象で判断している可能性も否めない。

表 4. Excerpt of Student A

Duration	33sec	Fluency	40 wpm	Accuracy	0% (0/2)	Complexity	2.98 (64%)
Content Evaluation	2.0		VGU	3.0		CASEC	623
[f] Mr. and [f] Ms., Mrs. Sasaki send to their [f] grand, [f] send apples to their grandson [f] their grandson, and his mother use apples to make apple pie,							

#### Student B 流暢さが無い為に正確さはあっても Content、VGU の評価が低かった受験者

書き起こしデータでは文法はおおむね正確だったが VGU は 2.0、Content Score は 2.0 であった。発話に言い直しやポーズが多く、流暢さも 25wpm と平均値 (48.18wpm) よりもかなり低い。流暢さが無いためにこのような結果になったのではないかと想像される。評価自体が流暢さに影響されている可能性も考えられる事例である。

表 5. Excerpt of Student B

Duration	127sec	Fluency	25 wpm	Accuracy	80% (4/5)	Complexity	4.52 (58%)
Content Evaluation	2.0		VGU	2.0		CASEC	491
One day [f] [s] [f] One day Mr. and Mrs. Sasaki were picking apples on their farm [s] [f] They, they brought them at their home [f] They, they thought that they want to send their [f]grandchild, grandchildren							



They, [f] they send them to their grandchildren  
 [s] and their [f] grandchil grandchild [s] received them  
 Three days later, and his, his mother cooked apple, cooked apple cake  
 [f] by, by using these apples

**Student C** 妥当に評価されたと思われる受験者

発話語数はあるものの、3コマの絵の中で言及した内容は限られており、内容スコア 2.0 であった。流暢さは 52wpm と平均的であり評価者もメッセージを受け取りやすい話し方であったが、文法が不正確で複雑さも平均的であり VGU は 2.0 の評価であった。書き起こしデータを見てみても、Accuracy では述語動詞を正しく使用できたのは 20%、Complexity の Giraud Index も 4.09 と高くはなく、ほぼ正しく評価されていると思われる事例である。

表 6. Excerpt of Student C

Duration	159sec	Fluency	52 wpm	Accuracy	20% (1/5)	Complexity	4.09 (45%)
Content Evaluation	2.3		VGU	2.0	CASEC	546	
<p>One day Mr., Mr., [f] Mr. and Mrs. Sasaki were picking the apples on their farm          [s] [f] They, they make some delicious apples in their garden          They, they were picking apples [f] [s] So after that,          they have, they have some apples to their house          That evening, then, they send, they send the apples          that they, that they, [s] that they          [s] some apples send to people for example,          they are [f] some [f], they are families, for people who, who are sent apples,          make [f]some, something, something that use, using is using apples</p>							

**Student D** 高い評価を受けた受験生

内容スコア 4.3、VGU4.0 と高い評価を受けた受験生である。流暢さは 78wpm で、書き起こしデータを見てみると、フィラーはあるもののポーズや言いよどみ、言い直しもあまりなく、流暢に発話を行っている。また、述語動詞も正しく使えており、Complexity も 5.98 と平均よりも高かった。正確で複雑さも高かったこの受験者は、妥当に評価されていたとすることができる。

表 7. Excerpt of Student D

Duration	54sec	Fluency	78 wpm	Accuracy	100% (5/5)	Complexity	5.98 (78%)
Content Evaluation	4.3		VGU	4.0	CASEC	565	
<p>One day Mr. and Mrs. Sasaki were picking apples on their farm</p>							

[f] She told her husband, [f] let's send some apples to our grandson  
He agreed. That That evening, she she called her grandson son  
and [f]the, they started packing apples  
Three days later, [f] the package [f] was send [f] sent by by them  
and their son, [f] their grandson opened that and [f]  
his mother tried to make apple pie next time  
That's all

以上のデータに見られるように、Content Score や VGU の評価と書き起こしデータをもとに作成した言語分析指標とが整合しているケースと、そうでないケースが存在することがわかる。今後 Q2 においては、Content Score 及び VGU と流暢さ、複雑さ、正確さに不整合があるケースには何が影響しているか、さらなる検討が必要である。

Q2 の分析から得られたことは下記のとおりである。

- ①Content Score と VGU の間には中程度の相関がみられた。そのことから 2 つの指標を独立して評価することの難しさが課題として示された。
- ②Content Score と言語分析指標 (Fluency, Accuracy, Complexity) の間には、中程度の相関がみられた。
- ③Accuracy と Fluency が Content Score を予測する数値として有意であり、2 つの指標で 46.3% を説明していることがわかった。Accuracy は 33% を説明しており、正確さが絵描写の内容評価において重要であることがわかった。
- ④VGU は Fluency と強い相関を示し、Complexity では相関はみられなかった。
- ⑤Q2 では、オンサイトで Content Score と VGU の両方を評価することの難しさが示された。

## 5.2 Q3 および Q4 について

### 5.2.1 記述統計

受験者の考えや意見を問う Q3 と Q4 について、インタビュー・テストの実施と同時に行った評価 (Score) と、書き起こしデータを分析した産出言語の流暢さ (Fluency)、語彙の複雑さ (Complexity) の 3 つの変数の記述統計と変数間の相関係数を算出した (表 8、表 9)。記述統計から、Q3 のスコア平均は 2.67、Q4 のスコア平均は 2.79 であった。テスト後、受験者が「Q3 のほうが難しかった」との感想を述べていたことから、同じ能力を問うテスト項目であったが、課題の特性がスコアの平均に影響を与えたものと考えられる。このことから、受験者の課題に対する印象と評価者の評価には、妥当性があつたと言えると思われる。Fluency を示す WPM は Q3 が 43.46 語、Q4 が 41.38 語と、Q2 よりもさらにゆっくりとしたものであり、Q3 が Q4 よりも若干高い数値を示した。Complexity を示す Giraud Index

の値は、Q3が3.83、Q4が3.75であった。Q3は受験者にとって「難しい」という印象があり、スコア平均も低かったが、流暢さと語彙の複雑さで比較すると、Q4よりも若干高い数値を示した。

表 8. Q3 スコア、流暢さ、語彙の複雑さ間の相関と記述統計 (N= 41)

Q3	Score	Fluency	Complexity
平均	2.67	43.46wpm	3.83
標準偏差	0.94	18.96wpm	0.87
Fluency	.554**	-	-
Complexity	.526**	.581**	-

\*\*  $p < .01$ , \*  $p < .05$

表 9. Q4 スコア、流暢さ、語彙の複雑さ間の相関と記述統計 (N= 41)

Q4	Score	Fluency	Complexity
平均	2.79	41.38wpm	3.75
標準偏差	0.87	17.75wpm	0.87
Fluency	.714**	-	-
Complexity	.675**	.598**	-

\*\*  $p < .01$ , \*  $p < .05$

### 5.2.2 相関分析

ここでは、表 8、表 9 に示した、ピアソンの積率相関係数を用いた相関係数をもとに、スコア、流暢さ、語彙の複雑さの間の関係を検討する。Score と Fluency の間には、Q3 で.554 の中程度、(図 8、9) Q4 で.718 の強い相関 (図 10、11) が見られた。同様の関係は、Score と Complexity にも見られ、Q3 で.526 の中程度、Q4 で.675 の比較的強い相関が見られた。Fluency と Complexity の間には、大きな違いはなく、Q3 で.581、Q4 で.598 と、中程度の相関が見られた。

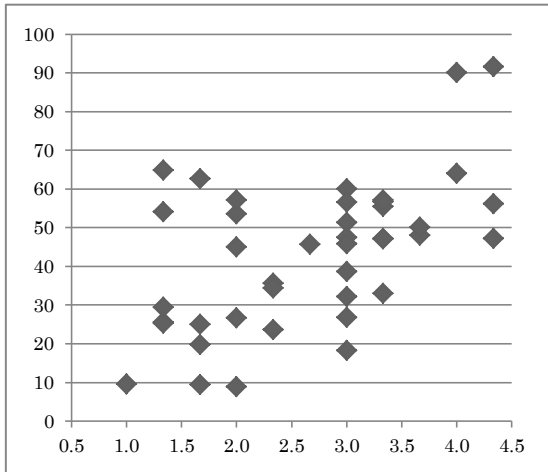


図 8 Q3Score と Fluency

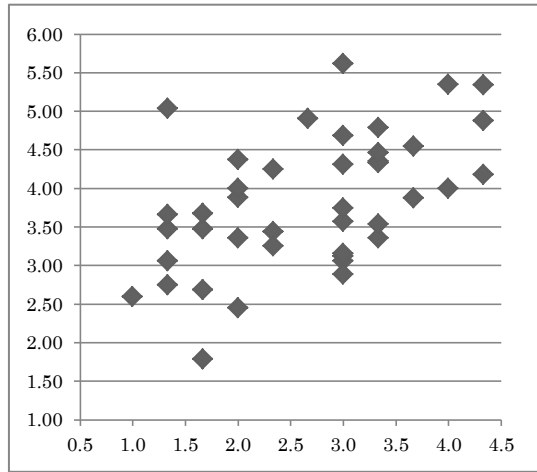


図 9 Q3Score と Complexity

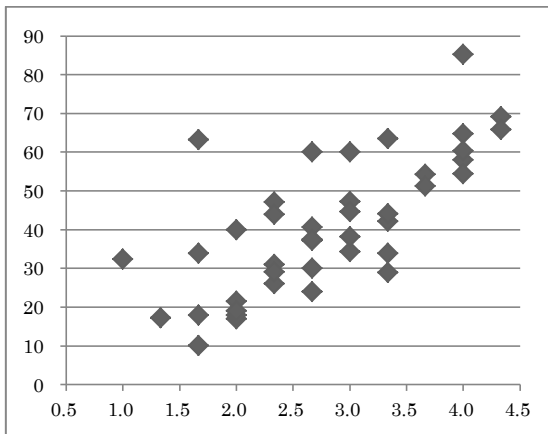


図 10 Q4Score と Fluency

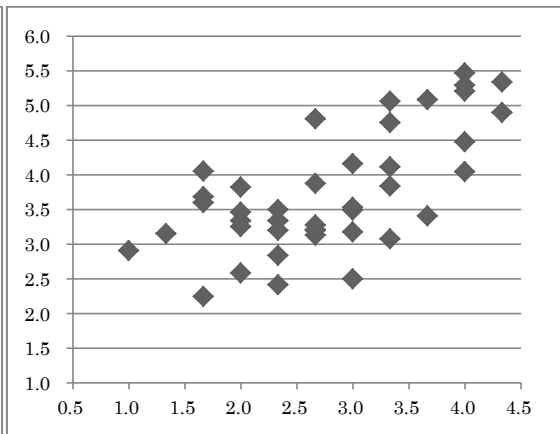


図 11 Q4Score と Complexity

Score とふたつの言語分析指標との相関は全般的に Q4 が高い。Q4 ではスピーキングテストの実施時における評価と、書き起こして分析した結果の関係性が強く、このことから、Q3 と比較すると Q4 の方が、実施時の評価が、「より正確であった」と言えるのではないかとと思われる。記述統計の考察でも述べたように、課題の特性とスコア平均のあいだには、関係性が見られた。今回使用した課題を具体的に見てみると、Q3 の質問は、“Some people say that because of convenience stores, supermarket will become less popular in the future. What do you think about that?”で、convenience stores と supermarket のふたつの事項についての意見を問う課題であり、さらに less+比較級を含む、日本人学習者にとっては、やや難しいと思われる表現が含まれていた。そのため、受験者には less+比較級の理解にためらいが生じて、質問に対して自分の意見をどのように展開していくかを考える以前に、正確に質問を理解するための負荷が生じていたのではないかと推測する。一方 Q4 は、“Today, some movies have a lot of violent scenes. Do you think people should stop making movies with violent scenes?”というものであり、ひとつのアイデアについて、

Do you think people should ~?の形式で、agree or disagree を問う課題であったため、質問の理解にはそれほど負荷がかからなかったと思われる。そのため、受験者は自分の意見を展開することに集中できる課題であったと言える。テスト課題のコントロールと評価の正確さという観点について、より深い検証が必要であると考ええる。

ふたつの言語分析指標と Score の関係を見ると、Q3 で、Fluency (.554,  $p < .01$ ) > Complexity (.526,  $p < .01$ )、Q4 で Fluency (.714,  $p < .01$ ) > Complexity (.675,  $p < .01$ )、という、同一の順番であり、Fluency と Score の間に、強い関係が見られた。今回使用したスピーキングテスト実施時の評価基準では、主に受験者が自分の意見とその理由・説明を「論理的」に「一貫して」、「正しい表現方法で」述べているかとの観点をあげているが、実際に評価は「流暢さの印象」をその評価に強く反映させていたのではないかと考えられる。Q2 では.340と低い相関を示した Fluency と Complexity の関係では、Q3 で.581、Q4 で.598 の中程度の相関が見られたことに関しては、Q3、Q4 の open ended な特性が影響を与えたものと考えられる。しかし、Q2 での考察と同様に、その場で面接を行いながら発話の採点を行う場面では、語彙の複雑さや文法の正確さも含めて内容の評価は困難であり、「流暢さの印象」が、「隠れた評価基準」として機能しているのではないかと考えられる。

### 5.2.3 個別の発話データ精査

Q3、Q4 で、高い Score を得た受験生について、その発話内容の書き起こしデータを検討し、どのような点が評価されたのかを以下で検討する。

#### Student E

全体的にバランスが良く、発話時間も長く、Fluency と Complexity も高い。特に Fluency は 91.2 wpm で、受験者全体平均 (48.18wpm) の倍程度の早さで話している。繰り返しや、文法的な誤りも散見されるが、Q3 では、最後に、“Because of these reasons...”という出だしで自分の意見をまとめて、Conclusion を述べている点が高い評価に結びついていると思われる。Q4 は Q3 に比べると言い直しも多いが、自分の意見（暴力的な映画を作ることはやめなくてよい）をきちんと主張していることが高い評価を得ることになったと考えられる。書き起こしデータを見てみると、動詞の使用に誤りは見られるものも見られるが、ネクサスの関係が伝わる表現で、チャンクとしてのまとまりが見られるものも多い。また、書き起こしデータからは観察できないが、英検スピーキングテストの評価基準の 1 つである「アティチュード」が高い受験者だったことも評価に影響しているように思われる。声のトーンや大きさ、伝えようとする態度などが非常に良く、英語のコミュニケーションとしてきちんと成立していたことが評価者に好印象を与えたことは否めない。

表 10. Q3: Excerpt of Student E

Score	4.3	CASEC	629	Duration	61 sec	Fluency	91.2 wpm	Complexity	4.87
<p>[f] I'm not [f] I'm disagree.            Because [f] convenience store [f] are always open            [f] They [f] in in 24 hours.            But the a a little a little a little expensive than supermarket.            So, the people want to want to buy more cheap one.            [f] yeah. So, that is that is true            that convenience store is more convenient for            [f] super [f] [f] than supermarket.            But [f] prices is prices sometimes not good.            [f] Many people think that [f]            people want to buy more cheap one.            [f] Because of these [f] reasons            I I don't I don't think the supermarket            [f] will become not [f] not popular.</p>									

表 11. Q4: Excerpt of Student E

Score	4.0	CASEC	629	Duration	76 sec	Fluency	85.2 wpm	Complexity	5.29
<p>No. I don't I don't think            [Interviewer: Okay. Could you tell me why?]            [f] Because that [f] some experts say that [f]            people were influenced by that such a movies            involve violent scene.            But I think [f] it is            they should think [f]            they should it's they should think            these things for them separated separate questions            and it is not same thing.            [f] Even if the [f] we see            even even if we see we see            [f] we watch a movie like which which            which has a lot of violent scene,            [f] we [f] we we [f] [s]            I I think I think we don't            we we must we will [f] will not will not influence</p>									

we we are not influenced by such scenes.

### Student F

Q3, Q4ともに4.3と高い評価を得た受験者である。Q3では、回答の前に less から more の言い換えを自分で確認している。発話中にはためらいの時間があったと思われ、そのために70秒の発話時間がありながら、総語数が少なく Fluency 低い値を示している。しかし、自分の考えをよくまとめており、同じ言葉を繰り返すような言い直しは見られない為、Complexity は5.4と高い数値となったと考えられる。発話速度はそれほど速くないが、全体的にすっきりと明確に意見を述べていることが高評価につながった例である。Q4ではためらいの時間は観測されず、発話時間は短いが簡潔に自分の意見を述べている。また、Q3、Q4ともに、まとめの部分で So を使用し、一貫性のある発話を保持していることが発話にまとまりを持たせており、評価者も理解しやすい内容であった。

表 12. Q3: Excerpt of Student F

Score	4.3	CASEC	645	Duration	70 sec	Fluency	48.0 wpm	Complexity	5.4
<p>[f] [f] That's mean convenience store more            People think convenience store more use convenience store than            Yes.yes, yes. Right.            [f] okay.okay.            [f] [f] I disagree with.            [Interviewer: Okay. Could you tell me why?]            Yes. [f] Convenience store open 24 hours.            and it's very convenienced for me and for us.            But [f] its [f] its cost            for example, [f] snacks or mineral waters and so on.            And its cost so more expensive than we than [f] supermarket.            So, [f] I in my case, I [f] I often go supermarket            to save money.</p>									

表 13. Q4: Excerpt of Student F

Score	4.3	CASEC	645	Duration	52 sec	Fluency	67.8 wpm	Complexity	4.0
<p>No.            [Interviewer: Okay. Could you tell me why?]            [f] When I was a child [f], that I don't like that movies.            Because it's very scared.            It's so [f]</p>									

But, [f] there are some interesting movies [f]  
including that violent scene.  
But that these movies are very nice or [f] very exciting one.  
So, [f] [f] it does not  
it don't have to stop making movie with violent scene.

以上、ふたつの例を見ると、open ended な課題である Q3、Q4 で高い評価を得た受験生は、外部基準である CASEC でも高いスコアを得ている。また、流暢さの印象が、スピーキングテストの評価に強い影響を持つと考えられることをすでに述べたが、Student F の Q3 で見られるように、極端に流暢さを欠いていない場合（それほど流暢でない場合でも）、発話の一貫性が際だっているために、高く評価されているケースもあった。文法的な誤りについては、インタビュー・テスト実施時に評価することは困難である点を指摘したが、Q3、Q4 の例を見ても、流暢さや一貫性ほどには、評価者の評価に影響を与えていないと言える。

Q3、Q4 の分析から得られたことは以下のとおりである。

- ①Score と Fluency の間には中程度もしくは強い相関がみられ、流暢さが評価に影響していることがわかった。
- ②Fluency と Complexity の間には中程度の相関がみられた。
- ③評価において、発話の一貫性やまとまりが影響していることがわかった。

## 6. スピーキングの指導へ

今回の研究から、スピーキングテスト及びスピーキング全般の指導についての提案を 3 つ上げたい。1 つ目はある程度の流暢さをもって話せるようになることである。今回の研究においても、Fluency とスコアの関係性は Q2, Q3, Q4 すべてで観察された。スピーキングでは文章とは異なり情報は短時間しか存在しないため、相手は聞きながら情報を処理して意味を理解していく。そのため、情報の処理を妨げるような長い沈黙や度重なる言い直しは、コミュニケーションを阻害しかねない。もちろん英語学習者の場合は心的不安や正確に話さなければいけないというプレッシャーもあるので、言いよどみやためらいもあるのだが、意味のまとまりごとに流れを持って話すことが望まれる。流暢さを高めると言われているタスクには、ワードカウンターを使用したペアワークで wpm を測定したり、同じトピックについて、話す時間を徐々に短くしながら複数回話させるペアワークなどがあげられる。また、準備時間が流暢さに影響を与えるという研究結果もあり（藤田, 2012）、話したいと思うことをどのようにプランすると流暢に話せるようになるのかについてのトレーニングも効果的であろう。

2 つ目は、チャンクで考えを発話できることである。Q3, Q4 の結果から、スコアの良い



受験者は、1つの考えを1つのまとまりとして話していた。チャンクで発話できるようになるためには、チャンクがなにかという意識を持たせることが重要である。例えば、チャンクで区切って英文を音読することで、1つのまとまりを一息で発話する感覚がつかめるであろう。そして、まとまりのある発想で発話をするためには、前述したことの繰り返しになるが、発話の前のプランニングのトレーニングが必要であろう。話したいことをなんとなくまとめるのではなく、話のポイントはいくつか、どの順番で話せばよいか、また各トピックでのメインポイントは何かなど、「どうやって考えるか」を学ぶことにより、コミュニケーション力は養われていく。これは英語に限ったことではなく、日本語にも言えることであり、普段からまとまりを意識しながらスピーチ等を行うことで力をつけることが可能であると考えられる。

3つ目は、チャンクをつなげて1つのまとまりとして情報を提供できる力である。ここでは、ネクサスをとらえて文章化する力が必要である。さらに、接続詞や連結語を使い、「論理的」で「一貫した」表現を用いて、相手に分かりやすく伝える練習が必要である。そのためには、わかりやすく表現するための論理的思考力を育成する必要があるだろう。問題解決型のタスクや、ディベートなどをも論理的思考力を必要とする活動である。また、スピーキングとライティングは、ともに発信する技能であることから関係も深い。論理的にわかりやすく話せるためには、論理的にわかりやすく書く訓練も必要である。スピーキングの指導をエッセイライティングと合わせて行うことも可能であろう。

## 7. おわりに

2年間の研究を通し、スピーキングテストというものを改めて考える機会を与えられたことは幸いであった。普段の指導の中でもなじみのある英検のスピーキングテストであるが、テストの項目がどのような役割を担っているのか、テストに対して受験者がどのように思っているのかについては疑問に思いながらも、詳しく分析などを行ったことがなかったため、今回得られた結果は今後の指導において大きな役割を担うものとなる。

今回は2年間という限られた時間であったが、本研究をさらに発展させるとすれば、下記の3点に取り組む必要があるだろう。1点はさらに幅広い受験者の調査を行うことである。今回は2級程度の受験者のみを調査したが、スピーキングテストは3級から1級まで行われており、今回得た結果がどのレベルの受験者にも当てはまるかは検討する必要がある。

2点目は書き起こしデータの分析において、受験者のエラーを述語動詞の使用以外においても、またQ3、Q4においても調査してみることである。今回は実験的にQ2のみにおいて調査を行ったが、応答問題においての結果も合わせて検討してみたい。

3点目はオンサイトでの評価の難しさとその対策である。今回は録音したものをきき直しながら採点を行ったが、それでも書き起こしたデータとつきあわせて検討してみると、整合性がとれていないものも存在した。特にQ2に関しては、内容のチェックをしつつ、文法や語彙使用まで採点をするとなると、かなりの熟達した評価者が必要になると想像する。

現在の英検スピーキングテストにおいては適切なトレーニングが行われていると思うが、今後とも現状の分析研究、トレーニングの両面で正しい評価を支えていく必要があると考える。

<引用文献>

- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press
- 和泉恵美・内元清貴・伊佐原仁志編 (2004) 『日本人 1200 人の英語スピーキングコーパス』  
アルク：東京
- 小泉利恵. (2009) 「常磐大学国際学部の学生の英語スピーキングの特徴：日本人英語上級者と英語母語話者との比較から」. 『常磐国際紀要 第 13 号』 .53-69
- 杉浦正利 (2008) 「英文ライティング能力の評価に寄与する言語的特徴について」 成田真澄 (代表) 『学習者コーパスに基づく英語ライティング能力の評価法に関する研究』 (pp.33-58). 平成 17 年度--平成 19 年度科学研究費補助金 (基盤研究 (C)) 研究成果報告書. URL <http://oscar.gsid.nagoya-u.ac.jp/paper/sugiura-naritakaken2008.pdf>  
(2014 年 3 月 22 日取得)
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, 1-27.