

ケンブリッジ英語検定の導入と そのスピーキング活動への応用

－新学習指導要領「話すこと [やりとり]」の指導と評価の試み－

英語科 金子麻子

1. はじめに

学習指導要領（平成30年）では、外国語の学習において重要視されるべきは知識の習得ではなく、知識・技能が実際のコミュニケーションにおいて活用されるなかで思考・判断・表現することを通して育成されることである、としている。このため、英語の授業においては、コミュニケーション活動がその中核を成すことになるのだが、コミュニケーション活動の評価に関してはその問題点を指摘する声も多く（小泉，2022；山本，近藤，鳴海，多田，吉田，2021；米崎，2016），文部科学省（2020）の調査によると、公立学校（普通科）の36.7%は年間を通してスピーキングテスト・ライティングテストなどのパフォーマンステストを実施していない。また、授業で生徒が行うコミュニケーション活動としては、生徒同士のやりとりがその大半を占めるが、その評価に関しては、生徒同士のやりとりのパフォーマンスではなく、教員と生徒のやりとりを評価することが多い点を懸念する声がある（小泉，2015；山本，他，2021）ものの、生徒同士のやりとりを評価しようという試みはいまだ少ない。日本で受験可能な主な英語検定試験においては、受験者と面接官という立場の異なる二者のやりとりを評価する試験は複数あるが、受験者同士のやりとりを評価の対象に含めているのはケンブリッジ英語検定（以下、ケンブリッジ英検）のみである。このような状況のなか、本校では2021年度よりケンブリッジ英検を導入し、校内で一斉実施をしている（2021年度は1・2年生のみ実施）。本稿では、ケンブリッジ英検の導入にあたり、その準備のために授業で取り組んだスピーキング活動とスピーキングテストを紹介し、新学習指導要領を見据えたその評価の方法を検証する。

2. ケンブリッジ英語検定

ケンブリッジ英検は、イギリスのケンブリッジ大学英語検定機構（Cambridge Assessment English）が開発し、130カ国以上で実施されている世界標準の英語検定試験である。5つの級があり、各級はCEFR（Common European Framework of Reference for Languages；ヨーロッパ言語共通参照枠）に基づく。CEFRは「外国語の運用能力を同一の基準で測ることが出来る国際標準」（British Council）であり、「目標言語で何が出来るか」を基準に構成された国際的な言語運用力の指標である（各レベルの運用能力に関しては表1を参照）。ケンブリッジ英検はリーディング、リスニング、ライティング、スピーキングの四技能を個別に測定する試験となっているが、本稿では、そのなかでもスピーキングに関するテスト（以後、「スピーキングテスト」）を取り扱う。

表 1. CEFR のレベルと熟達度 (British Council ホームページより)

段階	CEFR レベル	熟達度 (対象言語で「何ができるか」)	実用英語技能検定
熟達した言語使用者	C2	聞いたり読んだりした、ほぼ全てのものを容易に理解することができる。いろいろな話し言葉や書き言葉から得た情報をまとめ、根拠も論点も一貫した方法で再構築できる。自然に、流暢かつ正確に自己表現ができる。	
	C1	いろいろな種類の高度な内容のかなり長い文章を理解して、含意を把握できる。言葉を探しているという印象を与えずに、流暢に、また自然に自己表現ができる。社会生活を営むため、また学問上や職業上の目的で、言葉を柔軟かつ効果的に用いることができる。複雑な話題について明確で、しっかりとした構成の詳細な文章を作ることができる。	1 級
自立した言語使用者	B2	自分の専門分野の技術的な議論も含めて、抽象的な話題でも具体的な話題でも、複雑な文章の主要な内容を理解できる。母語話者とはお互いに緊張しないで普通にやり取りができるくらい流暢かつ自然である。幅広い話題について、明確で詳細な文章を作ることができる。	準 1-1 級
	B1	仕事、学校、娯楽などで普段出会うような身近な話題について、標準的な話し方であれば、主要な点を理解できる。その言葉が話されている地域にいるときに起こりそうな、たいいていの事態に対処することができる。身近な話題や個人的に関心のある話題について、筋の通った簡単な文章を作ることができる。	2- 準 1 級
基礎段階の言語使用者	A2	ごく基本的な個人情報や家族情報、買い物、地元の地理、仕事など、直接的関係がある領域に関しては、文やよく使われる表現が理解できる。簡単で日常的な範囲なら、身近で日常の事柄について、単純で直接的な情報交換に応じることができる。	準 2-2 級
	A1	具体的な欲求を満足させるための、よく使われる日常的表現と基本的な言い回しは理解し、用いることができる。自分や他人を紹介することができ、住んでいるところや、誰と知り合いであるか、持ち物などの個人的情報について、質問をしたり、答えたりすることができる。もし、相手がゆっくり、はっきりと話して、助けが得られるならば、簡単なやり取りをすることができる。	3- 準 2 級

※最右列は参考として、文部科学省 (2015) の対照表 (資料 1) を使用して実用英語技能検定の該当級を掲載した。

2.1. ケンブリッジ英検スピーキングテスト

ケンブリッジ英検のスピーキングテストは、試験官 2 人と受験者 2 人の対面式で行われ、所要時間 12 ～ 14 分のテストである。試験官 2 人のうち、1 人は受験者と会話をする面接官 (interlocuter) であり、もう 1 人は会話には介入せずに評価のみを行う評価官 (assessor) である。表 2 は、日本で受験可能な主な英語検定試験のスピーキングテストの形式を比較したものだが (各種スピーキングテストの比較は、小泉, 2021 や、村岡, 2019 などに詳細がある)、対面式のテストを実施しているのはケンブリッジ英検、実用英語検定試験、IELTS、TEAP のみであり、他のテストはコンピューター等の機器に録音する形式のスピーキングテストを実施している。録音式と対面式のス

スピーキングテストを比較した場合、対面式の方が受験者の発話は流暢であり (Zhou, 2008), 受験者は情意的には対面式を好む傾向にある (米崎, 2016) という報告がされており, 対面式は実際のコミュニケーションにより近い形態である (鈴木, 2020) ことから, 対面式のスピーキングテストは望ましい形式であると考えられる。

表 2. 主な英語検定試験のスピーキングテストの形式 (各団体の HP より)

	対面式	PC 機器などへの録音式		面接官・受験者	所要時間
ケンブリッジ 英語検定	●	—		2 対 2	12 ~ 14 分
実用英語検定 試験 1 ~ 3 級	●	● (S-CBT)	1 級 準 1 級 2 級 準 2 級 3 級	2 対 1 1 対 1 同 同 同	10 分 8 分 7 分 6 分 5 分
同 4 ~ 5 級	—	●	4・5 級	—	5 分
GTEC (中学・ 高校向け)	—	●		—	25 分
IELTS	●	—		1 対 1	11 ~ 14 分
TEAP	●	● (TEAP CBT)		1 対 1	10 分
TOEFL	—	●		—	17 分
TOEIC	—	●		—	20 分

ケンブリッジ英検のスピーキングテストは 4 パートから構成されており (表 3), なかでもパート 3 および 4 の「受験者同士のやりとり」が特徴的である。表 2 の試験のうち, 受験者同士のやりとりを採用しているのはケンブリッジ英検のみである。本校の授業では生徒同士による英語のやりとりが重視されており, その学習成果を測定するには同テストが適していると考えたことが, 本校におけるケンブリッジ英検導入の要因の一つとなった。Hughes (2003) は, 生徒同士がペアになりやりとりを行う会話テストの方が, 面接官との対話式よりも受験者が不安を感じにくいと論じており, 白鳥 (2018) もケンブリッジ英検のスピーキングテストの構成や形式は高校生にとって馴染みのあるものであるがゆえに応答しやすいと評価している。その一方で, 山本, 他 (2021) は, 「やりとり」のパフォーマンスを限られた時間で評価する方法が確立されていないという高校の英語教育の現状を指摘している。

表 3. ケンブリッジ英検スピーキングテストの構成

パート 1	身近な事柄に関する質問への応答
パート 2	写真描写
パート 3	イラストを使用し、受験者同士がやりとりして与えられたタスクを遂行する
パート 4	パート 3 のトピックに関して、受験者が回答（またはディスカッション）する

2.2. ケンブリッジ英検スピーキングテストの信頼性

テストにおける信頼性を示す指標の一つに、信頼性係数がある。これは同一のテストを同じ受験者が複数回受験した場合に、その結果が一致する程度を指し、0 から 1 の数値で表される。Hughes (2003) は、スピーキングテストの信頼性係数に関しては、目安として .85 という数値ならば十分であると考えており、ケンブリッジ英検スピーキングテストの信頼性係数は表 4 に示した通り、級によって異なるが、.84 から .87 と、十分な信頼性が確保されている (Cambridge Assessment English, 2020a)。

表 4. ケンブリッジ英検スピーキングテスト各レベルの信頼性係数

各レベル	信頼性係数
KET (A2)	.87
KET for Schools (A2)	.86
PET (B1)	.84
PET for Schools (B1)	.85
FCE First (B2)	.84
CAE (C1)	.82
CPE (C2)	.85

2.3. 日本人高校生英語学習者の英語運用能力

2015 年 6 月に文部科学省が実施した全国の高等学校および中等教育学校後期課程の第 3 学年 8.1 万人（「話すこと」は 1.8 万人）を対象とした英語力調査（文部科学省, 2016）によると、全国の国立・公立学校で学ぶ生徒の英語運用能力は、「話すこと」に関しては最も下のレベルである A1 レベルが 87.2% を占めており、A2 は 11%、B1 は 1.8% であり、特に発信型技能に課題が大きい（白鳥, 2018）。一方で、文部科学省 (2013) は、第二期教育振興基本計画において、生徒が目指す高等学校卒業時の英語力の目標を、「英検準 2 級から 2 級程度 (CEFR の A2 ~ B1 程度) 以上を達成する生徒の割合が 50%」としている。現状に対して非常に高い目標を定めていると言えるが、本校生徒の場合は 3 年生時点で GTEC のスピーキングスコアの平均が B1.1 (CEFR-J) に達していることから、B1 レベルは適切な目標と言える。

3. 本校におけるケンブリッジ英検

3.1. 本校におけるケンブリッジ英検の実施

2021年11月に、本校1年生117名、2年生117名がケンブリッジ英検 B1 for Schools (PET) を受験した。B1 for Schools とは、B1 レベルのなかでも特に学校で学ぶ生徒を対象としたタイプのテストであり、テスト内で扱われる語彙や場面設定が学校生活を中心とするものとなっており、生徒にとってより馴染みのある内容になっている。試験は本校で実施され、ケンブリッジ英検が派遣する試験官によるスピーキングテストも同日に行われた。

3.2. 本校生徒のケンブリッジ英検の結果

本校生徒受験者のスピーキングテスト(170点満点)の平均スコアは1年生154.3、2年生155.8となっており、これらのスコアはいずれもB1レベルを表す。四技能を統合した全体スコアで見ると、1年生は受験者全体の76.1%、2年生は86.3%がB1レベルを合格した。

4. 授業での取り組み

本セクションでは、上記ケンブリッジ英検受験に向けて、授業で行った活動を紹介する。ケンブリッジ英検は表2の英語検定試験のなかでは受験者やテストセンターの数も数も少なく(2022年3月現在、国内8ヶ所)、本校生徒の大部分が未受験であるため、事前に準備が必要であると考え、授業でその形式を紹介し練習を行った。授業ではリーディング、リスニング、ライティング、スピーキングの四技能に関して準備をしたが、本稿においてはそのうち、2年生の英語表現IIの授業(英語母語話者の講師とのチームティーチングを行っている)でのスピーキング活動の取り組みのみを取り扱う。

前述の通り(表3)、スピーキングテストは4つのパートに分かれている。授業では、まず生徒に同テストのデモ映像を見せ、その全体像と評価方法を説明した。その後、各パート1時間計4時間を使用して、各パートの出題形式に沿った活動を行った。生徒は主に2人1組のペアになり、質問作成・応答・相互フィードバックを行うなど、主体的に活動に取り組んだ。

4.1. パート1：身近な事柄に関する質問への応答

最初のパートは、受験者が英語での発話に慣れるための導入部分であり、面接官が受験者一人ずつに身近な話題に関する質問をする。質問の一例は、“What do you study?” “Who do you live with?” “Tell us about the teacher you like.” など、受験者が答えやすい質問である。授業では、デモテストの映像を再度視聴し、質問および答え方を確認したあと、まずは教員が用意した質問を用いてペアで質問・応答の練習を行った。その後、回答の仕方に関して生徒同士でフィードバックを与え合い、文法的な誤りの指摘やさらに詳細な回答をするための助言を与え合った。続いて、他にどのよう

な質問が想定されるかを考え、自分たちで考えた質問をペアで聞き合い、同様のフィードバックを行った。



図1. 授業中の活動の様子

4.2. パート2：写真描写

次のパートは、受験者一人ずつにそれぞれ異なる写真が提示され、受験者はその写真を描写するように指示される。その後、写真に関する2,3の質問が続く。授業では、まずはこのパートのデモテストで使用された写真を生徒に描写させた後、デモテストで受験者が回答する映像を見せた。デモテストの受験者は高得点を得られる回答をしているため、その回答のどのような点が優れているのかを生徒達に分析させた。回答の際に気をつけるべき点などをまとめた後、教員が用意した写真数点に関して、描写の練習をした。この際、生徒はペアとなり、①写真描写をする役、②描写した生徒の英語に関してフィードバックを与える役、の二つの役割を交互に果たした。

4.3. パート3：協同タスク

パート3は、collaborative task と呼ばれ、受験者同士がやりとりしながら与えられたタスクを遂行することを目指す。受験者には図2のようなプロンプトが提示され、面接官が状況設定を説明したのち、タスクが与えられる。図2に関しては、次のような設定およびタスクが与えられている。“Some students from a small village school are going on a trip to the capital city. Here are some activities they could do there. Talk together about the different activities they could do in their capital city, and say which would be most interesting. 「農村部にある小規模校が、首都への遠足を企画しています。

A school visit to a capital city



図2. ケンブリッジ英検パート3プロンプト

このイラストは、遠足で可能なアクティビティを表わしています。これらのアクティビティについて二人で話し合い、どれが最も面白いかを教えてください。」(著者による和訳)

授業では、同パートのデモテスト映像を視聴した後、受験者の発話の SCRIPT を分析し、どのような点に注意すべきかをグループで話し合った。教員は

注意点や採点の際のポイントを提示し、利用可能な英語表現をリストにして生徒に与えた(資料2)。その後、いくつかのプロンプトに関してペアで練習を行った。その際、2つのペアを統合して4人一組となり、2人は協同タスクを行い、2人はそのタスクを見ながらメモをとり、のちに改善点や文法的誤りに関するフィードバックを行った。

4.4. パート4：ディスカッション

最後のパートでは、パート3で取り扱ったトピックに関する質問が聞かれた後、受験者がお互いの意見を述べ合うディスカッションが行われる。面接官は、まずパート3の内容に関する簡単な質問を受験者一人ずつに問いかけ、その後やや発展的な質問に関して、“Discuss together.”と指示をする。受験者は2人で考えを述べ合う。4.3で紹介したプロンプトの場合は、“Would you like to have more school trips?” “What do you need to take when you go on a trip?”といった質問の後、“Which do you think is more interesting, visiting a country side or visiting a city?”などのように、パート3と関連あるトピックに関する意見を求められる。ここでは受験者同士のやりとりが評価の対象となっているので、受験者は一方的に自分の意見を述べるのではなく、相手の意見に対して賛成や反対といった自分の立場を表明し、その理由や根拠を述べる必要がある。先に自分の意見を述べた場合でも、相手の意見に相槌を打ったり、同意するかしないかおよびその理由を述べるなど、相手の意見に対する自分の考えを表わす必要がある。その際、アイコンタクトや相槌など、非言語的なコミュニケーションの要素も評価の対象となる。授業では、これらの点に注意しながら、パート3で用意したプロンプトに関して、どのような発展的な質問が想定されるかを生徒に考えさせ、2つのペアを合体させ、①②受験者、③面接官、④フィードバックを与える人の役割を交互に行った。

5. パフォーマンステスト（授業でのスピーキングテスト）とその評価

上記4の活動後、パート3および4に関して、授業内でパフォーマンステストを行った。その際、新学習指導要領(文部科学省, 2018)の「話すこと[やり取り]」における目標及び評価規準を、ケンブリッジ英語検定で使用される分析的評価項目(analytical assessment scales)に対応させて評価した。また、全活動後終了後に生徒に対して質問紙調査を実施し、生徒の反応を調査した。

5.1. スピーキングテストの内容

本テストは、上述の通り、ケンブリッジ英検のスピーキングテストの形式を踏襲しながら、そのパート3および4のみを実施した。問題およびプロンプトは、ケンブリッジ英検のデモテストのものを参考に著者が新たに作成した5問を、無作為に出題した。図3は、プロンプトの一例である。このプロンプトに関して、面接官が以下のような指示を与えた。



図3. パート3 プロンプトの一例

“First, you are going to talk about something together for 2 minutes. I’m going to describe you the situation. The school festival is coming. Your class needs to decide what they will do for the school festival. Here are some things your class can do for the school festival. Talk together about the different things your class could do for the festival and say which one is the best.”

本校の生徒達はこの約1カ月前に文化祭を終えており、上記のトピックの様な活動は実際に自ら体験しており、身近で話し易いトピックであると言えるだろう。

パート4では、これに関連した質問(“Schools in UK or US don’t have events like ‘school festival.’ Do you think a school festival is a meaningful event for high school students? Why (not) ?”)に関して、生徒同士が意見を交換した。

5.2. スピーキングテストの実施

上記テストを、試験官1人受験者2人の対面式で実施した。本来ならば、試験官は2人(面接官1人、評価官1人)であるが、十分な数の教員の確保が困難であるために、面接官が評価も同時に行うとともに、テストをビデオ録画した。著者と英語母語話者の講師(両者とも授業者)が面接官となり、クラスを二つに分け、同時進行で別室でテストを実施した。所要時間はそれぞれのパートに約2分で、合計4分間だった。テストは2021年10月26日に本校2階面談室および3階書道室にて実施した(図4)。当日は欠席者が4人いたが、後日追試を受け、最終的には2年生118名全員が受験した。

受験者は本テストの形式と同様の活動を授業内で行っていたため、スムーズにタスクを遂行した。受験者同士のやりとりに関しても、アイコンタクトやジェスチャーを交えながら、相手の意見に対して自分の意見を述べるなど、適切なインタラクションが行われていた。だが、ペアによっては、積極的な受験者がより多く発話し、消極的な受験者は聞き手にまわるなど、発話の量に差がでてしまう組み合わせもあった点が

課題として挙げられた。また、4分という時間は十分ではなく、受験者の発話が終わらない間に終了を知らせるベルがなることも少なくなかった。



図 4. 本校でのスピーキングテストの様子

5.3. 評価

5.3.1. 新学習指導要領「話すこと [やりとり]」の評価

学習指導要領（平成 30 年）は、(1) 学習評価の充実、(2) カリキュラム・マネジメントの一環としての指導と評価、(3) 主体的・対話的で深い学びの視点からの授業改善と評価の 3 点において改訂されており、評価に関しては「知識・技能」「思考・判断・表現」「主体的に学習に取り組む態度」の三観点の評価へと整理された。この三観点の評価に関しては「内容のまとまり」ごとに評価規準を作成することになっている。高等学校外国語における「内容のまとまり」は、「英語コミュニケーション I」（および同 II, III）においては「聞くこと」「読むこと」「話すこと [やりとり]」「話すこと [発表]」「書くこと」の四技能五領域、「論理・表現 I」（および同 II, III）では「話すこと [やりとり]」「話すこと [発表]」「書くこと」の三領域を指す。本稿で扱うのは、このうち「話すこと [やりとり]」となるが、国立教育政策研究所（2021 年）は、「話すこと [やりとり]」の三観点の評価規準を表 5 の様に提案している。

表 5. 「話すこと [やりとり]」の評価規準（国立教育政策研究所，2021，p. 38）

	知識・技能	思考・判断・表現	主体的に学習に取り組む態度
話すこと [やりとり]	知識：英語の特徴やきまりに関する事項を理解している。 技能：コミュニケーションを行う目的や場面、状況などに応じて、日常的话题や社会的な問題について、情報や考え、気持ちなどを、理論性に注意して伝え合う技能を身に付けている。	コミュニケーションを行う目的や場面、状況などに応じて、日常的话题や社会的な問題について、情報や考え、気持ちなどを、話して伝え合うやり取りを続けたり、論理性に注意して話して伝え合ったりしている。	外国語の背景にある文化に対する理解を深め、聞き手、話し手に配慮しながら、主体的、自律的に英語を用いて伝え合おうとしている。

また、事例（国立教育政策研究所，2021）のなかでは、「論理・表現 I」の「話すこと [やりとり]」に関して、事例とともに表 6 の評価規準を紹介している。

表 6. 事例「論理・表現 I」の評価規準（国立教育政策研究所，2021，p. 68）

知識・技能	思考・判断・表現	主体的に学習に取り組む態度
<ul style="list-style-type: none"> ・ 賛成・反対の意見を，論理の構成や展開を工夫して話して伝え合うために必要となる表現を理解している。 ・ 日常的な話題（余暇の過ごし方）や社会的な話題（海外研修先の選択）について，賛成・反対の意見を論理の構成や展開を工夫して話して伝え合う技能を身に付けている。 	<p>自分の意見を，相手によりよく理解してもらえるように，日常的话题（余暇の過ごし方）や社会的な話題（海外研修先の選択）について，聞いたり読んだりしたことを活用しながら，相手の意見に応じて，賛成・反対の意見を論理の構成や展開を工夫して話して伝え合っている。</p>	<p>自分の意見を，相手によりよく理解してもらえるように，日常的话题（余暇の過ごし方）や社会的な話題（海外研修先の選択）について，聞いたり読んだりしたことを活用しながら，相手の意見に応じて，賛成・反対の意見を論理の構成や展開を工夫して話して伝え合おうとしている。</p>

上記事例において想定されているパフォーマンステストは，生徒同士の実際の会話のやりとりではなく，予め録音されたある生徒（ケン）の意見を聞いて，そのケンの意見に対して自分の考えを述べ，それをタブレットなどの録音機器に録音する形式となっている。表 7 が，そのパフォーマンステストの採点基準である。この事例においては，評価の対象は「思考・判断・表現」「主体的に学習に取り組む態度」の二観点のみとなっている。テストにおいて生徒が表 8 の条件を両方とも満たせば b が与えられる。また，「主体的に学習に取り組む態度」に関しては，「思考・判断・表現」と一体的に評価するとしている。これは，「思考・判断・表現」の能力が「…しようとしている」という態度に結びつけられるという考え方に基づく措置である。

表 7. 事例のパフォーマンステストの採点基準（国立教育政策研究所，2021，p. 73）

評価結果	思考・判断・表現	主体的に学習に取り組む態度
a	二つの条件を満たした上で，意見を，論理に矛盾や飛躍がなく，より適切な理由とともに述べている。	二つの条件を満たした上で，意見を，論理に矛盾や飛躍がなく，より適切な理由とともに述べようとしている
b	二つの条件を満たしている。	二つの条件を満たそうとしている。
c	「b」を満たしていない。	「b」を満たしていない。

表 8. 事例のパフォーマンステストにおいて課せられた条件

条件 1	相手の意見の要点を述べている。
条件 2	相手とは反対の意見を，適切な理由とともに述べている。

5.3.2. ケンブリッジ英検スピーキングテストの評価

ケンブリッジ英検のスピーキングテストの評価は，2人の試験官（面接官および評価官）によって行われる。面接官は，受験者のコミュニケーションを評価する

global achievement scale (表 9) を用いて、全体的な印象から総合的な評価 (global assessment) を行い、評価官は①語彙・文法, ②ディスコースマネジメント, ③発音, ④相互コミュニケーションの 4 点を analytical assessment scale (表 10) を用いて分析的に評価 (analytical assessment) する。どちらのスケールも、最低が 0, 最高が 5 の尺度評価である。

表 9. ケンブリッジ英検総合的評価 (Global Achievement Scale)

B1	Description
5	Handles communication on familiar topics, despite some hesitation. Organises extended discourse but occasionally produces utterances that lack coherence, and some inaccuracies and inappropriate usage occur.
4	Performance shares features of Band 3 and 5.
3	Handles communication in everyday situations, despite hesitation. Constructs longer utterances but is not able to use complex language except in well-rehearsed utterances.
2	Performance shares features of Band 1 and 3.
1	Conveys basic meaning in very familiar everyday situations. Produces utterances which tend to be very short – words or phrases – with frequent hesitation and pauses.
0	Performance below Band 1.

表 10. ケンブリッジ英検分析的評価 (Analytical Assessment Scale)

B1	Grammar and Vocabulary	Discourse Management	Pronunciation	Interactive Communication
5	Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on familiar topics.	Produces extended stretches of language despite some hesitation. Contributions are relevant despite some repetition. Uses a range of cohesive devices.	Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly.	Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.
4	Performance shares features of Bands 3 and 5.			
3	Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about familiar topics.	Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, but there may be some repetition. Uses basic cohesive devices.	Is mostly intelligible, and has some control of phonological features at both utterance and word levels.	Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.
2	Performance shares features of Bands 1 and 3.			
1	Shows sufficient control of simple grammatical forms. Uses a limited range of appropriate vocabulary to talk about familiar topics.	Produces responses which are characterised by short phrases and frequent hesitation. Repeats information or digresses from the topic.	Is mostly intelligible, despite limited control of phonological features.	Maintains simple exchanges, despite some difficulty. Requires prompting and support.
0	Performance below Band 1.			

5.3.3. 学習指導要領とケンブリッジ英検の評価の比較

本セクションでは、学習指導要領の「話すこと [やりとり]」の評価と、ケンブリッジ英検スピーキングテストの分析的評価を比較する。両者の評価はそれぞれ、5.3.1 および 5.3.2 に既述している通りだが、これを観点別に対応させたものが図5である。学習指導要領の観点「知識・技能」は、「きまり」や「表現」の理解および技能を測るものであり、ケンブリッジ英検の Grammar & Vocabulary に対応する。「思考・判断・表現」は、「やり取りを続ける」「論理性に注意して話し合う」などが、ケンブリッジ英検の Discourse Management (会話を適切に続けていくための力) と対応する。「主体的に学習に取り組む態度」に関しては、学習指導要領の評価規準には「聞き手・話し手に配慮しながら」「伝え合おうとしている」とあり、ケンブリッジ英検の Interactive Communication に対応すると考えた。

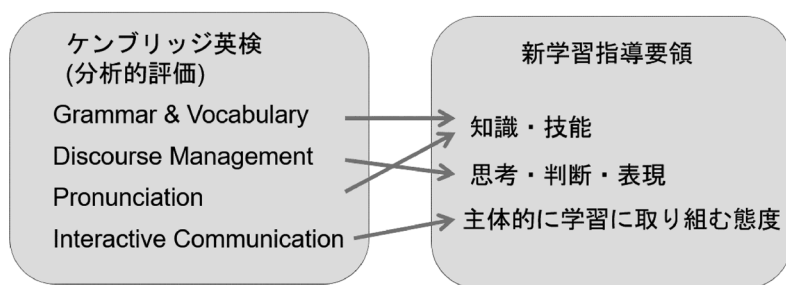


図5. 学習指導要領とケンブリッジ英検スピーキングテストの評価の対応

しかしながら、両者の評価はおおむね各項目には対応関係があるものの、完全に一致するものではない。特に、ケンブリッジ英検は Grammar & Vocabulary の項目において Shows a good degree of control of simple grammatical forms, and attempts some complex grammar forms (単純な文法形式を確実に使用することができ、複雑な文法形式を使用しようという試みがみられる) の様に、受験者が使用する文法項目のレベルに言及している。このため、例えば、現在形の「主語＋述語」から成る単純な文 (I live with my parents.) などは誤りがなく使用できるが、関係代名詞や後置修飾などの複雑な文法の使用を試みる際には誤りをおかすこともあるような受験者がレベル5となる。一方で、学習指導要領における「知識・技能」の評価規準には、テスト中に受験者が使用する文法形式の難易度への言及はない(「事例」では、「知識・技能」の評価は行っていないため、具体的な評価規準は紹介されていない)。

5.3.4. 本スピーキングテストの評価

授業内で実施したスピーキングテストに関しては、ケンブリッジ英検の分析的評価(表10)を用いた評価に加えて、学習指導要領に基づく表11の採点基準を用いた評価も行った。

表 11. 本スピーキングテストの採点基準

	知識・技能	思考・判断・表現	主体的に学習に取り組む態度
目標	<ul style="list-style-type: none"> 賛成・反対の意見を、論理の構成や展開を工夫して話して伝え合うために必要となる表現を理解している。 日常的な話題や社会的な話題について、賛成・反対の意見を論理の構成や展開を工夫して話して伝え合う技能を身に付けている。 	自分の意見を、相手によりよく理解してもらえるように、日常的な話題や社会的な話題について、相手の意見に応じて、賛成・反対の意見を論理の構成や展開を工夫して話して伝え合っている。	自分の意見を、相手によりよく理解してもらえるように、日常的な話題や社会的な話題について、相手の意見に応じて、賛成・反対の意見を論理の構成や展開を工夫して話して伝え合おうとしている。
a	上記を十分達成し、 <ul style="list-style-type: none"> 語彙や表現が適切に使用されている。 聞き手にわかりやすい音声等で話して伝えている。 の2点を満たしている。	上記を十分達成し、 <ul style="list-style-type: none"> 自分の意見や主張における論理に矛盾や飛躍がない 理由や根拠がより適切なものとなっている の2点を満たしている。	上記を十分達成し、 <ul style="list-style-type: none"> 相手とコミュニケーションを図ろうしている態度が見られる。
b+	b 評価ではあるものの、極めて a に近い場合は b+ とする。		
b	上記をおおむね達成し、 <ul style="list-style-type: none"> 多少の誤りはあるが、理解に支障のない程度の語彙や表現を使って話して伝えている。 理解に支障のない程度の音声等で話している。 の2点を満たしている。	上記をおおむね達成し、相手とは反対の意見の場合は、その理由を適切な表現で述べている。	上記をおおむね達成している。
c+	c 評価ではあるものの、極めて b に近い場合は c+ とする。		
c	「b」を満たしていない。	「b」を満たしていない。	「b」を満たしていない。

表 11 は、国立教育政策研究所 (2021) の「論理・表現 I」の「話すこと [やり取り]」の評価規準 (p. 68) のおよび「事例 3」(p. 73) の採点基準を参考に、著者が本テストのために作成した採点の基準である。

学習指導要領による評価では、目標の達成度合いは、a (十分満足できる)、b (おおむね満足できる)、c (努力を要する) の三段階で評価することになっているが、本テストにおいてはケンブリッジ英検の評価との比較を試みたため、5 段階の数値に換算し易いように、b+ および c+ を設定した。数値に換算する場合は、a = 5, b+ = 4, b = 3, c+ = 2, c = 1 として換算した。

5.3.5. 採点の方法

5.1 で述べた通り、本テストは 1 クラスを二つに分け、授業者である 2 人の教員 (教員 A : 英語母語話者の講師, 教員 B : 著者) が、別々の部屋でテストを行い、それぞれ

れの教員が面接官と評価官の役割を同時に行った。この際の評価は、ケンブリッジ英検の評価基準を用いた。テストはビデオカメラを使って録画した。受験者はテスト当日の座席順に2人1組のペアとなってテスト教室に入室し、教員AおよびBは、面接官としてスピーキングテストをしながら、評価官としてもテストの採点を行い、テスト終了後に受験者が退室した後、採点シートに得点を記入した。その後、教員Bは録画されたスピーキングテストを視聴して、教員Aが採点した受験者53名分を採点した。表12は、教員AとBの採点者間信頼性 (inter-rater reliability) である。

表 12. 本スピーキングテストの採点者間信頼性

	クロンバッファ α
Grammar & Vocabulary	.58
Discourse Management	.81
Pronunciation	.71
Interactive Communication	.62

各項目のクロンバッファアルファの値は .58 から .71 とそれほど高くなく、採点者間信頼性が十分に高いとは言い難い結果だった。スピーキングテストの採点に関しては、事前に各レベルのサンプルパフォーマンスを吟味し、その特徴を明確化する作業 (benchmarking) を行うべきだが、ケンブリッジ英検が公開しているスピーキングテストのデモビデオは、レベル5を得た受験者のものに限られており、レベル3や4の受験者のパフォーマンスを吟味することはできない。また、今回のテストは本校にとって初めてのケンブリッジ英検のスピーキングテスト形式によるテストであり、本校生徒のパフォーマンスレベルも不確定であった。本来はテスト実施後に録画映像を見ながら各レベルのパフォーマンスの典型を洗い出す benchmarking を行うべきだったが、時間的な制約のために実施しなかった。このため、信頼性を確保するために録画映像をもとに採点した教員Bの評価を採択した。教員Bはケンブリッジ英検 A2, B1 レベルの評価官の資格と、米国で広く利用されているスピーキングテストである ACTFL-OPI (Oral Proficiency Interview) の日本版の Standard Speaking Test (SST: 現在は TSST) の採点官の資格を有しており、スピーキングテストの評価の経験も豊富である。

5.4. 本スピーキングテストの結果

本スピーキングテストのスコアは表13と表14の通りである。前述の通り、ケンブリッジ英検の分析的評価基準で評価し、それとは別に、授業の評価の一環としては学習指導要領に基づく三観点でも評価した。学習指導要領に基づく評価は a, b, c に加えて、b+ および c+ を設けることにより、5段階評価として数値化した。

表 13. ケンブリッジ英検の分析的評価の結果 (N=112)

	平均	標準偏差	分散	最大値	最小値
Grammar & Vocabulary	4.07	.48	.23	5	3
Discourse Management	4.10	.53	.28	5	3
Pronunciation	4.17	.41	.17	5	3
Interactive Communication	4.51	.50	.25	5	3

表 14. 学習指導要領による三観点の評価の結果 (N=112)

	平均	標準偏差	分散	最大値	最小値
知識・技能	3.98	.64	.41	5	2
思考・判断・表現	4.24	.62	.39	5	2
態度	4.47	.53	.28	5	3

5.4.1. 本スピーキングテストにおける2つの評価の比較

前述の様に本研究では、ケンブリッジ英検の問題形式を利用した授業内のスピーキングテストを、①ケンブリッジ英検の分析的評価と、②学習指導要領による三観点の評価の2つの方法を利用して評価した。5.3.3で述べたように、両者の評価には対応する項目はあるものの、完全に一致するものではない。表15は、両者の採点者内信頼性(ICC 級内相関係数)である。「知識・技能」に関しては、ケンブリッジ英検のGrammar & VocabularyとPronunciationが対応すると考え、両者の平均値を利用した。

表 15. ケンブリッジ英検と学習指導要領による評価の採点者内信頼性 (N=112)

	ICC
知識・技能× Grammar & Vocabulary, Pronunciation	.37
思考・判断・表現× Discourse Management	.66
態度× Interactive Communication	.62

山本, 他 (2021) は ICC 級内相関係数が .70 以上で「信頼性がある」と言えるとしており, .37 から .66 という相関係数は, 高くないと言えるだろう。このことから, 両評価は2つの異なる規準を用いた独立した評価であり, テスト実施の際には使い分ける必要があると言える。例えば, ケンブリッジ英検を受験する直前にその準備として行う授業内スピーキングテストの場合は, ケブリッジ英検の評価項目を用いて評価し, 得点が低かった項目に関して具体的なフィードバックを与えることが望ましいし, 形成的評価の一環として授業内で行うスピーキングテストの場合は, 学習指導要領によ

る評価を利用するというように、テストの目的に応じて、評価方法を選択することができるだろう。

6. 生徒の反応

授業でのスピーキングテスト終了後、Google フォームを用いて事後アンケート（質問紙調査）を行った（回答者数 96）。生徒は一連の活動やスピーキングテストに関して、おおむね肯定的な感想をもっており、自己の英語学習にとって役だったと感じていた（表 16）。

表 16. 生徒による質問紙調査の結果① (N=96)

	そう思う	ややそう思う	どちらとも言えない	あまりそう思わない	そう思わない
授業中のスピーキングテスト対策の活動は、自分のスピーキング力の向上に役だった	47 (49.0%)	43 (44.8%)	4 (4.2%)	2 (2.1%)	0
授業中のスピーキングテスト対策の活動へは、積極的に取り組むことができた	46 (46.9%)	45 (47.9%)	2 (2.1%)	3 (3.1%)	0
テスト中は自分の力が十分にだせたと感じますか	5 (5.2%)	33 (34.4%)	32 (33.3%)	24 (25%)	2 (2.1%)

2人1組での受験形態に関しては、「緊張しすぎてあまり話せなかったけれど、2人の方が楽しかった」「とても緊張しましたが、ペアでの受験方式の方が個人的にはリラックスして受けられました」など、テストを2人で受けることに関しても肯定的だった（表 17）。

表 17. 生徒による質問紙調査の結果② (N=96)

	特に問題なく受験できた	やりにくかった	ペアの組み方がやりにくかった	受験を機に相手と話せて嬉しかった
2人1組のペアでの受験形態に関して、どう感じましたか	85 (89.5%)	8 (8.4%)	1 (1.1%)	1 (1.1%)

自由記述には、「すごく楽しかったし、力になったと思います！これからもいっぱいやりたいです！」「スピーキングの練習をこれほどしたことがなかったのでたくさんの学びがありよかったです。」など、スピーキング活動に関して肯定的なコメントが多く、また、「本番前にこのような練習ができるのは嬉しかった」「本番前にすごい参考になった」など、ケンブリッジ英検受験前の準備の一環としてのスピーキングテスト

の実施を評価する声も多かった。テストに関して、2人1組のペアでの受験形態を「やりにくかった」と感じたのはなぜかを聞いた問いは、「自分が英語があまり出来ないので、相手に迷惑をかけていないか心配だから」「集中できない上に、余計に緊張するから」「相手との会話を意識するのが大変」「相手に分かりやすく伝えられるかも考えたから」などのコメントがあった。

7. 考察・課題

指導が生徒の学びに結び付いたかを確認するためには、指導内容に応じた適切な評価を行うべきである。生徒同士のペア活動（やりとり）は英語の授業で多用される活動ではあるものの、その具体的な評価方法に関しては課題が多い。本研究では、ケンブリッジ英検のスピーキングテストの形式に応じたスピーキング活動を行ったのち、学習指導要領の評価規準とケンブリッジ英語検定の評価項目を対応させながら、生徒同士のやりとりを評価した。課題としては、Hughes (2003) や Norton (2005) も指摘している通り、受験者の組み合わせにより生徒の発話量に差がでてしまうことや、複数の採点者で採点する場合の採点者間信頼性の低さなどが挙げられた。今後、このようなパフォーマンステストを実施する場合には、テストの目的・形式に応じてルーブリックを作成し、実際の受験者のパフォーマンスを数人分を評価して、採点者間に共通の指標 (benchmark) が確立したのちに採点を行うことが望ましい (スピーキングテストの作り方に関しては、小泉, 2022 が詳しい)。受験者の組み合わせに関しては、どちらかの受験者があまりに不利な条件になった場合にはペアの組み合わせの変更や、再試験などの措置を考慮する必要もあるだろう。学習指導要領の三観点の評価規準に関しては、三観点全てを評価する必要は無く、適宜評価する観点を選択してテストをデザインすることも考えられる。また、教員はこのようなパフォーマンステストの評価を、その後の指導に活かしていくことも視野にいたした指導の全体像を念頭に授業をデザインするべきであり、教員による授業計画には綿密な準備が必要だと言えるだろう。

参考文献 (URL に関しては、全て 2022 年 3 月 28 日に最終アクセスした)

British Council. CEFR (ヨーロッパ言語共通参照枠)

<https://www.britishcouncil.jp/programmes/english-education/updates/4skills/about/cefr>

Cambridge Assessment English. (2020a) .

<https://www.cambridgeenglish.org/research-and-validation/quality-and-accountability/>

Cambridge Assessment English. (2020b) . *B1 Preliminary for School Handbook*.

<https://www.cambridgeenglish.org/exams-and-tests/preliminary-for-schools/preparation/>

Hughes. (2003) . *Testing for Language Teachers*. Cambridge University Press.

- Norton, J. (2005) . The paired format in the Cambridge speaking tests. *English Language Teaching Journal*, 59, 287–297. doi:10.1093/elt/cci057
- Zhou, Y. (2008) . A comparison of speech samples of monologic tasks in speaking tests between computer-delivered and face-to-face modes. *JLTA Journal*. 11, 189-208.
https://www.jstage.jst.go.jp/article/jltaj/11/0/11_KJ00008662563/_article
- 小泉利恵 . (2015) . 「スピーキングの評価：スピーキングテスト作成・実施を中心に」望月昭彦, 深澤真, 印南洋, 小泉利恵 (編著) 『英語 4 技能評価の理論と実践：CAN-DO・観点別評価から技能統合的活動の評価まで』大修館書店 .
- 小泉利恵 . (2022) . 『実例でわかる英語スピーキングテスト作成ガイド』大修館書店 .
- 国立教育政策研究所 . (2021) . 「『指導と評価の一体化』のための学習評価に関する参考資料 高等学校 外国語」東洋館出版社 .
- 白鳥金吾 . (2018) . 「『英語教育改善のための英語力調査』及び『ケンブリッジ英語検定：B1 Preliminary』を利用した日本人英語学習者の英語力・技能別の特徴, 及び仕様がテスト結果に及ぼす影響に関する研究」『北里論集』第 16 号. 1-12.
- 鈴木卓 . (2020) . 「大学入試における英語外部テスト利用の再検討」清泉女子大学言語教育研究所『言語教育研究』第 12 号. 43-65.
- 村岡有香 . (2019) . 「英語民間試験におけるスピーキングテスト問題形式と評価の分析」聖学院大学総合研究所 Newsletter. Vol. 29. 13-20. <http://doi.org/10.15052/00003733>
- 文部科学省 . (2013) . 第 2 期教育振興基本計画
https://www.mext.go.jp/a_menu/keikaku/detail/1336379.htm
- 文部科学省 . (2015) . 各資格・検定試験と CEFR の対照表
https://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afiedfile/2019/01/15/1402610_1.pdf
- 文部科学省 . (2016) . 「平成 27 年度英語教育実改善のための英語力調査事業 (高等学校) 報告書」
https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1377767.htm
- 文部科学省 . (2018) . 「高等学校学習指導要領解説外国語編・英語編」
- 文部科学省 . (2020) . 「令和元年度公立高等学校における英語教育実施調査」
https://www.mext.go.jp/content/20200715-mxt_kyoiku01-000008761_6.pdf
- 山本大貴, 近藤暁子, 鳴海智之, 多田ウエインディ, 吉田達弘 . (2021) . 「日本の高等学校での活用を想定したペア型スピーキングテストの開発」*Language Education & Technology*. 58. 99-128
- 米崎里 . (2016) . 「スピーキングテストの形式が学習者の情意面に及ぼす影響：直接テストと半直接テストの比較を通して」『鳴門英語研究』第 26 巻. 161-173.

資料2 スピーキングテストパート3に利用可能な英語表現リスト

Starting	Shall we start with this one?
	Let's start with this one, shall we?
Suggesting	I think this would be good because...
	This wouldn't be such a good idea because...
Asking your partner	So, what do you think?
	Do you think that's a good idea?
Showing preferences	I prefer...
	I think this is better because...
Agreeing	Yes, that's a good point.
	I see what you mean.
Disagreeing	Well, I'm not so sure about that.
	I don't think that's such a good idea.
Choosing/ summing up	Which one shall we choose?
	So, shall we choose this one?
	Do we agree that this is the best one?

Disagree する時は、相手を傷つけないように配慮しましょう！例えば…
 - I know what you mean, but I think...
 I understand your point, but my idea is...
 などを使うと柔らかい表現となります