

能力記述文による自己評価と実際のスピーキング能力の関係

—英検 Can-do リストと CEFR-J を使って—

茨城県／筑波大学大学院在籍 横内 裕一郎

概要

本調査は、英語スピーキング能力を英検 Can-do リストや CEFR-J の記述文を用いた自己評価が、再話課題への回答で得られた実際のスピーキングパフォーマンスとどのような関係があるかを調査したものである。近年、英語スピーキング能力の評価を大学入試でも行おうという動きがある中、実際に中・高等学校でスピーキング能力を評価する機会は多くない。また、各中・高等学校が学習指導要領に基づいた学習到達目標を CAN-DO リストの形で具体的に設定するよう文部科学省が提言したことにより、今後 CAN-DO リストの開発と妥当性の検証を行うことが多くなると考えられる。本調査では、Shojima (2007a, 2007b) による潜在ランク理論を用い、自己評価と再話課題のパフォーマンスから得られた潜在ランク推定値をクロス集計によって比較した。その結果、英検 Can-do リストと CEFR-J の能力記述文を使用した自己評価は同じ方向性を示したものの、再話課題によるパフォーマンスとの関連性は低かった。

1

はじめに

学習指導要領の改訂に伴い、これまで以上に英語教育の現場では 4 技能の統合的な指導が望まれるようになった。また、一部の大学では TOEFL を大学入試に組み込む計画があり、上智大学と日本英語検定協会（以下、英検）の共同開発による TEAP が 2015 年度入試から一部の大学で実施される。しかし、教育現場ではスピーキング能力の向上を目的とした指導や評価がそれほど行われていないようであ

る。実際、Fulcher (2003) はスピーキングの評価は他の技能の評価よりも難しいことを指摘しており、中学や高校の教育現場でスピーキング能力を適切に評価することは、試験の実施や採点上のコストが高いため、非常に難しい。石川他 (2010) が日本大学生の英語スピーキング能力は低く、8 割以上が CEFR 換算で A1 以下のレベルにとどまることを指摘しており、複雑なタスクを学習者に課してもスピーキング評価に適当なパフォーマンスを引き出すことが難しいことが想定される。その一方、容易すぎるタスクを試験に設定した場合、パフォーマンスを適切に弁別できないことも予想される。

TOEFL iBT® や TOEIC® Speaking Writing test のスピーキングセクションを始めとするテストではさまざまなタスクを使用し、この懸念点を解決しているが、中学や高校などの教育現場で外部の試験を使用するにはコストがかかりすぎる。一方、複数のスピーキングタスクを含んだテストを作成し、そういったテストの実施、評価を行うことは上述のとおり教員への負担が大きすぎる上に、テストの妥当性を確保することも難しく、実現可能性が高くはない。

アルク社開発の SST は、訓練を受けた試験官が受験者のパフォーマンスに応じて出題内容を変え、次々に質問を投げかける形式のインタビュー試験で、パフォーマンスの評価を行う際に適切なタスクを課すことが可能になるが、これも教師がインタビュー役を務める場合に負担が大きく、実際に教育現場で似たような方式で試験を年間に何度も実施することは難しいと思われる。英検二次試験の形式でスピーキングテストを実施するにしても、やはり時間的負担が大きい。例えば英検 2 級の二次試験の

時間は約7分間に設定されており、これと同様のインタビューをクラス単位で実施するとすると、時間的な負担だけでなく、受験者間で試験内容が漏洩する可能性も否めないため、こうしたインタビューテストを学内の到達度テストなどで実施することは難しい。

ACTFLに基づく Oral Proficiency Interview computer (OPIc) では、試験開始の前に自身の興味のあることや余暇の過ごし方などを事前インタビューとして行い、その内容に関連した出題がされ、試験中盤にそれまでに回答した問題の難易度が適切であったかどうか質問が設けられている。これらの質問によって、受験者にとって適切な難易度のタスクを課すような設定となっている。

この他に、今井・吉田(2007)による中高生を対象とした英語スピーキングテストである HOPE (High school Oral Proficiency Examination) や Pearson 社による Versant など、さまざまなスピーキングテストがあるものの、外部試験を実施するには金銭的なコストがかかりすぎ、中学校や高校がそれぞれで独自のスピーキングテストを行う際には人的・時間的なコストが大きく、どちらのケースでも年間に複数回のテストを実施するには負担がかかる。

そこで、スピーキングテストの実施前に、OPIcのように事前アンケートを行うことで、学習者自身が「できるかできないか」のラインにあるタスクを自己申告し、それに基づいた試験課題を課すことで適切なパフォーマンスを引き出すことができるのではないかと考えた。

投野他(2013)による CEFR-J や公益財団法人日本英語検定協会(n.d.)による英検 Can-do リストなどの能力記述文を使用し、試験前に自己評価を行い、どのタスクを使用すれば実際にパフォーマンスを適切に引き出すことができるのかを知ることができれば、試験に使用されるタスク数を減らし、スピーキングテストの実施と採点にかかる人的コストを削減できるものとする。

本調査では、英検 Can-do リストの記述文(can do statement や can do descriptor など呼称はさまざまあるが、本稿ではすべて記述文と表記する)と CEFR-J の記述文を使用した、自身の英語スピーキング能力の自己評価とリテリング課題によって得られたスピーキングパフォーマンスの関係を調査し、

試験前に自己評価を行うことの利点を探ることを目的としている。また、英検 Can-do リストを使用した自己評価と CEFR-J の能力記述文を使用した自己評価の傾向に差が生じるかどうかについて併せて調査を行った。

2

先行研究

2.1 英検 Can-do リストと CEFR-J

2.1.1 英検 Can-do リスト

英検(n.d.)によると、英検 Can-do リストは延べ20,000人を超える英検1級から5級の合格者に対し「具体的にどのようなことができる可能性があるか」について調査したものをリスト化したもので、「英検合格者の実際の英語使用に対する自身の度合い」を示したものである。

臼田(2009)は英検 Can-do リストに書かれていることを、学習者が実際にできるのかどうかを調査することで、英検 Can-do リストの妥当性の検討を行っている。英検準2級合格者を対象に、4つの英検 Can-do リストの能力記述文を基に作成したタスクに回答させることで、英検 Can-do リストの能力記述文と実際のパフォーマンスの関係について比較したところ、「自分の将来の夢や希望について話することができる」という能力記述文に関してはやや妥当性が低いようにも思われる結果となったが、話す内容があればタスクを完成できたのではないかと考察している。この結果、部分的ではあるものの、英検 Can-do リストの記述には妥当性があると見込まれると報告している。

また、木村(2012)は英検 Can-do リストに基づく自己評価の結果と能力記述文の関係を調査したところ、リーディングでは自身の能力を過小評価する傾向があり、リスニングの能力を過大評価する傾向があったものの、高い相関が見られたことを報告している。このように、自己評価の際にやや不安定な部分があることがわかるものの、英検 Can-do リストに基づく自己評価と、実際の能力との間にはある程度の関連性があることがわかる。

2.1.2 CEFR-J

CEFR-J は2012年に公表された CEFR (Common European Framework of Reference for Languages)

型の Can do descriptor 群で、日本人英語学習者 5,000名以上を対象に、さまざまなタスクと自己評価を課し、作成された。CEFR-J の特徴として、本家 CEFR のレベルが A1, A2, B1, B2, C1, C2 の 6 段階であるのに対し、Pre A1 として A1 のレベルに届かない学習者の能力をカバーしていることや、A1 レベルを 3 段階、A2 から B2 レベルを 2 段階に分割し、計 12 段階としたことで、より詳細に学習者のレベルを分けることができるようになっている。また、「話すこと」に関する Can-do statements が「やり取り」と「発表」の 2 つに分かれており、より詳細にスピーキング能力についての記述がなされていることがわかる。

2.2 潜在ランク理論

潜在ランク理論は Shojima (2007a) によって提唱された新しいテスト理論で、古典的テスト理論や項目応答理論で見られるような間隔尺度でテストの結果を示すのではなく、5～20 段階の比較的少ない順序尺度でテストの結果を解釈するところに特徴がある。荘島 (2010) はテストの得点の持つ意味について以下のように示しており、これまでのテストのように細かく評価するのではなく、漠然と分類できれば十分であると主張している。

荘島 (2010, p.84)

身長計や体重計が、ほとんど同じ 2 人の身長や体重の違いを見ぬくことができる解像度（精度）の高い測定道具であるのに対し、テストは、同じくらいの学力の持ち主の違いを見分けるほど解像度が高い測定道具ではない。（中略）体重計がグラムの単位までの精細な解像度を持つのに対し、100 点満点のテストが 1 点 1 点までの細かい解像度を持っているとは考えにくい。すなわち、テストは、学力をせいぜい 5～20 段階に分けるくらいの分解能しかない。

また、荘島 (n.d.) は、「潜在ランク理論は Can-do chart の作成支援ツールであると言っても過言ではありません」と述べている。項目参照プロファイル (Item Reference Profile ; IRP) を調査することによってそれぞれの潜在ランクに該当する受験者がどのような問題に正答し、どのような問題で間違いを起こしているのかを分析することで、Can-do

statements を作成することができる、と主張している。

潜在ランク理論を利用した分析には、自己組織化マップ (self-organizing map ; SOM) と生成トポグラフィックマッピング (generative topographic mapping ; GTM) のメカニズムを用いてモデルを推定する 2 種類のアプローチがある。前者は毎回の計算が微小に異なる特徴があり、後者は毎回の推定で同じ結果になる。また、SOM はサンプルサイズが比較的小さくてもモデルに適合し、項目参照プロファイルが GTM に比べてやや滑らかになる傾向がある。一方、GTM はサンプルサイズが 10,000 以上のケースで使用するべきであるとされている (荘島, n.d.)。SOM を使用する際の最低サンプル数については特に言及がないが、小泉・飯村 (2010) は 147 名のデータを分析する際に SOM による推定を行っている。

潜在ランク理論の分析で得られる指標はさまざまあるため、Shojima (2007a, 2007b, 2008a)、小泉・飯村 (2010)、木村 (2013)、小山・木村 (2011)、荘島 (2010) を参考に以下に概要をまとめた。

a. テスト参照プロファイル (Test referenced profile ; TRP)

特定の潜在ランクに属する受験者がテストで取ることが予想される期待得点。

b. 項目参照プロファイル (Item referenced profile ; IRP)

項目応答理論の項目特性曲線のようなもので、特定の潜在ランクにある受験者がその項目に正答する確率を示したもの。

c. 潜在ランク (Latent rank)

一言で言い換えると熟達度や学力のこと。事前に決定したランク数に応じてテストの結果を分割することができる。潜在ランクの個数はサンプルサイズが大きければ 10 程度となってもよいが、サンプルサイズが小さい場合は潜在ランクの数も小さくするべきである。

d. 潜在ランク分布 (Latent rank distribution ; LRD)

特定の潜在ランクに何名の受験者が該当するかを示す潜在ランクの分布のことで、分析前に事

前分布を仮定して分析をする。分析ソフトの Exametrika では一様分布と正規分布を事前分布として仮定することができる。

e. ランクメンバーシッププロファイル (Rank membership profile ; RMP)

受験者が各潜在ランクに当てはまる確率を示したものの。木村 (2013) によると、「潜在ランクの推定値が同じだったとしても、RMP の違いによって、受験者に異なるフィードバックを返すことができる」ということで、同じ潜在ランクに属する複数の受験者のパフォーマンスを比較することも可能。

f. ランクメンバーシップ分布 (Rank membership distribution ; RMD)

特定の潜在ランクに受験者が何人該当するか示す分布。受験者ごとの RMP の単純和 (小泉・飯村)。

g. 境界カテゴリ参照プロファイル (Boundary category reference profile ; BCRP)

潜在ランク尺度を通して各カテゴリ以上を選択する確率 (荘島, 2010)。

h. 項目カテゴリ参照プロファイル (Item category reference profile ; ICRP)

カテゴリごとの IRP。多値モデルによって潜在ランクを推定する場合に使用する。

潜在ランク理論のうち、リッカート尺度のような心理尺度を使用した分析をする場合、拡張モデルの 1 つである段階的ニューラルテスト (Graded neural test ; GNT) モデルを使用した分析を使用する。

荘島 (n.d.)

GNT モデルは、多値の順序データに対応するための LRT (Latent rank theory) / NTT (Neural test theory) モデルです。テストレット項目やリッカートタイプの心理質問紙に運用することが可能です。カテゴリ数が 2 の時は 2 値モデルになります。

2.3 再話課題

再話課題は、読解もしくは聴解後にどのような話だったのか覚えている内容を口頭で再生させる課題である (中川, 2012)。この再話課題をスピーキングテストに昇華させたものが、Hirai and Koizumi (2009, 2013), Koizumi and Hirai (2010) による Story Retelling Speaking Test (SRST) である。SRST では 100 語から 150 語程度の文章を 2 分間で読んだ後、2 分 30 秒でその内容と意見・感想を述べる課題である。また、他の再話課題との大きな差違として、実際には採点されないものの、内容理解問題を 3 問程度含むことが挙げられる。

英検 Can-do リストでは準 1 級の記述文の中に、「読んだ本や見た映画について、そのあらすじを述べるができる」という記述があり、CEFR-J にも「短い読み物か短い新聞記事であれば、ある程度の流暢さを持って、自分の感想や考えを加えながら、あらすじや要点を順序立てて伝えることができる (B1.2)」と記載されており、何らかのインプットを基に、その内容や意見・感想について話す能力が上級の英語学習者には必要とされている。

3 調査 1

3.1 調査の目的

本調査の全体の目標は、日本人英語学習者が英検 Can-do リストと CEFR-J の記述文を使用し、自身の英語スピーキング能力をどの程度自己評価できるかを明らかにすることである。その第 1 段階として、英検 Can-do リストと CEFR-J の記述文の自己評価の関係性を探る。

RQ1-1 英検 Can-do リストの記述文を使用した英語スピーキング能力自己評価と、CEFR-J を使用した英語スピーキング能力自己評価の結果に差異はあるか。

3.2 マテリアル

英検 Can-do リストのうち、3 級、準 2 級、2 級、準 1 級の「話すこと」に関する記述文、計 26 問と CEFR-J Ver 1.1 の「話すこと (発表)」の記述文 22 問 (Pre. A1 から C2 まで) を使用し、「4」= 自信がある、「3」= やや自信がある、「2」= あまり自

信がない,「1」＝自信がない,の4件法で回答させた。なお,回答時間は約15分だった。

3.3 協力者

18歳から29歳までの日本人大学生63名,高等専門学校生12名,社会人7名の計82名が本調査に参加した。そのうち,適切な回答が得られていないデータ(欠損が多すぎる場合,明らかにいかにげんにマークした場合など)と調査2で行ったスピーキングテストのタスクすべてに回答していないデータを除外し,最終的に75名分のデータが分析の対象となった。

3.4 手順

2013年5月に自己評価アンケートを実験協力者に課し,3.2に記載の内容について自己評価を行ってもらった(資料1)。大学生を対象としたアンケートではマークシートを利用し,質問文に対する自信度という観点で自己評価を行ってもらい,高専生ならびに社会人にはマークシートではなく質問紙に直接選択肢の番号を記入してもらう形で自己評価アンケートを行った。

3.5 分析

本調査の分析は小山・木村(2011)の潜在ランク理論を使用した調査を参考に分析を行った。自己評価アンケートでは潜在ランク理論のうち,自己評価アンケートに4件法のリッカート尺度を使用しているため,拡張モデルの1つであるGNTモデルを採用した。英検Can-doリストの3級から準1級の4つの級の記述を自己評価アンケートに使用したため,英検Can-doリストとCEFR-Jの両方の調査で潜在ランクを4段階に分けることとした。また,事前分布は一様分布を仮定している。なお,潜在ランク理論による分析にはExametrika Version 5.3を使用し,クロス集計にはSPSS Version 21を使用した。

3.6 調査1の結果

3.6.1 英検Can-doリストを使用した自己評価

26項目の質問がこの自己評価アンケートにあったため,自己評価の合計点は26点から104点の範囲内に収まる。表1に英検Can-doリストとCEFR-Jを使用した自己評価の本調査の記述統計を,表2に英検Can-doリストを使用した自己評価のテスト参照

プロファイル,潜在ランク分布,ランクメンバーシップ分布の情報を記載する。

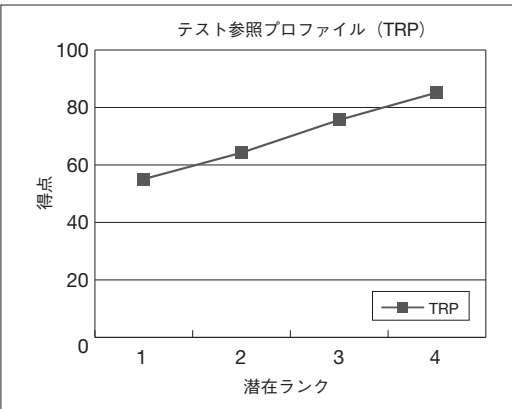
■表1:英検Can-doリスト,CEFR-Jを使用した自己評価の記述統計

	英検Can-doリスト	CEFR-J
受検者数	75	75
項目数	26	22
最小値	36	26
最大値	104	85
平均値	70.15	54.75
分散	261.10	157.92
標準偏差	16.16	12.57

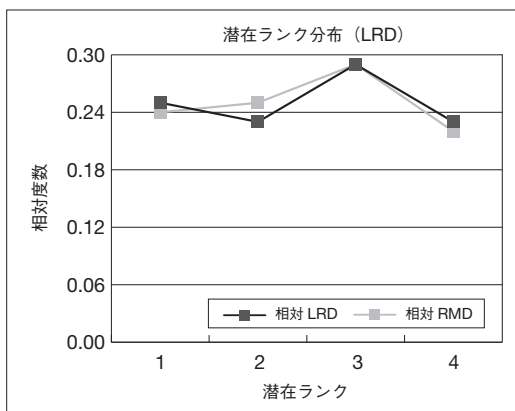
■表2:英検Can-doリストを使用した自己評価のTRP,LRDとRMD

	Rank 1	Rank 2	Rank 3	Rank 4
TRP	55.08	64.26	75.6716	85.20
LRD	19	17	22	17
RMD	17.87	19.01	21.56	16.56
相対TRP	0.53	0.62	0.73	0.82
相対LRD	0.25	0.23	0.29	0.23
相対RMD	0.24	0.25	0.29	0.22

図1は英検Can-doリストを使用した英語スピーキング能力自己評価のテスト参照プロファイル(TRP),図2がLRDとRMDである。この結果から,およそ10点刻みとなるが,55点程度記録した受験者は潜在ランク1に,65点程度の受験者は潜在ランク2,75点で潜在ランク3,85点で潜在ランク4となった。中には58点で潜在ランク1に割り振られる受験者もいれば,同じ点数で潜在ランク2に割り



▶図1:英検Can-doリストを使用した英語スピーキング能力自己評価のTRP



▶ 図2：英検 Can-do リストを使用した英語スピーキング能力自己評価のLRDとRMD

振られる受験者もいた。出力された RMP を確認したところ、前者が潜在ランク1に該当する確率が70.9%であったのに対し、後者が12.7%であった。このように、自己評価の得点と同じ得点だからといって、必ずしも同じ潜在ランクに該当するわけではないことが確認された。

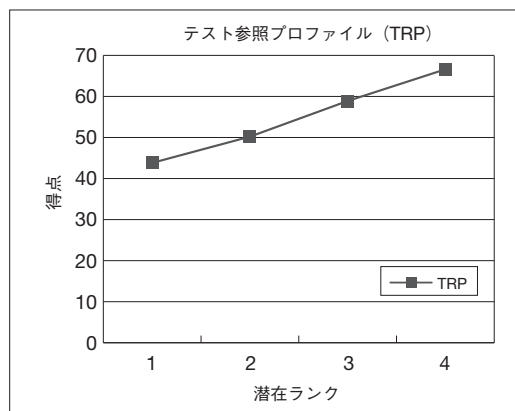
3.6.2 CEFR-J を使用した自己評価

CEFR-J を使用した自己評価についても同様の手法で調査した。表3は CEFR-J を利用した自己評価の TRP, LRD と RMD を示したものである。

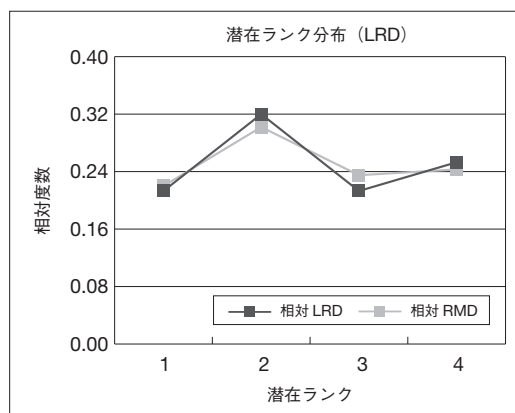
図3は CEFR-J を使用した英語スピーキング能力自己評価の TRP, 図4が LRD と RMD を示している。英検 Can-do リストを使用した場合と同様、それぞれの潜在ランクに25～30%程度ずつ受験者が割り振られたこととなる。

■ 表3：CEFR-Jを使用した自己評価のTRP, LRDとRMD

	Rank 1	Rank 2	Rank 3	Rank 4
TRP	43.813	50.202	58.886	66.603
LRD	16	24	16	19
RMD	16.520	22.652	17.628	18.199
相対TRP	0.509	0.584	0.685	0.774
相対LRD	0.213	0.320	0.213	0.253
相対RMD	0.220	0.302	0.235	0.243



▶ 図3：CEFR-Jを使用した英語スピーキング能力自己評価のTRP



▶ 図4：CEFR-Jを使用した英語スピーキング能力自己評価のLRDとRMD

3.6.3 英検 Can-do リストによる自己評価と CEFR-J による自己評価の比較

英検 Can-do リストを使用した自己評価では、Rank 3に該当するデータが最頻値となった一方で、CEFR-J を使用した自己評価では Rank 2の該当者が最も多くなった。Dunlea (n.d.) によると、CEFR と英検の関係は以下の表4のようになっている。

■ 表4：英検とCEFRの関係について

CEFR	英検
C2	
C1	1級
B2	準1級
B1	2級
A2	準2級
A1	3級・4級・5級

(注) Dunlea (n.d.) より引用

るため、英検1級レベル以上に相当するC1とC2の記述を含むCEFR-Jを使用した自己評価でRank2が最頻値となった結果は妥当なものであると言えるだろう。

次に、英検Can-doリストとCEFR-Jの潜在ランクをクロス集計し、自己評価の関係について調査を行った。

■ 表5：英検Can-doリストとCEFR-Jの潜在ランクのクロス集計

	CEFR-J					
		1	2	3	4	total
英検	1	13	4	1	1	19
	2	2	12	3	0	17
	3	1	6	10	5	22
	4	0	2	2	13	17
	total	16	24	16	19	75

(注) χ^2 検定 $p < .001$

3.7 調査1のまとめ

英検Can-doリストの潜在ランク1のデータはCEFR-Jでも潜在ランク1に該当することが最も多く、ランク2、3、4でも同様の傾向が見られた。英検Can-doリストとCEFR-Jの記述文を使用した自己評価では、どちらかの自己評価で自身の英語スピーキング能力を高く評価している受験者はもう一方の自己評価でも高く自己評価をしており、どちらを使用してもある程度同じような結果が得られるものと推測される。しかしながら、CEFR-Jを使用した自己評価では潜在ランク2に該当する結果が最も多くなり、英検Can-doリストを使用した自己評価では潜在ランク3に該当する結果が最多となった。これは、CEFR-Jを使った自己評価の場合、質問の項目数自体は英検Can-doリストに比べて少ないものの、より熟達度の幅が広がったためであると考えられる。

4

調査2

4.1 目的

調査2は、英検Can-doリストとCEFR-Jを利用した自己評価と、実際のスピーキングパフォーマンスの関係を調査する目的で行われた。今回の調査では、英検Can-doリスト準1級の項目の1つである「読んだ本や見た映画について、そのあらすじを述

べることができる」と、CEFR-J「話すこと（発表）」のB1.2記述文である「短い読み物か短い新聞記事であれば、ある程度の流暢さをもって、自分の感想や考えを加えながら、あらすじや要点を順序だてて伝えることができる」に対する回答と、再話課題のパフォーマンスを比較した。

RQ 2-1 英検Can-doリストとCEFR-Jによる自己評価潜在ランクと再話課題の潜在ランクはどのような関係か。

RQ 2-2 英検Can-doリスト準1級の「読んだ本や見た映画について、そのあらすじを述べるができる」に対し、「4＝自信がある」、「3＝やや自信がある」と回答した受験者は再話課題で高いパフォーマンスを発揮しているか。

RQ 2-3 CEFR-J B1.2の「短い読み物か短い新聞記事であれば、ある程度の流暢さをもって、自分の感想や考えを加えながら、あらすじや要点を順序だてて伝えることができる」に対し、「4＝自信がある」、「3＝やや自信がある」と回答した受験者は再話課題で高いパフォーマンスを発揮しているか。

4.2 マテリアル

Hirai and Koizumi (2009, 2013), Koizumi and Hirai (2010) の Story Retelling Speaking Test (SRST) を参考に、再話課題を4問作成した。本調査に使用した再話課題の条件は、2分間で問題文を読み、続いて2分30秒で読んだ内容を話し、さらにその話題についての意見・感想を述べる課題である。また、録音された英文を2度聞かせ、その内容と意見感想を2分30秒で話してもらう課題も2問課した。それぞれのパッセージの持つ情報量や難易度がパフォーマンスに影響しないよう、表6のように総語数、Flesch-Kincaid Grade Level (FKGL), Flesch-Kincaid Ease (FKE), JACET 8000のレベル1の語が含まれる割合、リスニングパッセージとした場合のWPMと再生時間をできる限り均等にした。

再話課題の評価基準として Hirai and Koizumi (2013) の empirically derived, binary-choice, boundary-definition (EBB) 2 scale for SRST を使用し、

■ 表 6：再話課題のマテリアル情報

	総語数	FKGL	FKE	JACET 8000 Level 1 (%)	WPM	再生時間 (sec)
Passage 1	149	7.8	65.0	79.87	146.1	62.1
Passage 2	149	7.8	63.0	78.52	146.9	60.9
Passage 3	98	7.6	64.8	88.00	146.8	40.5
Passage 4	98	7.6	59.3	80.81	143.7	40.9

■ 表 7：採点者間信頼性

	Communicative Efficacy	Grammar and Vocabulary	Pronunciation	Total score
α	.987	.933	.914	.978

Communicative Efficiency, Grammar & Vocabulary, Pronunciation の3観点について採点を行った。採点の信頼性確保のため、英語教育を専攻する大学院生1名にトピック1について同じ採点基準で採点してもらい、クロンバックの α を用いて採点者間信頼性を算出した。その結果、どの項目でも極めて高い採点者間信頼性を得ることができた(表7)。

4.3 協力者

調査2の協力者は調査1の調査協力者と同じである。

4.4 手順

調査1のデータ収集の後、2013年5月から6月下旬にかけて毎週1題ずつ再話課題を実施した。社会人の調査協力者を除いた75名は毎週の授業の際に1問ずつ再話課題を行い、社会人の調査協力者7名は2013年5月中旬に一齐に実験を実施した。各受験者はグループ1とグループ2に分けられ、グループ1の受験者はパッセージ1と3をリーディングのインプットで、2と4をリスニングのインプットとして再話を行い、グループ2の受験者はその反対の組み合わせで再話を行った。今回はそれぞれのパッセージやインプットモードの影響は考慮しなかった。

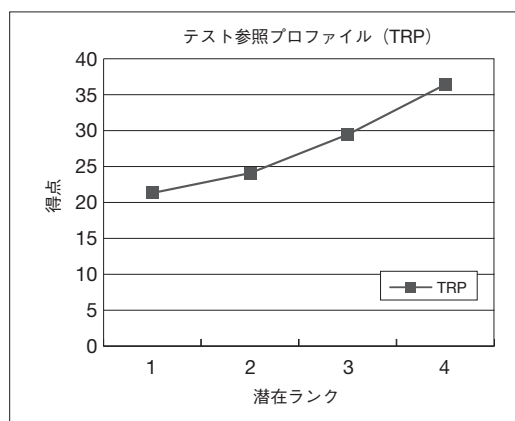
4.5 分析

初めに再話課題のパフォーマンスをSOMによって潜在ランクに分割した。続いて、クロス集計によって英検Can-doリストを使用した自己評価と再

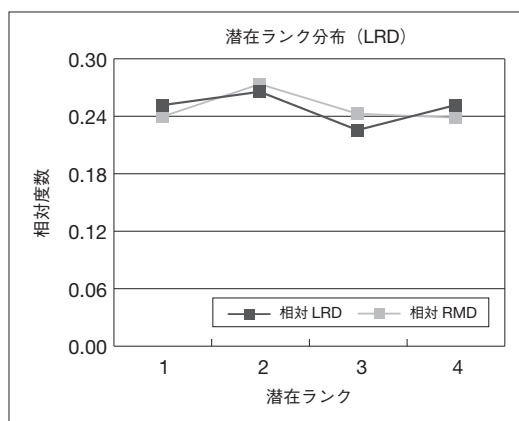
話課題、CEFR-Jを使用した自己評価と再話課題の潜在ランクを比較した。さらに、英検Can-doリストの準1級、CEFR-J B1.2の再話に関する記述文への回答とパフォーマンスの潜在ランクの関係についてクロス集計を用い、比較した。

4.6 調査2の結果

再話課題の得点をもとに、SOMによって4段階の潜在ランクに分割した。図5は再話課題のTRP、図6はLRDとなる。図6のLRDを見る限り、微かな差はあるものの、各ランクに約25%ずつ受験者が割り振られている。



▶ 図 5：再話課題のTRP



▶ 図 6：再話課題の LRD

続いて、自己評価の潜在ランクと再話課題の潜在ランクをクロス集計によって比較した。その結果、英検、CEFR-J いずれもかなりのばらつきがあり、自己評価と実際のパフォーマンスに関連があると言いつける結果とはならなかった。

次に、自己評価の実際の回答と再話課題の潜在ランクについても同様にクロス集計を行った。その結果、英検 Can-do リストを使用した自己評価、CEFR-J を使用した自己評価のどちらも得点が分散しており、自己評価は有効ではなかったと思われる結果となった。

次に、英検 Can-do リスト準 1 級「読んだ本や見た映画について、そのあらすじを述べることができ

■ 表 8：英検 Can-do リストによる自己評価と再話課題の潜在ランククロス集計

英検	再話					
		1	2	3	4	total
	1	6	5	6	2	19
	2	5	6	2	4	17
	3	7	6	4	5	22
	4	1	3	5	8	17
	total	19	20	17	19	75

(注) χ^2 検定 $p = .256$

■ 表 9：CEFR-J による自己評価と再話課題の潜在ランククロス集計

CEFR-J	再話					
		1	2	3	4	total
	1	4	6	4	2	16
	2	5	7	8	4	24
	3	7	4	0	5	16
	4	3	3	5	8	19
	total	19	20	17	19	75

(注) χ^2 検定 $p = .133$

る」と、CEFR-J B1.2「短い読み物か短い新聞記事であれば、ある程度の流暢さをもって、自分の感想や考えを加えながら、あらすじや要点を順序だてて伝えることができる」についての回答（「4」＝自信がある、「3」＝やや自信がある、「2」＝あまり自信がない、「1」＝自信がない）と再話課題の潜在ランクをクロス集計で比較してみたところ、どちらのケースでも的確に自己評価ができているとは言えない結果となった。

■ 表 10：英検 Can-do リストによる自己評価（実測値）と再話課題の潜在ランクのクロス集計

再話	英検					
		1	2	3	4	total
	1	7	8	4	0	19
	2	5	10	2	3	20
	3	5	9	2	1	17
	4	3	5	6	5	19
	total	20	32	14	9	75

(注) χ^2 検定 $p = .176$

■ 表 11：CEFR-J による自己評価（実測値）と再話課題の潜在ランクのクロス集計

再話	CEFR-J					
		1	2	3	4	total
	1	6	10	3	0	19
	2	8	10	1	1	20
	3	3	12	2	0	17
	4	2	9	3	5	19
	total	19	41	9	6	75

(注) χ^2 検定 $p = .042$

4.7 調査 2 のまとめ

調査 2 では英検 Can-do リストと CEFR-J を使用した自己評価の潜在ランクと再話課題のパフォーマンスの潜在ランクを比較し、さらに再話に関する記述文（英検 Can-do 準 1 級、CEFR-J B1.2）への自己評価回答とパフォーマンスの関係を、クロス集計を使用して調査した。その結果、自己評価の潜在ランクと再話課題の潜在ランクは一致しなかった。また、自己評価の回答と再話課題の潜在ランクの関係も一貫性の乏しいものとなった。この結果は、今回の調査に参加された方の多くが大学生が高専生であったため、再話課題への不慣れが自己評価とパフォーマンスの不一致の原因の 1 つであると思われる。

5

結論と今後の課題

学習者が「どのようなことであれば英語を使って話せるか」ということを試験実施前に質問し、その回答に応じて学習者に与える試験課題を変えることでより適切なパフォーマンスを引き出すことができるようになると、スピーキングテスト実施の負担を軽減することになると考え、本調査を実施した。英検 Can-do リストを利用した自己評価と CEFR-J を利用した自己評価の関係を調査し、また、再話課題のパフォーマンスと自己評価の関係についての調査を行った。その結果、英検 Can-do リストを使用した自己評価と CEFR-J を使用した自己評価では同じような傾向が見られ、お互いに関連性が強かったことが明らかになった。一方、再話課題への回答によって得られたスピーキングパフォーマンスと自己評価の関係は薄く、自己評価によってパフォーマンスを推定することは難しいことがわかった。自己評価には年齢、性別、学習期間、情意要因などが影響を与えていることは明らかであり (e.g. MacIntyre, Noels, & Clément, 1997; Underhill, 1987)、これらの要因が今回の調査結果にも影響を与えたものと思われる。

今回の調査では、スピーキングパフォーマンスの評価については評価者による評価のみを分析の対象とし、流暢さや文法・語彙使用の正確さ、文法・語彙使用の複雑さといった観点での調査は行わなかった。この観点を組み込んだ調査を行うことで、自己評価の高い学生が流暢さや正確さ、複雑さの観点で、どの要因を重視しているかを推定することが

可能となると思われるため、今後、この観点についても調査が必要である。このような研究がなされることによって、それぞれの記述文の妥当性の向上に寄与できるものと考えられる。また、今回は再話課題で得られたスピーキングパフォーマンスのみを分析に使用したが、英検 Can-do リストや CEFR-J の記述に含まれるさまざまなタスクを使用し (例. 道案内、特定の話題を扱ったスピーチ、ロールプレイなど)、そのパフォーマンスと自己評価の結果を比較する必要がある。文部科学省初等中等教育局 (2013) は「各中・高等学校が学習指導要領に基づき、生徒に求められる英語力を達成するための学習到達目標を CAN-DO リストの形で具体的に設定すること」について提言がなされたことを踏まえ、「各中・高等学校の外国語教育における『CAN-DO リスト』の形で学習到達目標設定のための手引き」を公表した。これによって今後能力記述文を作成し、検討する機会が増えることが予想されるため、自己評価と実際のパフォーマンスの関係についての調査を重ね、能力記述文の妥当性の検証を重ねることが、今後重要になるだろう。

謝 辞

今回、この研究の機会をくださった公益財団法人日本英語検定協会と選考委員の先生方、特にアドバイスをいただいた和田稔先生に、心より御礼申し上げます。また、実験計画書の作成、実験の実施に協力くださった筑波大学大学院の平井明代先生、採点に協力していただいた皆様、そして何より実験に参加してくださった皆様に厚く御礼申し上げます。

参考文献 (*は引用文献)

- * Dunlea, J. (n.d.). 「英検と CEFR との関連性について
研究プロジェクト報告」. Retrieved from https://www.eiken.or.jp/eiken/group/result/pdf/report_02.pdf
- * Fulcher, G. (2003). Testing second language speaking. Edinburgh Gate, England: Pearson Education.
- Fulcher, G., & Reiter, M. R. (2003). Task difficulty in speaking tests. *Language Testing*, 20 (3), 321-344.
- * Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, 6, 151-167.
- * Hirai, A., & Koizumi, R. (2013). Validation of empirically derived rating scales for a story retelling speaking test. *Language Assessment Quarterly*, 10:4, 398-422.
- * 今井裕之・吉田達弘 (編著). (2007). 『HOPE 中高生のための英語スピーキングテスト』. 東京：教育出版.
- * 石川祥一・中村優治・杉森直樹・伊藤素子・シュナイダーデニス・金子恵美子. (2010). 『英語スピーキング能力の測定と評価法の研究』平成 18～21 年度文部科学省科学研究費補助金 基盤研究 (A) 研究成果報告書.
- * 木村哲夫. (2012). 「能力記述文による自己評価」. 日本言語テスト学会 (JLTA) 第 16 回全国研究大会発表 (2012 年 10 月 27 日；於専修大学).
- * 木村哲夫. (2013). 「潜在ランク理論を用いたコンピュータ適応型テストのためのアルゴリズムの提案と実装」. 早稲田大学審査学位論文.
- * Koizumi, R., & Hirai, A. (2010). Exploring the quality of the Story Retelling Speaking Test: Roles of story length, comprehension questions, keywords, and opinions. *Annual Review of English Language Education in Japan (ARELE)*, 21, 211-220.
- * 小泉利恵・飯村英樹. (2010). 「ニューラルテスト理論の特徴：古典的テスト理論・ラッシュモデリングとの比較から」. *JLTA (Japan Language Testing Association) Journal*, 13, 91-109.
- * 公益財団法人日本英語検定協会 (n.d.). 英検 Can-do リスト. Retrieved from <http://www.eiken.or.jp/eiken/exam/cando/>
- * 小山由紀江・木村哲夫. (2011). 「Neural Test Theory を使った Can-do Statements の分析」. 統計数理研究所共同研究レポート, 254, 59-77.
- * MacIntyre, P.D., Noels, K.A., & Clément, R. (1997). Biases in Self-Ratings of Second Language Proficiency: The Role of Language Anxiety. *Language Learning*, 47, 265-287.
- * 文部科学省初等中等教育局. (2013). 『各中・高等学校の外国語教育における「CAN-DO リスト」の形での学習到達目標設定のための手引き』. Retrieved from http://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2013/05/08/1332306_4.pdf
- * 中川知佳子. (2012). 「リーディングテストと他技能の関連」. 卯城祐司 (編著). 『英語リーディングテスト考え方と作り方』. 東京：研究社.
- * Shojima, K. (2007a). Neural test theory. *DNC Research Note*, 07-02. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/jpaper.htm>
- * Shojima, K. (2007b). The graded neural test model: A neural test model for ordered polytomous data. *DNC Research Note*, 07-03. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/jpaper.htm>
- * Shojima, K. (2008a). Neural test theory: A latent rank theory for analyzing test data. *DNC Research Note*, 08-01. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/jpaper.htm>
- Shojima, K. (2008b). The batch-type neural test model: A latent rank model with the mechanism of generative topographic mapping. *DNC Research Note*, 08-06. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/jpaper.htm>
- * 荘島宏二郎. (2010). 「ニューラルテスト理論—学力を段階評価するための潜在ランク理論」. 植野真臣・荘島宏二郎 (編著), 『学習評価の新潮流』. 東京：朝倉書店.
- 荘島宏二郎. (2014). Exametrika (Version 5.3) [Computer software]. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/exmk/jindex.htm>
- * 荘島宏二郎. (n.d.). 「潜在ランク理論 (ニューラルテスト理論)」. Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/jindex.htm>
- * 投野由紀夫 (編). (2013). 『英語到達度指標 CEFR-J ガイドブック』. 東京：大修館書店.
- * Underhill, N. (1987). *Testing Spoken Language: A handbook of oral testing techniques*, New York: Cambridge University Press.
- * 臼田悦之. (2009). 「英検 Can-do リストのスピーキング分野における Can-do 項目の妥当性検証」. *STEP BULLETIN*, vol.21, 262-273.

資 料

資料1：評価アンケート質問用紙

問1. 以下の質問について、別紙解答用紙に「4＝自信がある」・「3＝やや自信がある」・「2＝あまり自信がない」・「1＝自信がない」として回答してください。

	質問
質問1	自分の好きなことについて、短い話をするができる。(趣味、クラブ活動など)
質問2	物ごとの「好き」「嫌い」とその理由を簡単に述べるができる。(動物、食べ物、スポーツなど)
質問3	日常生活の行動について言うことができる。(例. I got up at seven. / I ate some bread for breakfast.)
質問4	自分の予定を簡単に言うことができる。(例. I'm going to meet my friends.)
質問5	簡単な頼み事をするができる。(例. Can you open the window, please?)
質問6	身近なことで相手を誘うことができる。(例. Let's go to a movie tonight.)
質問7	簡単な相づちを打つことができる。(例. I see. / Really?)
質問8	興味・関心のあることについて、自分の考えを述べるができる。(好きなスポーツ、趣味に関する事など)
質問9	自分の将来の夢や希望について、話すことができる。(訪れたい国、やりたい仕事など)
質問10	自分の気持ちを表現することができる。(うれしい、悲しい、さびしいなど)
質問11	簡単な約束をすることができる。(会う場所や時間など)
質問12	ファストフード・レストランでメニューを見ながら注文することができる。(食べ物、飲み物、サイズなど)
質問13	電話で簡単な表現や決まり文句を使って応答することができる。(例: Please wait a moment. / Hold on. / Speaking.)
質問14	日常生活の身近な状況を説明することができる。(遅刻や欠席の理由など)
質問15	印象に残った出来事について、話すことができる。(旅行、イベントなど)
質問16	自分の学校(会社)について、簡単な紹介をすることができる。(場所、人数、特徴など)
質問17	簡単な道案内をすることができる。(例: Go straight and turn left at the next corner.)
質問18	買い物で店員に欲しいものや好みを伝えたり、簡単な質問をすることができる。(色、サイズ、値段など)
質問19	簡単な伝言をすることができる。(例: Tell Jane to call me back. / Tell John I can't go to the meeting today.)
質問20	調べたことについて、まとまりのある話をするができる。(課題の発表、仕事のプレゼンテーションなど)
質問21	自分の仕事や専門分野に関する講義や発表などを聞いて、それについて質問したり自分の考えを述べたりすることができる。
質問22	商品やサービスについて、苦情を言うことができる。(商品の故障、サービスの内容など)
質問23	公共の施設で簡単な用を足すことができる。(郵便局で手紙を出す、図書館で本を借りるなど)
質問24	病院などで健康状態を伝えることができる。
質問25	簡単な内容であれば、電話で用を足すことができる。(歯医者や美容院の予約など)
質問26	読んだ本や見た映画について、そのあらすじを述べるができる。
質問27	簡単な語や基礎的な句を用いて、自分についてのごく限られた情報(名前、年齢など)を伝えることができる。
質問28	前もって話すことを用意した上で、基礎的な語句、定型表現を用いて、人前で実物などを見せながらその物を説明することができる。
質問29	基礎的な語句、定型表現を用いて、限られた個人情報(家族や趣味など)を伝えることができる。
質問30	基礎的な語句、定型表現を用いて、簡単な情報(時間や日時、場所など)を伝えることができる。
質問31	前もって発話することを用意した上で、限られた身近なトピックについて、簡単な語や基礎的な句を限られた構文を用い、簡単な意見を言うことができる。
質問32	前もって発話することを用意した上で、日常生活の物事を、簡単な語や基礎的な句を限られた構文を用い、簡単に描写することができる。
質問33	前もって発話することを用意した上で、限られた身近なトピックについて、簡単な語や基礎的な句を限られた構文に用い、複数の文で意見を言うことができる。

質問 34	前もって発話することを用意した上で、日常生活に関する簡単な事実を、簡単な語や基礎的な句を限られた構文を用い、複数の文で描写できる。
質問 35	一連の簡単な語句や文を使って、自分の趣味や特技に触れながら自己紹介をすることができる。
質問 36	写真や絵、地図などの視覚的補助を利用しながら、一連の簡単な句や文を使って、身近なトピック（学校や地域など）について短い話をするができる。
質問 37	写真や絵、地図などの視覚的補助を利用しながら、一連の簡単な語句や文を使って、自分の毎日の生活に直接関連のあるトピック（自分のこと、学校のこと、地域のことなど）について、短いスピーチをすることができる。
質問 38	一連の簡単な語句や文を使って、意見や行動計画を、理由を挙げて短く述べるができる。
質問 39	使える語句や表現をつないで、自分の経験や夢、希望を順序だて、話を広げながら、ある程度詳しく語ることができる。
質問 40	自分の考えを事前に準備して、メモの助けがあれば、聞き手を混乱させないように、馴染みのあるトピックや自分に関心のある事柄について語ることができる。
質問 41	短い読み物が短い新聞記事であれば、ある程度の流暢さをもって、自分の感想や考えを加えながら、あらすじや要点を順序だてて伝えることができる。
質問 42	自分の関心事であれば、社会の状況（ただし自分の関心事）について、自分の意見を加えてある程度すらすらと発表し、聴衆から質問が出れば相手に理解できるように答えることができる。
質問 43	ある視点に賛成または反対の理由や代替案などをあげて、事前に用意されたプレゼンテーションを聴衆の前で流暢に行うことができ、一連の質問にもある程度流暢に対応ができる。
質問 44	ディベートなどで、そのトピックが関心のある分野のものであれば、論拠を並べ自分の主張を明確に述べることができる。
質問 45	要点とそれに関連する詳細の両方に焦点を当てながら、流暢にプレゼンテーションができ、また、あらかじめ用意されたテキストから自然に離れて、聴衆が興味のある点に対応してプレゼンテーションの内容を調整し、そこでもかなり流暢に容易に表現できる。
質問 46	ディベートなどで、社会問題や時事問題に関して、補助的観点や関連事例を詳細に加えながら、自分の視点を明確に展開することができ、話を続けることができる。
質問 47	複雑なトピックを、派生的問題にも立ち入って、詳しく論ずることができ、一定の観点を展開しながら、適切な結論でまとめ上げることができる。
質問 48	状況に合った文体で、はっきりと流暢に記述・論述ができる。効果的な論理構成によって聞き手に重要点を把握させ、記憶にとどめさせることができる。

資料 2：再話課題の例

〈本文省略〉

＜次頁＞

Retell as much of the story as you can in English in two and half minutes. You can look at the keywords while you are retelling. At the end of your retelling, be sure to include your opinions about the story. You will hear a signal 30 seconds before closing.

今読んだ内容をできるだけ詳しく、2分30秒間英語で話してください。話しながら、キーワードを見てもかまいません。読んだ内容を話し終えたら、必ず、その内容についての感想や意見を英語で述べてください。終了30秒前に合図をします。感想を始める目安にできます。話す前に、名前と学籍番号を言うてから話してください。

Keywords:

load, nutrient, transportation, supermarket