

# 「GTEC」スコアと CEFR・CEFR-J レベル 関連付け調査報告

2022 年

はじめに

「GTEC」(Global Test of English Communication)とは、株式会社ベネッセコーポレーションによって開発されたスコア型英語4技能検定試験である。「GTEC」には、小学生～中学1年生向けの「GTEC Junior」、中学生・高校生向け「GTEC」、大学生・社会人向け「GTEC」がある。これらのうち、中学生・高校生向け「GTEC」については、2016年度から2017年度にかけてCEFRレベルとの関連付けを、2018年度にはCEFR-Jレベルとの関連付けを行った。本報告書は、この中学生・高校生向け「GTEC」のうち、Advanced、Basic、Coreタイプにおけるライティングとスピーキングの2技能について、前回までに設定したCEFR・CEFR-Jレベルを見直し、調整を行った際の報告である。

## 1. 調査の背景

大学入試における英語の4技能外部検定の活用や、言語運用能力を客観的に定義するために用いられるCEFR(Common European Framework of Reference for Languages)レベルを起点とした指導と評価の広がりを受け、株式会社ベネッセコーポレーションでは、「GTEC」4技能それぞれのスコアについて、2016年度から2017年度にかけてCEFRの閾値の検証を実施した。また2018年度にはCEFR-Jの閾値の検証を実施した。「GTEC」のスコアを、日本の英語教育での利用を目的に構築されているCEFR-Jと関連付けることで、日本の中高生の指導や学習により活用しやすいフィードバックが可能になった。

この「GTEC」とCEFRとの関連付け調査を行ってから5年以上が経過している。Council of Europeの*Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*には、スタンダードセッティングの信頼性を維持するためには、「パネリストや手法を変えながら設定した閾値を見直すべきである」という記載(2009, p.87)や、「実受検者のパフォーマンスデータを用いてスタンダードセッティングを実施することが望ましい」という記載(2009, p.94)がある。また、CEFRにおいても、各CEFRレベルが各技能・各タスクで実際にできることの記述(can-do ディスクリプタ)が、2年前に改訂された。「GTEC」Advanced、Basic、Coreタイプのライティングとスピーキングでは、2019年度の出題と採点基準の改訂から約2年が経過し、中高生の受検者の十分な解答データが集まったこの段階で、CEFR・CEFR-Jの閾値について再度見直しをする必要があると考えた。そのため本調査では、2019年度以降における中高生の受検者のパフォーマンスに基づき、ライティングとスピーキングそれぞれの技能についてCEFR・CEFR-Jの閾値を検証することとした。

## 2. 出題・採点基準改訂の概要

「GTEC」の詳しい実施形式やIRTを用いたスコア算出方法は、CEFRレベルとの関連付けについての報告書「GTECスコアとCEFRレベル関連付け調査報告」(2017)にて詳細を記しているため、本報告書では割愛する。ここでは、2019年度の「GTEC」

Advanced、Basic、Coreタイプの出題と採点基準の改訂内容を説明する。

ライティングはこれまで意見展開問題一題のみを出題していたが、より多角的に書く力を測定し、特にCEFR Aレベルの受検者に合ったタスクを出題できるよう、相手のメールに返信するというEメール問題を追加した。この改訂によりライティングは、Part AのEメール問題、Part Bの意見展開問題という構成となった。なお、Part Bの意見展開問題は、細かな採点基準の文言に一部修正を加えているが、出題や採点段階数に変更はない。

#### ライティングにおける出題改訂の概要

出題改訂前	出題改訂後
出題なし	<b>Part A Eメール問題 (1題・2問)</b> Task Achievement 1 (×,○) Task Achievement 2 (×,○) 構成 (0～3) 正確さ (0～3)
<b>意見展開問題 (1問)</b> Goal Achievement (意見) (×,○) Goal Achievement (理由) (×,○) 語彙 (0～8) 文法 (0～8) 構成・展開 (0～8)	<b>Part B 意見展開問題 (1問)</b> Goal Achievement (意見) (×,○) Goal Achievement (理由) (×,○) 語彙 (0～8) 文法 (0～8) 構成・展開 (0～8)

スピーキングはPart Aが音読問題、Part Bが質問を聞いて応答する問題、Part Cが4コマのストーリーを英語で話す問題、Part Dが自分の意見を述べる問題となっている。このうちPart Aの音読問題については、従来は短い対話文の片方を読み上げる形式であったが、改訂後はモノログを読み上げる形式に変更した。これは、状況に応じて一定の長さの英文を相手に伝わるように読み上げる力を測定することを目的としている。採点基準の改訂内容としては、Part Aの採点観点を「発音」から、「発音・流ちょうさ」に変更し、Part C・Dの採点観点のうち、「発音」「流ちょうさ」の2つの観点を「発音・流ちょうさ」の1観点到、「語彙」「文法」の2つの観点を「語彙・文法」の1観点到に統合した。また、採点段階数については、「発音・流ちょうさ」を0～4の5段階から0～3の4段階に変更した。さらに、Advanced/Basicタイプのみ、Part Dの「理由が述べられているか」を評価する観点について、×・○の2段階から×・○・◎の3段階評価に変更し、意見をサポートする内容の発展性を評価する設計とした。Coreタイプについては、出題するテーマが日常生活に関する身近なトピックであるため、内容の発展性で差別化をすることが難しいと判断し、×・○の2段階評価を維持した。

スピーキングにおける出題改訂・採点基準変更の概要

出題改訂・採点基準変更前	出題改訂・採点基準変更後
<b>Part A 音読問題（対話文）（6問）</b> 発音（0～3）	<b>Part A 音読問題（モノログ）（2問）</b> 発音・流ちょうさ（0～3）
<b>Part B 質問を聞いて応答する（4問）</b> Goal Achievement 1～4（×,○）	<b>Part B 質問を聞いて応答する（4問）</b> Goal Achievement 1～4（×,○）
<b>Part C ストーリーを英語で話す（1問）</b> Goal Achievement 1～4（×,○） 語彙（0～4） 文法（0～4） 発音（0～4） 流ちょうさ（0～4）	<b>Part C ストーリーを英語で話す（1問）</b> Goal Achievement 1～4（×,○） 語彙・文法（0～4） 発音・流ちょうさ（0～3）
<b>Part D 自分の意見を述べる（1問）</b> Goal Achievement（意見）（×,○） Goal Achievement（理由）（×,○） 語彙（0～4） 文法（0～4） 発音（0～4） 流ちょうさ（0～4）	<b>Part D 自分の意見を述べる（1問）</b> Goal Achievement（意見）（×,○） Goal Achievement（理由）（×,○,◎*） *Coreは（×,○）から変更なし 語彙・文法（0～4） 発音・流ちょうさ（0～3）

これらの改訂を踏まえた2019年度以降のライティングとスピーキングの出題例は、付録として本報告書の巻末に掲載している。

### 3. 閾値設定範囲

今回の調査対象となるのは、中学生・高校生向け「GTEC」のAdvanced、Basic、Coreの3タイプである。その中で上限スコアが最も高いAdvancedタイプでB2下位までを測定するため、本調査の閾値設定範囲は、ライティングとスピーキングのPre-A1～B2までとした。調査前の閾値は、以下、表1のとおりである。

【表1. 調査前のCEFR-Jの閾値】

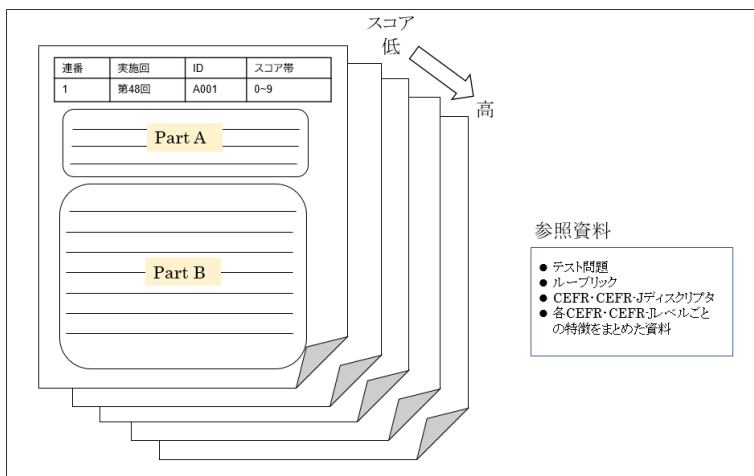
CEFR	CEFR-J	ライティング	スピーキング
B2	B2	300	320
B1	B1.2	260	300
	B1.1	240	280
A2	A2.2	220	220
	A2.1	190	190
A1	A1.3	140	140
	A1.2	100	100
	A1.1	60	80
Pre-A1	Pre-A1	～59	～79

#### 4. 使用データ

今回は、過去に実施した「GTEC」の検定版における受検者のライティングの答案データ、スピーキングの解答音声データを抽出して検証用のブックレットを作成した。Core タイプは第44回（2020年度第2回検定）・第49回（2021年度第2回検定）の2版を使用し、Advanced/Basic タイプでは、第46回（2020年度第3回検定）・第48回（2021年度第1回検定）の2版を使用した。

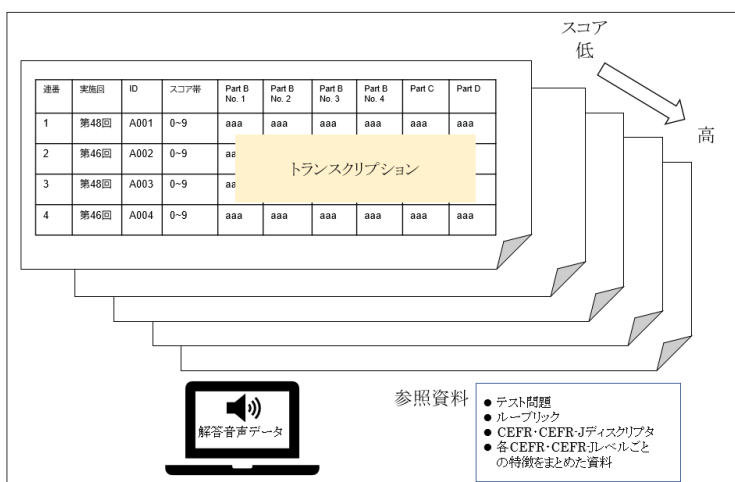
ライティングでは、「GTEC」のライティングスコアが付与されている受検者のPart A と Part B の答案を準備した。そしてそれらを Advanced、Basic、Core それぞれ10点刻みのスコアごとに4～5人分ずつ任意に選び、スコアが低い方から高い方へ順に答案を並べて、ブックレットを作成した。各タイプの抽出データ数は、Advanced タイプ44名、Basic タイプ66名、Core タイプ70名であった。なお、議論に影響を与えないよう、各答案への現在のCEFR・CEFR-J レベル表示はしなかった。また、閾値を10点刻みで設定するために、スコアは「0～9点」、「10～19点」というように10点ごとのスコア帯を表示させた。

## ライティングのブックレットのイメージ



スピーキングでは、「GTEC」のスピーキングスコアが付与されている受検者の Part A～D の解答のうち、「やりとり」として Part B、「発表」として Part C と Part D のトランスクリプション（受検者の発話の書き起こし）を用意した。音読という観点の CEFR および CEFR-J ディスクリプタは存在しないため、Part A の音読問題は本調査では除外した。これらのトランスクリプションを Advanced/Basic タイプと Core タイプそれぞれ 10 点刻みのスコアごとに 4～5 人分ずつ任意に選び、スコアが低い方から高い方へ順に答案を並べて、ブックレットを作成した。各タイプの抽出データ数は、Advanced/Basic タイプ 88 名、Core タイプ 51 名であった。なお、ブックレットを作成する際には、ライティングと同様に現在の CEFR・CEFR-J レベルの表示は伏せ、スコアは「0～9 点」、「10～19 点」というように 10 点ごとのスコア帯を表示させた。また、検証の際に「発音・流ちょうさ」を確認できるよう、トランスクリプションに加えて解答音声データを準備した。

## スピーキングのブックレットのイメージ



今回の調査では、これらAdvanced、Basic、Core（スピーキングではAdvanced/Basic、Core）のブックレットのうち、閾値を検討する各CEFR・CEFR-Jレベルと、出題しているタスクのレベルが近いタイプのブックレットを使用した。具体的には以下、表2のとおりである。

【表2. 各 CEFR・CEFR-J レベルの関連付けで使用した問題タイプ】

技能	CEFR・CEFR-J	タイプ
ライティング CEFR	Pre-A1/A1	Core
	A1/A2	Basic・Core
	A2/B1	Advanced
	B1/B2	Advanced
ライティング CEFR-J	A1.1/A1.2	Core
	A1.2/A1.3	Core
	A2.1/A2.2	Basic
	B1.1/B1.2	Advanced

技能	CEFR・CEFR-J	タイプ
スピーキング CEFR	Pre-A1/A1	Core
	A1/A2	Advanced/Basic
	A2/B1	Advanced/Basic
	B1/B2	Advanced/Basic
スピーキング CEFR-J	A1.1/A1.2	Core
	A1.2/A1.3	Core
	A2.1/A2.2	Advanced/Basic
	B1.1/B1.2	Advanced/Basic

## 5. パネリスト

パネリストは東京外国語大学大学院の投野由紀夫教授、清泉女子大学言語教育研究所の小泉利恵准教授の2名とした。投野由紀夫教授はCEFR-J科学研究費補助金研究（科研）の研究統括者であり、日本の英語教育へのCEFRベースの英語指標浸透に貢献している研究者でもある。2017年度のCEFR関連付け調査および2018年度のCEFR-J関連付け調査にも参画した。小泉利恵准教授はCEFRおよび英語テストにおける教育測定に精通した研究者であり、外国語の評価方法の研究を専門としている。CEFR-J科研に2016年度～2019年度に参画し、CEFR-JのCAN-DOに沿った「やりとり」のテスト・タスク作成を担った。2名とも共通参照枠としての各レベルの枠組みの理解のみならず、英語特有の言語特徴に

についての知識も深い専門家である。

また、パネリストの他に、「GTEC」の作問・採点に関わる編集メンバー（英語を母語とするメンバーを含む）も参画し、会議進行を行いながら、「GTEC」の出題や採点基準の説明、パネリストからの質問に対する回答、機器の操作（「解答音声を聞くための機器（コンピュータ）の操作」など）、議事録作成などの補助作業を行った。

## 6. 調査の方法

本調査では、Council of Europe (2009) のテストを CEFR と関連付けるためのマニュアル（以降、CEFR 関連付けマニュアル）に基づき、The Bookmark Method（しおり推定法）と The Body of Work Method（作品集合法）を合わせたスタンダードセッティングの手法を用いた（詳細は、Downing & Haladyna, 2008; 大友, 2009; 大友他, 2012 参照）。

The Bookmark Method はテスト項目中心であり、統計的なアプローチをベースに閾値を決めるスタンダードセッティングの手法である。困難度が低いタスクから高いタスクへと並べ、特定の CEFR レベルの受検者がどのくらいの確率でそれぞれのタスクに正答できるかを低い困難度のタスクから順に判断し、閾値を決めていく手法である。正答率が response probability（反応確率）を下回ると判断された最初のタスクのスコアが閾値となる。

The Body of Work Method は受検者中心であり、受検者の様々なタスクの解答から総合的に判断し CEFR を紐づけていくスタンダードセッティングの手法である。解答を2つのレベル（例えば B1 か B2）に振り分けた後、ロジスティック回帰分析を用いて、閾値を決定することが多い。

The Bookmark Method ではタスクを困難度の低い順に並べるのが特徴である。一方、「GTEC」では受検者の解答ごとに IRT を用いてスコアの算出が可能であり、それを用いる方がタスクを困難度順に並べた場合よりも、より詳細な閾値の検討が可能である。そのため、タスクではなく、受検者の答案データ・解答音声データをスコアが低いものから順に並べて検討することにした。さらに、The Bookmark Method では二値で評価されるタスクを扱うことが多く、本調査のように多値で評価される受検者のパフォーマンスを検証する場合、response probability（反応確率）を設定し、正答率を判断していくプロセスをとることが難しいため、代わりに The Body of Work Method の受検者のアウトプットを総合的に判断して CEFR レベルを紐付けていく手法を進めることとした。答案データ・解答音声データ確認の際は、The Body of Work Method に従い、受検者を3つのカテゴリ（例えば B1、B1 と B2 のボーダーライン、B2）に分けて討論を行った。そのスコア帯および前後の答案データ・解答音声データを確認し、最終的にボーダーとなるスコア帯を中心に、どこが閾値として適切かを議論し決定した。

また、The Body of Work Method ではロジスティック回帰分析を使用して閾値を算出することが書かれている（Council of Europe, 2009）。The Body of Work Method の場合、スコアはわからないブラインドの状態、複数の有識者が受検者の解答に CEFR レベルを



付けてグルーピングをする。有識者ごとのグルーピングの差異について、どのスコアを閾値と定めるのが確率的に妥当かを統計的に導き出すために、一般的にはロジスティック回帰分析が必要とされる。

一方、a training exercise (トレーニング段階)、a range-finding stage (範囲探索段階)、a pinpointing stage (正確な閾値決定段階) を経て議論し、ロジスティック回帰分析を使用しないで閾値を決定する手法も提唱されている (Zieky & Perie, 2006)。今回は The Bookmark Method をベースとしており、解答がスコアの低い順に並んだブックレットを使用し、どこに閾値があるかを見つけるプロセスをとった (範囲探索段階)。意見が分かれた場合も、CEFR に精通しているパネリスト同士で十分議論し、合意をとった上で閾値を決めることができるため (正確な閾値決定段階)、回帰式を作成して閾値を決定するロジスティック回帰分析は必要ではないと判断した。

またパネリストを2名にした大きな理由として、今回はすでに一度設定した閾値があり、その見直しというプロセスであるため、大きく範囲検索段階で変更が出ることはないだろうと予想されたことが上げられる。パネリスト2名も CEFR に精通し、スタンダードセッティングの経験もされている有識者であり、さらに小泉先生はスピーキングを専門としており、今回の発信技能の閾値決めにおいては、専門の知見が活かせることでパネリスト決定に至った。

## 7. 分析の流れ

CEFR関連付けマニュアルでは、スタンダードセッティングを行う前に、すべての参加者がCEFRの知識をつけておくためのFamiliarization (習熟化)、閾値を決める英語資格試験の出題内容について理解するSpecification (明確化) の2つの工程が必要だと述べられている (Council of Europe, 2009)。本調査でもその2段階を行ったうえで、スタンダードセッティングを行った。まず、パネリストへ事前にFamiliarizationとSpecificationに使用する資料を送付し、パネリストが個別に学習したうえで、オンラインによる2回の事前会議にてFamiliarizationおよびSpecificationを行った。

【表3. スタンダードセッティングまでのステップ】

事前送付	#	手順	目的
	1	Familiarization	かかわるメンバー全員がCEFRについて深い理解をもつ -GTECのスピーキング・ライティングのタスクや採点基準に関連するディスクリプタを共有 -他社が公開しているCEFR別解答例などを共有
	2	Specification	自分たちのテストや採点基準をCEFRと照らし合わせ、どのレベルに沿った設計になっているかを探る -GTECの出題内容や採点基準を共有
	3	Standardization and Benchmarking	スタンダードセッティングにあたり、実際のアウトプットを用いて理解の基準を揃える -前回のスタンダードセッティングで作成した、各CEFR・CEFR-JレベルのGTECの答案・解答の特徴を共有 -他社が公開しているCEFR別解答例などを再度参照
当日議論	4	Standard setting	CEFRとCEFR-Jの閾値を決める -前回のスタンダードセッティングで作成した、各CEFR・CEFR-JレベルのGTECの答案・解答の特徴を参照

Familiarizationの工程では、CEFR並びにCEFR-Jに記載されている各レベルの特徴（表4）の確認と、他社の英語資格試験のスコアとCEFRとの関係、公表されているCEFRレベルごとのライティングの答案データ・スピーキングの解答音声データなどの確認を行った。

【表4. CEFRディスクリプタの参照ページ】 (Council of Europe, 2020)

Writing	Speaking
<ul style="list-style-type: none"> <li>• Overall written production (p. 66)</li> <li>• Overall written interaction (p. 82)</li> </ul> <p>&lt;Part A (Email)&gt;</p> <ul style="list-style-type: none"> <li>• Correspondence (pp. 82–83)</li> <li>• Notes, messages and forms (pp. 83–84)</li> </ul> <p>&lt;Part B (Essay)&gt;</p> <ul style="list-style-type: none"> <li>• Creative writing (p. 67)</li> <li>• Reports and essays (p. 68)</li> </ul>	<ul style="list-style-type: none"> <li>• Overall oral production (p. 62)</li> <li>• Overall oral interaction (p. 72)</li> </ul> <p>&lt;Part B (Listening and responding)&gt;</p> <ul style="list-style-type: none"> <li>• Conversation (pp. 73–74)</li> <li>• Information exchange (pp. 78–79)</li> </ul> <p>&lt;Part C (Telling a story)&gt;</p> <ul style="list-style-type: none"> <li>• Sustained monologue: describing experience (pp. 62–63)</li> </ul> <p>&lt;Part D (Expressing your opinion)&gt;</p> <ul style="list-style-type: none"> <li>• Sustained monologue: putting a case (e.g. in a debate) (p. 64)</li> </ul>

Writing-Vocabulary / Writing-Grammar/ Writing-Organization / Speaking-Linguistic Control	Speaking Fluency
<ul style="list-style-type: none"> <li>• General linguistic range (pp. 130–131)</li> <li>• Vocabulary range (p. 131)</li> <li>• Grammatical accuracy (p. 132)</li> <li>• Vocabulary control (pp. 132–133)</li> <li>• Thematic development (pp. 139–140)</li> <li>• Coherence and cohesion (pp. 140–141)</li> </ul>	<ul style="list-style-type: none"> <li>• Phonological control (pp. 133–135)</li> <li>• Fluency (p. 142)</li> </ul>

Specificationの工程では「GTEC」の出題内容や採点基準をパネリストへ説明し、パネリストは実際の問題冊子で出題内容の確認を行った。また、今回は2019年度問題改訂後の実受検者の答案データと解答音声データを用いて検証する目的であったため、問題や採点基準がどのように改訂されたかという点についてもパネリストと確認した。パネリストの疑問点は、2回の事前会議で解消できるよう、スタンダードセッティング当日に向けて準備を行った。Standardization and Benchmarkingの工程において、2018年度に行ったスタンダードセッティングで作成した、「GTEC」の解答を利用したCEFR-J別の特徴をまとめた資料を使用した。（「5. 使用データ」参照）。

スタンダードセッティング当日の流れは表5のとおりである。また、タイムテーブルは表6のとおりであり、1日にわたる集合型のワークショップ形式で行った。

【表5. スタンダードセッティングの流れ（各技能共通）】

	内容	備考
①	ブックレットで現在の各CEFR・CEFR-Jレベルの閾値スコア前後約20点に該当する答案データ・解答音声データを確認し、3つのカテゴリ（閾値の下のレベルのグループ、ボーダーラインのグループ、閾値の上のレベルのグループ）に分ける。その際にCEFR・CEFR-Jのディスクリプタも参照する。	パネリスト 個人作業
②	パネリスト両者のグルーピングを共有する。	
③	ボーダーラインのカテゴリに分類されたスコア帯が異なる場合、パネリスト2名で討議を行い、双方合意のもと閾値を決定する。	パネリスト の討論
④	①～③を各技能、CEFR・CEFR-Jレベルで繰り返し行う。	

【表6. スタンダードセッティングの技能・CEFR・CEFR-Jごとのタイムテーブル】

開始	終了	項目	詳細
9:30		開始	
9:30	9:45	当日の流れ・プロセスの説明	
9:45	10:10	スピーキング CEFR 解答確認	Pre-A1/A1
10:10	10:40	スピーキング CEFR 解答確認	A1/A2
10:40	11:20	スピーキング CEFR 解答確認	A2/B1
11:20	12:00	スピーキング CEFR 解答確認	B1/B2
12:00	12:45	昼休憩	
12:45	13:05	ライティング CEFR 答案確認	Pre-A1/A1
13:05	13:25	ライティング CEFR 答案確認	A1/A2
13:25	14:05	ライティング CEFR 答案確認	A2/B1
14:05	14:45	ライティング CEFR 答案確認	B1/B2
14:45	15:00	小休憩	
15:00	15:25	ライティング CEFR-J 答案確認	A1.1/A1.2
15:25	15:50	ライティング CEFR-J 答案確認	A1.2/A1.3
15:50	16:15	ライティング CEFR-J 答案確認	A2.1/A2.2
16:15	16:55	ライティング CEFR-J 答案確認	B1.1/B1.2
16:55	17:10	小休憩	
17:10	17:30	スピーキング CEFR-J 解答確認	A1.1/A1.2
17:30	17:50	スピーキング CEFR-J 解答確認	A1.2/A1.3
17:50	18:10	スピーキング CEFR-J 解答確認	A2.1/A2.2

18:10	18:50	スピーキング CEFR-J 解答確認	B1.1/B1.2
18:50	19:00	終了	

表5の①の工程では、パネリストが各自で答案データ・解答音声データを確認し、ボーダーラインのグループとなるスコア帯を吟味する時間とした。②の工程では、パネリスト各自が考えた3つのカテゴリ（定める閾値の下のグループ、ボーダーラインのグループ、定める閾値の上のグループ）のスコア帯を発表し、その結論に至った背景や、ボーダーラインのグループに設定した理由などを参加者全員と共有する時間とした。③の工程では、パネリストの閾値が異なる場合、各パネリストが導き出したボーダーラインや定める閾値の上のグループのスコア帯の解答と CEFR・CEFR-J のディスクリプタを参照し、必要に応じて追加の答案データ・解答音声データを確認しながら、双方でどこを閾値とするか議論し、合意のうえに閾値を決定した。さらなるプロセスは、以降で詳述する。

#### 8. ライティング・スピーキングの検討結果

この章では、ライティング・スピーキングの技能ごと、および各 CEFR・CEFR-J レベルの閾値設定段階において、どのような検討や議論を経て、最終的に合意に至ったかについて説明する。当日の検討の順番に合わせて、CEFR のレベルの低い方から Pre-A1/A1、A1/A2、A2/B1、B1/B2 の順に、その後、CEFR-J のレベルの低い方から A1.1/A1.2、A1.2/A1.3、A2.1/A2.2、B1.1/B1.2 の順に記述する。

各技能、各 CEFR・CEFR-J レベルの閾値の検討内容において、7.分析の流れの表5にある①CEFR・CEFR-J のディスクリプタを参照しながら、3つのカテゴリ（定める閾値の下のグループ、ボーダーラインのグループ、定める閾値の上のグループ）に分けて、各パネリストが仮閾値を設定、②各パネリストの結果共有、③パネリスト2名の討議と閾値決定、というプロセスで行ったため、本章でも同様の流れで説明する。

また、7.で述べたように、閾値を決定する前段階として、パネリストは Familiarization の工程で各 CEFR・CEFR-J レベルの特徴やディスクリプタを確認している。当日も、Familiarization で使用した資料は手元に置いて参照しながら討議した。8. 1. ライティング、8. 2. スピーキングの各 CEFR・CEFR-J レベルに関する報告の冒頭に、Familiarization で確認した各 CEFR・CEFR-J レベルについての特徴を記載している。CEFR については、「GTEC」のタスクと合致した特徴をもとに閾値を検討することを目的として、CEFR レベルとの関連付けについての報告書「GTEC スコアと CEFR レベル関連付け調査報告」（2017）におけるパネリストの議論から、各 CEFR の特徴をまとめたものを引用した。また、CEFR-J については、東京外国語大学 投野由紀夫 研究室における“CEFR-J 本体：日本語版”より引用した。

## 8.1. ライティング

### CEFR の閾値検討

<ライティング Pre-A1/A1>

#### ◆Pre-A1 と A1 の特徴

	各 CEFR の特徴
Pre-A1	<ul style="list-style-type: none"> <li>● アルファベットの太文字・小文字、単語のつづりをブロック体で書くことができる。</li> <li>● 単語のつづりを1文字ずつ発音できれば、聞いてそのとおりに書くことができる。また書いてあるものを写すことができる。</li> </ul>
A1	<ul style="list-style-type: none"> <li>● いつも完全な文を産出できるという段階には至らず、語句やフレーズのみでの表現や、単文を主とした産出ができる。</li> <li>● 住所・氏名・職業などの項目がある表を埋めることができる。</li> <li>● 辞書を使えば、自分について基本的な情報を短い句または文で書くことができる。</li> </ul>

#### ◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
Pre-A1	～49	～69
Pre-A1/A1	50～69	70～79
A1	70～	80～

パネリスト A は、Pre-A1 は、文字の認識・書き写しができることが特徴であると述べ、スコア 49 以下の答案において、英語の文字は書けているものの、意味を成す語を書いているのではなく英文字の羅列であることが多いことから Pre-A1 の特徴にあてはまると判断した。また、A1 は、自分が知っている単語をいくつか書けるのが特徴だと述べ、その特徴が顕著であるスコア 70 以上の答案を A1 と判断した。スコア 50～60 台では、単語が書けている答案と、書けていない答案が混在していることから、Pre-A1 と A1 のボーダーとした。

パネリスト B も、Pre-A1 は、文字の認識・書き写しができることが特徴であると述べ、スコア 69 以下の答案が Pre-A1 にあてはまると判断した。A1 については、辞書を使えば短い句または文で書けることが特徴であると述べ、“I like go to the karaoke.”や“I have a book.”などと書けているスコア 80 以上の答案が A1 の特徴にあてはまると判断した。スコア 70 台については、単語のみ、または意味をなさない英文字の羅列が見られるため、Pre-A1 と A1 のボーダーとした。

#### ◆閾値の検討と結果

Pre-A1 と A1 の閾値：スコア 70

これらの要素を踏まえ、閾値を決定するための議論に移った。議論の中で、パネリスト A

は、スコア 60 台では、単語は書けているが短い句を作るまでに至らないことや、意味をなさない言葉が含まれている傾向に着目し、Pre-A1 の特徴がより強く現れていることから、改めてスコア 60 台は Pre-A1 と判断した。また、パネリスト B は、問題冊子で見ることができる Part A の E メール問題の出題文で使用されている英文の一部や単語を書き写していると思われる答案がスコア 60 台で見られることに着目し、これが Pre-A1 の特徴と合致しているとした。両者の意見を総合し、スコア 69 以下は Pre-A1 と結論付けた。両者とも、スコア 70 以上の答案では、知っている単語やフレーズを使って基本的な文を産出することができるという A1 の特徴が見られると判断し、スコア 70 が閾値として適切であると判断した。

#### <ライティング A1/A2>

##### ◆A1 と A2 の特徴

	各 CEFR の特徴
A1	<ul style="list-style-type: none"> <li>● いつも完全な文を産出できるという段階には至らず、語句やフレーズのみ表現や、単文を主とした産出ができる。</li> <li>● 住所・氏名・職業などの項目がある表を埋めることができる。</li> <li>● 辞書を使えば、自分について基本的な情報を短い句または文で書くことができる。</li> </ul>
A2	<ul style="list-style-type: none"> <li>● シンプルな文構造を使えると同時に、“and”や“because”などを使って文をつないで書くことができるため、完全な文が複数書ける。</li> <li>● 自分に関することや身近なことについて、いくつかの文で書くことができる。</li> <li>● 覚えた文を使っており、時制などの文法の誤りはあるが、意味は通る。</li> <li>● Email の書き方が身についておらず、単に聞かれた質問への解答のみを書いていることが多い。</li> <li>● 日常的・個人的な内容であれば、招待状、私的な手紙、メモ、メッセージなどを簡単な英語で書くことができる。</li> <li>● 文と文を簡単な接続詞でつなげるような書き方であれば、基礎的・具体的な語彙、簡単な句や文を使った英語で、日記や写真、事物の説明文などのまとまりのある文章を書くことができる。</li> <li>● 基本的な文は書けている。文と文を接続詞でつないで書くことができる。 また、接続詞などを用いて 1 文をより長めに書くことができる。 “I”で始まる文が多いが、使おうとする文型に少し広がりが見られる。 文章間に意味のつながりを持たせることができ、ある程度まとまった量の文章を書くことができる。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1	～169	～169
A1/A2	170～189	170～179
A2	190～	180～

パネリスト A は、スコア 170～180 台では、Part B の意見展開問題において、受検者の意見と補足情報（理由付けなど）がある程度まとまりをもっていくつか書いているという A2 の特徴が見られる答案と、主張があることは読みとれるものの、文の意味が理解しづらい答案や、文と文のつながりが見られないという A1 の特徴を持った答案が混在すると述べ、このスコア帯を A1 と A2 のボーダーと判断した。スコア 190 台の答案においては、接続詞や“**For these reasons**”のような意見展開問題のタスクに適した定型表現が使用されていることと、用いている語句の正確性があがることから、A2 と判断した。

パネリスト B は、スコア 170 台のスコア帯は、多くが定型表現で構成されている傾向を踏まえ、A1 と A2 のボーダーと判断した。また、スコア 180 台の答案において、仮定法が出現している点に着目した。覚えた定型表現をそのまま書くのではなく、文脈や意図に沿った表現を使用できているとし、スコア 180 を A2 と判断した。

◆閾値の検討と結果

A1 と A2 の閾値：スコア 190

これらの要素を踏まえ、閾値を決定するための議論に移った。スコア 170 台及び 180 台の答案では、定型表現をなぞるだけで、意見展開問題というタスクに合わせた表現を使えておらず、表現方法が限定的という A1 の特徴を併せ持つ答案もまだ一部見られることから、これらのスコア帯は閾値として適切ではないと判断した。スコア 190 台の答案を追加で確認したところ、共通して SVC・SVO の構造で文を書くことができたり、意見展開問題において理由などの補足情報を含めることができたりする特徴が確認された。従って、シンプルな文構造の使用や、身近なことをいくつかの文で書くことができるといった A2 の特徴が、スコア 190 台の答案で顕著に現れるという点で両者が合意し、ここを閾値と結論付けた。

<ライティング A2/B1>

◆A2 と B1 の特徴

	各 CEFR の特徴
A2	● シンプルな文構造を使えると同時に、“and”や“because”などを使って文をつないで書くことができるため、完全な文が複数書ける。



	<ul style="list-style-type: none"> <li>● 自分に関することや身近なことについて、いくつかの文で書くことができる。</li> <li>● 覚えた文を使っており、時制などの文法の誤りはあるが、意味は通る。</li> <li>● Email の書き方が身についておらず、単に聞かれた質問への解答のみを書いていることが多い。</li> <li>● 日常的・個人的な内容であれば、招待状、私的な手紙、メモ、メッセージなどを簡単な英語で書くことができる。</li> <li>● 文と文を簡単な接続詞でつなげるような書き方であれば、基礎的・具体的な語彙、簡単な句や文を使った英語で、日記や写真、事物の説明文などのまとまりのある文章を書くことができる。</li> <li>● 基本的な文は書けている。文と文を接続詞でつないで書くことができる。 また、接続詞などを用いて1文をより長めに書くことができる。 “I”で始まる文が多いが、使おうとする文型に少し広がりが見られる。 文章間に意味のつながりを持たせることができ、ある程度まとまった量の文章を書くことができる。</li> </ul>
B1	<ul style="list-style-type: none"> <li>● 文章の中でアイデアをうまくつなげて、パラグラフライティングができるようになっている。</li> <li>● 自分に直接関わりのある環境、職場、地域などでの出来事を、身近な状況で使われる語彙・文法を用いて、ある程度まとまりのあるかたちで、描写することができる。</li> <li>● 身近な状況で使われる語彙・文法を用いて、筋道を立てて、作業の手順などを示す説明文を書くことができる。</li> <li>● 形容詞節を用いた長い文を書くことができはじめる。 複数の段落の形を意識・構成して文章を書くことができる。ただし、段落内でアイデアを深めるまでには至っていない。 (身の回りのこと以上ではあるが) 自分の経験に紐づけて考えられる内容については、様々な文法や単語を用いて文章を作ろうとするが、日本語の影響を受けた語彙選択の誤りなどが見られる。使い方が適切ではないところがあり、わかりづらい部分がある。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A2	～229	～229
A2/B1	230～239	230～239
B1	240～	240～

パネリスト A は、スコア 230 台では、必ずしも正確ではないものの現在完了や関係代名詞などの複雑な文法を使用して、ある程度まとまりのある文章を書けている答案が散見されるため、B1 の特徴が見られるとし、A2 と B1 のボーダーと判断した。スコア 240 台の答案は、複雑な文法が使用され、書かれている内容の読みやすさが増しているとし、B1 と判断した。

パネリスト B は、スコア 230 台の答案では、関係代名詞を用いるなど、使用できる文法の幅が広がってきているという点において B1 の特徴があるとしたが、一方で文のまとまりが弱く、構文の複雑度合いが文によって異なる答案も確認できることから、A2 と B1 のボーダーと判断した。スコア 240 台の答案では、名詞や動詞を形容詞や副詞を用いて修飾している様子が確認できた。修飾語を使用することにより、文構造に複雑さが増し、また、詳細情報を伝えられるようになることから内容も読みやすくなっていた。そのため、B1 と判断した。

#### ◆ 閾値の検討と結果

##### A2 と B1 の閾値：スコア 230

これらの議論を踏まえ、閾値を決定するための議論に移った。両パネリストとも、スコア 230 台の答案で、一部 A2 らしさが残るものがある一方で、ある程度まとまりのある文章が書けていたり、関係代名詞などの複雑な文法を使えたりするなど、B1 の特徴が十分見られる解答が一定数存在するため、スコア 230 が閾値として適切だと判断した。

#### <ライティング B1/B2>

##### ◆ B1 と B2 の特徴

	各 CEFR の特徴
B1	<ul style="list-style-type: none"> <li>● 文章の中でアイデアをうまくつなげて、パラグラフライティングができるようになっている。</li> <li>● 自分に直接関わりのある環境、職場、地域などでの出来事を、身近な状況で使われる語彙・文法を用いて、ある程度まとまりのあるかたちで、描写することができる。</li> <li>● 身近な状況で使われる語彙・文法を用いて、筋道を立てて、作業の手順などを示す説明文を書くことができる。</li> <li>● 形容詞節を用いた長い文を書くことができはじめる。 複数の段落の形を意識・構成して文章を書くことができる。ただし、段落内でアイデアを深めるまでには至っていない。 (身の回りのこと以上ではあるが) 自分の経験に紐づけて考えられる内容については、様々な文法や単語を用いて文章を作ろうとするが、日本語の影響を受けた語彙選択の誤りなどが見られる。使い方が適切で</li> </ul>

	はないところがあり、わかりづらい部分がある。
B2	<ul style="list-style-type: none"> <li>● 文章内で自らの意見を主張し、議論を展開することができる。</li> <li>● 自分の意見とは対立する意見も交えて、それを否定する形で自分の意見を肯定するといった議論を構成することができる。</li> <li>● パラグラフ同士の流れまではうまくつながりを示すことができないことがあっても、パラグラフ内では、アイデアを一貫して述べることができる。</li> <li>● アカデミックライティング相当のライティングの力がある。</li> <li>● 興味のある分野に関して、幅広いテーマで、明快かつ詳細な文章を書くことができる。</li> <li>● 特定の意見に対して、賛成または反対の立場で情報を伝えたり理由を述べたりしながらエッセーやレポートを書くことができる。</li> <li>● 出来事や経験について、自分が大切だと思うことを強調しつつ、手紙やメールを書くことができる。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
B1	～279	～279
B1/B2	280～289	280～289
B2	290～	290～

パネリスト A は、スコア 280 台では、議論を展開したり詳細な記述ができたりするといった B2 の特徴が見られない答案が混在していることから、B1 と B2 のボーダーと判断した。一方、スコア 290 台の答案は、議論を深く展開し詳細なアイデアを含めて書けることから B2 と判断した。また、表現の側面からは、“come across”や“well-known”、“appropriately”などを使用して語彙が豊かである点や、単語の繰り返しが少ないこと、複雑な文法を誤りなく適切に使用できていることなどの特徴を挙げた。

パネリスト B は、スコア 280 台では、段落の中で一貫したアイデアを述べられている答案がある一方、簡素な意見を述べるのに留まり細かい説明が不足している答案も見られると述べ、ここが B1 と B2 のボーダーと判断した。スコア 290 台の答案では、“to my regret”などの様々な表現を活用し、意見を効果的に、かつ、わかりやすく整理して展開できていること、また、書かれている論に一貫性があるという点から、B2 と判断した。

◆閾値の検討と結果

B1 と B2 の閾値：スコア 290

これらの要素を踏まえ、閾値を決定するための議論に移った。この議論のなかで、パネリスト A は、スコア 280 台の答案は内容が個人の経験に限定されており、主張・例示方法に

アカデミックさが欠ける答案が多いことを指摘し、改めてスコア 280 台では B2 よりも B1 の特徴が色濃いため、閾値として適切ではないと判断した。パネリスト B は意見の展開・発展性に着目し、スコア 280 台は意見の展開が限定的であることから、B2 の要素が弱いと述べた。一方、スコア 290 台の答案では、複雑な文法構造が安定して使えていることと、“indicate”や“regardless”などの多様な語彙が使えていること、また、社会的な視点を含めて論を展開できていることやその読みやすさなどから、B2 の特徴が含まれていると判断した。以上の検討を踏まえ、スコア 290 が閾値として適切だと判断した。

### CEFR-J の閾値検討

<ライティング A1.1/A1.2>

#### ◆A1.1 と A1.2 の特徴

	各 CEFR-J の特徴
A1.1	<ul style="list-style-type: none"> <li>● 住所・氏名・職業などの項目がある表を埋めることができる。</li> <li>● 辞書を使えば、自分について基本的な情報を短い句または文で書くことができる。</li> </ul>
A1.2	<ul style="list-style-type: none"> <li>● 簡単な語や基礎的な表現を用いて、身近なこと（好き嫌い、家族、学校生活など）について短い文章を書くことができる。</li> <li>● 簡単な語や基礎的な表現を用いて、メッセージカード（誕生日カードなど）や身近な事柄についての短いメモなどを書ける。</li> </ul>

#### ◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1.1	～99	～99
A1.1/A1.2	該当なし	該当なし
A1.2	100～	100～

パネリスト A は、スコア 99 以下の答案において、産出している英語の量は多いものの単語の羅列で終始していたり、文を構成できていなかったりするため、A1.1 と判断した。スコア 100 台の答案では、Part B の意見展開問題において、簡単な語句を用いて自分の意見を表現できているため、A1.2 と判断した。

パネリスト B は、スコア 99 以下の答案において、産出している英語の量が少なかったり、英語の量は十分であっても意味を成さない箇所を多く含んだりするため、A1.1 と判断した。スコア 100 台の答案においては、基礎的な表現を使って意見を表現できているものも確認できるため、スコア 100 が A1.1 と A1.2 のボーダーと判断した。

◆ 閾値の検討と結果

A1.1 と A1.2 の閾値：スコア 100

これらの要素を踏まえ、閾値を決定するための議論に移った。両パネリストとも、スコア 100 台の答案で基礎的な表現を使い、Part B の意見展開問題において、課題文に対して“I have camera.”のように自分の意見を表現できていると述べ、改めて A1.2 の特徴が現れはじめることを確認したことから、スコア 100 が閾値として適切だと判断した。

<ライティング A1.2/A1.3>

◆ A1.2 と A1.3 の特徴

	各 CEFR-J の特徴
A1.2	<ul style="list-style-type: none"> <li>● 簡単な語や基礎的な表現を用いて、身近なこと（好き嫌い、家族、学校生活など）について短い文章を書くことができる。</li> <li>● 簡単な語や基礎的な表現を用いて、メッセージカード（誕生日カードなど）や身近な事柄についての短いメモなどを書ける。</li> </ul>
A1.3	<ul style="list-style-type: none"> <li>● 自分の経験について、辞書を用いて、短い文章で書くことができる。</li> <li>● 趣味や好き嫌いについて複数の文を用いて、簡単な語や基礎的な表現を使って書くことができる。</li> </ul>

◆ 各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1.2	～139	～139
A1.2/A1.3	140～149	140～149
A1.3	150～	150～

パネリスト A は、スコア 130 台の答案は、SVC・SVO 構造に限定された、基礎的な表現を使用したものが多いことから、A1.2 の特徴が強く出ていると述べた。スコア 140 台の答案では、文と文をつなぎ、少しまとまりをもった文章が出てきていることも確認できるため、A1.3 の特徴が散見されるとし、A1.2 と A1.3 のボーダーと判断した。スコア 150 台の答案では、Part B の意見展開問題において、文と文をつなげて自分の意見を表現できるようになっているため、安定して A1.3 の特徴が見られるスコア帯であると判断した。

パネリスト B は、スコア 140 台の答案は単語のスペルミスが多いために読み手が正しく単語の意味を推測しながら読まなければ内容理解が難しいものもあるが、一方で複数の文につながりを持たせて短い文章で自分の意見を表現できている答案も見られることから、A1.2 と A1.3 のボーダーと判断した。スコア 150 台の答案では、少しまとまりを持った文章を安定的に構成できているなど A1.3 の特徴が顕著であることから、A1.3 と判断した。

◆ 閾値の検討と結果

A1.2 と A1.3 の閾値：スコア 140

これらの要素を踏まえ、閾値を決定するための議論に移った。両パネリストとも、スコア 140 台の答案において、産出される英語が文章で構成されはじめるという A1.3 の特徴が確認できることから、スコア 140 が閾値として適切だと判断した。

<ライティング A2.1/A2.2>

◆ A2.1 と A2.2 の特徴

	各 CEFR-J の特徴
A2.1	<ul style="list-style-type: none"> <li>● 日常的・個人的な内容であれば、招待状、私的な手紙、メモ、メッセージなどを簡単な英語で書くことができる。</li> <li>● 文と文を “and”、“but”、“because” などの簡単な接続詞でつなげるような書き方であれば、基礎的・具体的な語彙、簡単な句や文を使った簡単な英語で、日記や写真、事物の説明文などのまとまりのある文章を書くことができる。</li> </ul>
A2.2	<ul style="list-style-type: none"> <li>● 身の回りの出来事や趣味、場所、仕事などについて、個人的経験や自分に直接必要のある領域での事柄であれば、簡単な描写ができる。</li> <li>● 聞いたり読んだりした内容（生活や文化の紹介などの説明や物語）であれば、基礎的な日常生活語彙や表現を用いて、感想や意見などを短く書くことができる。</li> </ul>

◆ 各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A2.1	～209	～199
A2.1/A2.2	210～219	200～209
A2.2	220～	210～

パネリスト A は、スコア 200 台の答案では、文脈に沿っていない単語が使用されているものや、Part B の意見展開問題において論理展開に一貫性が欠けるものが見られるとし、A2.1 の特徴が見られると述べた。一方、スコア 210 台の答案では、使用されている語彙・文法に誤りが見られるものの、論理展開に一貫性が見えはじめることから、A2.2 の特徴が現れているとし、このスコア 210 台を A2.1 と A2.2 のボーダーと判断した。スコア 220 台の答案では、安定して論理展開に一貫性が見られるため、A2.2 と判断した。

パネリスト B は、スコア 200 台の答案は、Part B の意見展開問題において、基本的な文法を用いて文章を書いているものの、導入文と結論文が一致していないなど、論理展開が整理されていない特徴があり、A2.1 と A2.2 の両方の特徴が見られることから、スコア 200 台を A2.1 と A2.2 のボーダーと判断した。スコア 210 台の答案では、“First”、“Second”など

を使用し、文章全体の構成が整理されている特徴が安定して見られるため、A2.2 と判断した。

◆ 閾値の検討と結果

A2.1 と A2.2 の閾値：スコア 210

これらの要素を踏まえ、閾値を決定するための議論に移った。パネリスト B は、スコア 200 台の答案の中の「先生に尋ねる」を“ask a teacher”ではなく“hear a teacher”と表現した答案を例に挙げ、意味理解を阻害する誤りを含む答案があるため、A2.1 に留まる答案も見られることを確認し、スコア 200 が A2.1 と A2.2 の閾値として適切ではないと判断した。スコア 210 台の答案では、パネリスト A は、平易な語句を用いて、意味理解を阻害するような誤りを含むことなく、概ね適切に自分の主張を書くことができることから、A2.2 の特徴が見えると改めて判断した。両者の意見を総合し、スコア 210 が A2.1 と A2.2 の閾値として適切だと判断した。

<ライティング B1.1/B1.2>

◆ B1.1 と B1.2 の特徴

	各 CEFR-J の特徴
B1.1	<ul style="list-style-type: none"> <li>● 自分に直接関わりのある環境（学校、職場、地域など）での出来事を、身近な状況で使われる語彙・文法を用いて、ある程度まとまりのあるかたちで描写することができる。</li> <li>● 身近な状況で使われる語彙・文法を用いれば、筋道を立てて、作業の手順などを示す説明文を書くことができる。</li> </ul>
B1.2	<ul style="list-style-type: none"> <li>● 新聞記事や映画などについて、専門的でない語彙や複雑でない文法構造を用いて、自分の意見を含めて、あらすじをまとめたり、基本的な内容を報告したりすることができる。</li> <li>● 物事の順序に従って、旅行記や自分史、身近なエピソードなどの物語文を、いくつかのパラグラフで書くことができる。また、近況を詳しく伝える個人的な手紙を書くことができる。</li> </ul>

◆ 各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
B1.1	249	249
B1.1/B1.2	250～259	250～259
B1.2	260～	260～

パネリスト A は、スコア 250 台の答案は、複雑な構造の文法を使用しているという点で、B1.2 の特徴が一部見られるとした。一方、その複雑な文法を使いこなせていないために読

み手が理解しづらい部分があったり、主張の整理が十分でなかったりするなど、B1.1 の特徴も残っていると判断した。スコア 260 台の答えは、複雑な文構造を用いながら、関連のある情報を複数の段落で構成し、また、段落ごとの関連性を意識しながら書くことができることから、B1.2 と判断した。

パネリスト B は、スコア 250 台の答えでは、“properly”や“occasionally”などの多様な語彙が使えているものや、使役動詞や“too ~ to ~...”のような文法を文脈に合わせて使えているものが見られるため、B1.2 の特徴が現れはじめると述べ、スコア 250 が B1.1 と B1.2 のボーダーと判断した。

#### ◆ 閾値の検討と結果

##### B1.1 と B1.2 の閾値：スコア 260

これらの要素を踏まえ、閾値を決定するための議論に移った。パネリスト A は、スコア 260 台の答えは、Part B の意見展開問題において、複雑な文構造を用いているものの語彙の用法に誤りを含むという、B1.1 の特徴がやや残っているものもあるが、B1.2 の特徴が色濃く出ているもののほうが多く、スコア 260 台から B1.2 であると述べた。パネリスト B は、スコア 260 台の答えでは、段落内のアイデアを詳細に説明するまでには至っていないものも見られるが、豊富な語彙や複雑な構造の文法を使用できている B1.2 レベルの特徴が出現していると、パネリスト A に同意した。従って、スコア 260 が B1.2 の閾値として適切だと判断した。

## 8.2. スピーキング

### CEFR の閾値検討

<スピーキング Pre-A1/A1>

#### ◆ Pre-A1 と A1 の特徴

	各 CEFR の特徴
Pre-A1	<ul style="list-style-type: none"> <li>● 名前や年齢などを単語や短い定型表現で言える。</li> <li>● “happy”や“tired”などの簡単な感情表現ができる。</li> <li>● 知っている語句を並べることができるが、自らの意図を伝えるために言語知識を「活用」できる様子は見られない。</li> </ul>
A1	<ul style="list-style-type: none"> <li>● 基礎的な語句、定型表現を用いて、限られた個人情報を伝えることができる。</li> <li>● 個人的なトピックについて、なじみのある表現を使って、質問したり答えたりすることができる。</li> </ul>



◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
Pre-A1	～59	～69
Pre-A1/A1	60～69	該当なし
A1	70～	70～

パネリスト A は、スコア 50 台は、知っている語句のみ、または同じ語句を繰り返す Pre-A1 の特徴が見られると判断した。スコア 60 台から“I like～.”などの定型表現を用いた解答、または語句をつなげて文を作りだそうと試みている解答が見られることから、スコア 60 台をボーダーのスコア帯であると判断した。そして、スコア 70 台は、知っている定型表現を応用できる解答が見られることから A1 と結論付けた。

一方、パネリスト B は、スコア 60 台は相手からの質問に対して、適切な解答を産出できていないことから Pre-A1 と判断し、スコア 70 台からは、Part B「質問を聞いて応答する問題」などで、単語や定型句などを用いて、相手に応答することができていることから A1 と判断し、ボーダーに分類されるスコア帯はないと述べた。

◆閾値の検討と結果

Pre-A1 と A1 の閾値：スコア 60

これらの要素を踏まえ、閾値を決定するための議論に移った。パネリスト A は、授業で学ぶ基礎的な定型表現を用いて英語を産出しているという A1 の特徴を持つ解答が多く見られるという理由で、改めてスコア 60 台が A1 の閾値として適切であると判断した。パネリスト B は、解答のグルーピングの際には「ボーダーはなし」という見解であった。しかし、その段階では、A1 のタスクにより近い Part B「質問を聞いて応答する問題」を主に見て判断していたことから、議論のなかでは Part C「ストーリーを英語で話す問題」、Part D「自分の意見を述べる問題」の解答もあわせて確認した。その結果、スコア 60 台のなかにも、例えば、“(Internet) is important”のように基本的な語句や定型表現を少し応用して、相手に自らの意図を伝えようとする A1 の特徴が見られると述べた。以上の検討を踏まえ、スコア 60 以上を A1 とすることで合意した。

<スピーキング A1/A2>

◆A1 と A2 の特徴

	各 CEFR の特徴
A1	<ul style="list-style-type: none"> <li>● 基礎的な語句、定型表現を用いて、限られた個人情報伝えることができる。</li> <li>● 個人的なトピックについて、なじみのある表現を使って、質問したり答えたりすることができる。</li> </ul>

A2	<ul style="list-style-type: none"> <li>● 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。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することができる。</li> <li>● <b>that</b> 節や、副詞や形容詞などを入れて、長い文を使うことができる。時制の不一致などは時々見られるが、意味理解を阻害するような語彙・文法の誤りは減少するため、内容は理解することができる。</li> <li>● 簡単な内容のストーリーならば、かろうじて全体像を伝えることはできるが、難しい内容だと、出来事の要因などを伝えられなかったり、英語の間違いが多くて理解できなかったりするため、聞き手に負荷がかかる。</li> </ul>
----	--

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1	～179	～169
A1/A2	180～189	170～179
A2	190～	180～

パネリスト A は、スコア 170 台は、質問に解答できている一方で、特に Part C においてストーリー展開を“I want”、“I get”、“I have”などのなじみのある定型句を用いて、文で話そうとしているが、文を正しく構成できず、誤りのために意味理解が阻害されてしまうという A1 の特徴が見られると判断した。スコア 180 台から、**that** 節や、“but”、“so”、“because”を用いた比較的長い文の中で自らの意図を伝えることができ、さらに三人称から始まる英文も発話の中に見られることから、A2 のボーダーと述べた。そして、意味理解を阻害する誤りが少ないスコア 190 台の解答が A2 であるとの見解を示した。

一方、パネリスト B は、スコア 160 台はタスクを理解できているが、基礎的な語句・定型表現の使用に留まっているため A1 と判断し、スコア 170 台は全体的に簡単な語句をつなげて文を作ろうとしているが、正しい文構造になっていない解答が含まれることから、A2 のボーダーと判断した。スコア 180 台になると意味理解を阻害する語彙・文法の誤りが減少するため、聞き手の負荷が少なくなることから、A2 の解答だと述べた。

◆閾値の検討と結果

A1 と A2 の閾値：スコア 180

これらの要素を踏まえ、閾値を決定するための議論に移った。その際には、両者がボーダーとしたスコア 170 台の解答とスコア 180 台の解答を比較した。パネリスト A は、単語をつなげて句を組み立てることはできているが、文を構成するまでには至っていないこと、文を構成できていたとしても意味理解を阻害する誤りを含む解答が多いことからスコア 170

台の解答は、A1 とした。文法の誤りが見られても、Part C などではストーリーを4コマ目まで触れることができていることや、図示したことが伝わる解答が存在するという理由でスコア 180 台から A2 の特徴が見られるという意見であった。さらに他の解答も追加で確認したところ、スコア 170 台は正しい文構造を使えていないことから、意味理解を阻害するものが多いことがわかり、A1 の領域を脱していないとパネリスト B は判断し、パネリスト A が述べた A1 の特徴が見られるという点に同意した。また、スコア 180 台の解答には、それ以下のスコアの解答と比較すると発話量が多い特徴が見られた。Part D において自分の意見に対する理由を述べる“because”以降の文構造が正確で、意図や意味が相手に伝わるという特徴から、これを閾値とすることに合意した。以上の検討を踏まえ、スコア 180 台を A2 と結論付けた。

<スピーキング A2/B1>

◆A2 と B1 の特徴

	各 CEFR の特徴
A2	<ul style="list-style-type: none"> <li>● 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。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することができる。</li> <li>● that 節や、副詞や形容詞などを入れて、長い文を使うことができる。時制の不一致などは時々見られるが、意味理解を阻害するような語彙・文法の誤りは減少するため、内容は理解することができる。</li> <li>● 簡単な内容のストーリーならば、かろうじて全体像を伝えることはできるが、難しい内容だと、出来事の要因などを伝えられなかったり、英語の間違いが多くて理解できなかつたりするため、聞き手に負荷がかかる。</li> </ul>
B1	<ul style="list-style-type: none"> <li>● 使用している語や文法の幅が広がる。</li> <li>● 使える語句や表現をつないで、自分の経験や夢、希望を順序立て、話を広げながら、ある程度詳しく語るすることができる。</li> <li>● 短い具体例を入れるなど、ある程度詳しく意見や理由を言うことはできるが、内容は深められておらず、発話している情報量が少ない。</li> <li>● 抽象的な意見を言う場面で適切な語が選べていないなど、自分の言いたいことを的確に表現できていない部分がある。</li> <li>● ある程度順序立てて情報を整理して話すことができる。</li> <li>● 身近なトピックであれば、適切な語を使ってわかりやすく説明できるが、社会的なトピックに対する意見や理由をメモなしで話す場合には、適切な語や文法が選べていないなど、自分の言いたいことを的確に表</li> </ul>

	現できていない部分がある。
--	---------------

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A2	～259	～249
A2/B1	260～269	250～269
B1	270～	270～

パネリスト A は、スコア 250 台は、基本的な定型表現を使用し、身の回りのことについて説明できるが、社会的な内容になると、英語の間違いが多く、聞き手に内容や意図を伝えることができていない特徴が多く見られることから A2 と判断した。スコア 260 台は、“but”、“however”、“in the end”、“then”などを使用し、聞き手に話の展開が明確に伝わるような語彙や文法の幅の広がりが見られること、また、発話内容をある程度整理できていることで、意図することを理解してもらえるように伝えられているという B1 の特徴が見えはじめることから、ボーダーであると述べた。スコア 270 台からは、誤りも見られるものの、関係代名詞や、形式主語構文・使役構文などを用いて、自分の意見や理由を効果的に表現できている、という特徴が見られることから B1 と判断した。

一方、パネリスト B は、スコア 250～269 をボーダーと判断した。全体的に平易な語句を組み立てることができ、文構造の正確性はあがってきているが、産出した英文量は限られていることからボーダーと判断。スコア 270 台になると、自分の主張を具体例を用いて深めることに課題があるが、それでも使用している語句や文法に広がりが見られることや、情報を順序立てながら述べられていることから、B1 レベルの解答と判断した。

◆閾値の検討と結果

A2 と B1 の閾値：スコア 260

これらの要素を踏まえ、閾値を決定するための議論に移った。パネリスト A は、A2 レベルの語句や表現を正確に使用でき、情報を順序立てながら自分の考えを述べている解答が散見されるという理由でスコア 260 台から B1 の特徴が顕著になると判断した。パネリスト B は、スコア 270 台と比較すると、スコア 260 台は的確に表現できていない部分もあるが、自分の経験に基づき、ある程度詳細な内容を含めて話しているという理由で B1 の特徴が現れているとした。「GTEC」の試験は準備時間が長いもので 1 分、かつメモもとれないことから、即興に近いタスクであり、その点を考慮し、スコア 250 台の解答音声とスコア 260 台の解答音声を流しように着目しながら聞き比べた。スコア 260 台になると少しためらいや言いよどみが確認できる話し方ではあるが、聞き手に伝わるように話している、文脈に合わせて適切に解答に必要な文を応用できている特徴に着目して、スコア 260 台を B1 レベルとすることで合意した。

<スピーキング B1/B2>

◆B1 と B2 の特徴

	各 CEFR の特徴
B1	<ul style="list-style-type: none"> <li>● 使用している語や文法の幅が広がる。</li> <li>● 使える語句や表現をつないで、自分の経験や夢、希望を順序立て、話を広げながら、ある程度詳しく語るができる。</li> <li>● 短い具体例を入れるなど、ある程度詳しく意見や理由を言うことはできるが、内容は深められておらず、発話している情報量が少ない。</li> <li>● 抽象的な意見を言う場面で適切な語が選べていないなど、自分の言いたいことを的確に表現できていない部分がある。</li> <li>● ある程度順序立てて情報を整理して話すことができる。</li> <li>● 身近なトピックであれば、適切な語を使ってわかりやすく説明できるが、社会的なトピックに対する意見や理由をメモなしで話す場合には、適切な語や文法が選べていないなど、自分の言いたいことを的確に表現できていない部分がある。</li> </ul>
B2	<ul style="list-style-type: none"> <li>● ある程度なじみのあるトピックならば、新聞・テレビ・インターネットなどで知り得たニュースの要点について議論することができる。</li> <li>● 母語話者同士の議論に加われないこともあるが、自分が学んだトピックや自分の興味や経験の範囲内のトピックなら、抽象的なトピックであっても、議論できる。</li> <li>● ある視点に賛成または反対の理由や代替案などを挙げて、事前に用意されたプレゼンテーションを聴衆の前で流ちょうに行うことができ、一連の質問にもある程度流ちょうに対応ができる。</li> <li>● 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ができる。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
B1	～319	～309
B1/B2	該当なし	310～319
B2	320～	320～

パネリスト A は、スコア 310 台は、ある程度順序立てて情報を整理して話すことはできているが、社会的な内容に対する意見や理由を適切な語彙や文法を用いて表現する力が不足している特徴が見られることから B1 と判断した。スコア 320 台の解答からは、“until”など時の副詞節を導く接続詞や、関係副詞の“where”を用いるなど、使用語句や文法が豊かになり、また、複雑な文構造を取り入れながら社会的な内容に対応した主張を明確に述べるこ

とができているという特徴が見られることから B2 と判断し、ボーダーはないという結論であった。

一方、パネリスト B は、スコア 300 台は、発話する情報量は増えているものの、主張の展開が浅く、また適切でない語彙・文法の誤りを含むなどの理由で、自分の考えを的確に表現できていない解答が多いことから、B1 と判断した。スコア 310 台は、抽象的な内容を伝えようとしている箇所で、適切な表現を使用できていない面もあるが、要点は伝わるように発話できているという特徴から、B2 の要素が見えはじめているとし、ボーダーと判断した。スコア 320 台からは正確に要点を伝えることができ、論拠を並べながら、自分の意見と理由を述べるができている特徴から B2 と判断した。

#### ◆ 閾値の検討と結果

##### B1 と B2 の閾値：スコア 320

これらの要素を踏まえ、閾値を決定するための議論に移った。両パネリストは、スコア 310 台の解答音声とスコア 320 台の解答音声を聞き比べ、スコア 310 台とスコア 320 台で明らかに流ちょうさに差があることに同意した。スコア 310 台は、目立ったためらいや言いよどみが見られる、難しい語句や文法を正確に使用できていない解答が目立つという理由で B1 であると判断。スコア 320 台は、長いためらいや言いよどみがなく、“without～”で条件を付加するなど難しい文法を正確に使用でき、頭の中で話の内容を組み立てながら、途切れずに話を続けることができているという理由から B2 と判断した。以上の検討を踏まえ、スコア 320 台を B2 とすることで合意した。

#### CEFR-J の閾値検討

<スピーキング A1.1/A1.2>

#### ◆ A1.1 と A1.2 の特徴

各 CEFR-J の特徴	
A1.1	<ul style="list-style-type: none"> <li>● 基礎的な語句、定型表現を用いて、限られた個人情報を伝えることができる。</li> <li>● 個人的なトピックについて、なじみのある表現を使って、質問したり答えたりすることができる。</li> </ul>
A1.2	<ul style="list-style-type: none"> <li>● 前もって発話することを用意したうえで、限られた身近なトピックについて簡単な語や基本的な句、限られた構文を用い、簡単な意見を言うことができる。</li> <li>● 簡単な語を使った短い文や定型表現を使って、アイデアを伝えることができるが、語彙や文法に誤りが多いため、適切に伝わらないところがある。</li> </ul>

	<ul style="list-style-type: none"> <li>● 複数のアイデアを伝えることができるが、一つひとつのアイデアが短くシンプルであり、アイデアのつながりが弱い。</li> <li>● “because”を正しく使い、理由を述べることができる。</li> </ul>
--	--

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1.1	～89	～89
A1.1/A1.2	90～99	該当なし
A1.2	100～	90～

パネリスト A は、スコア 80 台は、基礎的な定型表現を使用して解答できていることから A1.1 と判断した。スコア 90 台は、文構造にまとまりが見られるようになり、また、“What do you usually do～?”などの質問に対して“I usually do～.”と出だしのフレーズを言えるが、一方で出だしに続く内容が伝わらないような特徴が見られることからボーダーだと述べた。スコア 100 台からは 2 文以上の短い発話がつながり、“I play”、“I like it”、“It’s exciting”などの比較的なじみのある語や表現を使って、アイデアを伝えることができることから A1.2 と判断した。

一方、パネリスト B は、スコア 80 台は語句の羅列や“Carrot get. Look rabbit. Carrot rabbit eat.”などと限定的な単語の羅列で解答していることから A1.1 であるとの見解を示した。スコア 90 台は、Part D の解答で文を生成しながら意見が言えているという特徴から A1.2 と判断した。

◆閾値の検討と結果

A1.1 と A1.2 の閾値：スコア 90

これらの要素を踏まえ、閾値を決定するための議論に移った。議論のなかで、両パネリストは、スコア 80 台は、語句の羅列や限定的な発話にとどまっているため A1.1 であること、スコア 90 台からは“because”を正しく使い、理由を述べることができていることや、スコア 80 台と比較すると、意味が理解できる文の割合が増え、いくつか発話をつなげようと試みながら、文を生成して意見を伝えられるようになっていることで合意した。よってスコア 90 台を閾値として結論付けた。

<スピーキング A1.2/A1.3>

◆A1.2 と A1.3 の特徴

	各 CEFR-J の特徴
A1.2	<ul style="list-style-type: none"> <li>● 前もって発話することを用意したうえで、限られた身近なトピックについて簡単な語や基本的な句、限られた構文を用い、簡単な意見を言うことができる。</li> <li>● 簡単な語を使った短い文や定型表現を使って、アイデアを伝えることができるが、語彙や文法に誤りが多いため、適切に伝わらないところがある。</li> <li>● 複数のアイデアを伝えることができるが、一つひとつのアイデアが短くシンプルであり、アイデアのつながりが弱い。</li> <li>● “because”を正しく使い、理由を述べることができる。</li> </ul>
A1.3	<ul style="list-style-type: none"> <li>● 前もって発話することを用意したうえで、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、限られた構文を用い、複数の文で意見を言うことができる。</li> <li>● 定型表現を脱し、オリジナルの文で伝えることができている。しかし、語彙選択や文法に誤りがあり、意味理解を阻害するものもある。</li> <li>● “because”や“so”のような接続詞を使い、複数の短い文が意味的なつながりを持ち、アイデアの羅列ではなくなってくる。しかし、描写が粗いために、うまく伝えられなかったり、意見や理由もシンプルな内容にとどまっていたりする。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A1.2	～129	～129
A1.2/A1.3	130～139	130～139
A1.3	140～	140～

パネリスト A は、スコア 120 台は、“because”を使用して理由を述べることができるが、複数の短い文で意味的につながりのある構成ができるという A1.3 の特徴がまだ見られないことから A1.2 だと述べた。スコア 130 台は、平易な語や基礎的な句を組み合わせる意見を言うというような A1.2 の特徴がまだ見られるが、短い発話をいくつかつなげることができる解答もあることから A1.3 とのボーダーと判断。スコア 140 台からは“because”や“so”のような接続詞を使用して、短い発話をいくつかつなげることができることから A1.3 と判断した。

一方、パネリスト B は、スコア 120 台は、平易な語や基本的な句、限られた構文を用いて単純な意見を産出しているという特徴から A1.2 と判断した。スコア 130 台は、定型表現



を脱し、自分の考えを交えて複数の文で意見を言うことができるという A1.3 の特徴をもつ解答も散見されることからボーダーと判断した。スコア 140 台からは様々な動詞を用いた表現を使える特徴や、与えられた課題文に対する意見や理由を、定型表現や基本的な接続詞を用いて簡潔に言える特徴が見られることから A1.3 と述べた。

#### ◆ 閾値の検討と結果

##### A1.2 と A1.3 の閾値：スコア 130

これらの要素を踏まえ、閾値を決定するための議論に移った。パネリスト A は、課題文に関連する単語と、定型表現や基礎的な構文を組み合わせて自分が伝えたい意見・理由を産出できるという特徴が見られるという理由でスコア 130 台で A1.3 の特徴が見えてくるとした。パネリスト B は、自分のしたいことを自分の意見として相手に伝え、また、定型表現や基礎的な構文を正確に組み合わせて理由も伝えることができる特徴が見られるという理由でスコア 130 台から A1.3 というパネリスト A の意見に同意した。また、両パネリストは、スコア 120 台の解答音声と、スコア 130 台の解答音声を聞き比べ、スコア 130 台になると簡単な内容ではあるものの、言いよどみ少なく意見と理由を述べることができていることを確認した。以上の検討を踏まえ、スコア 130 台を A1.3 の閾値とすることで合意した。

#### <スピーキング A2.1/A2.2>

##### ◆ A2.1 と A2.2 の特徴

	各 CEFR-J の特徴
A2.1	<ul style="list-style-type: none"> <li>● 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。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することができる。</li> <li>● <b>that</b> 節や、副詞や形容詞などを入れて、長い文を使うことができる。時制の不一致などは時々見られるが、意味理解を阻害するような語彙・文法の誤りは減少するため、内容は理解することができる。</li> <li>● 簡単な内容のストーリーならば、かろうじて全体像を伝えることはできるが、難しい内容だと、出来事の要因などを伝えられなかったり、英語の間違いが多くて理解できなかったりするため、聞き手に負荷がかかる。</li> </ul>
A2.2	<ul style="list-style-type: none"> <li>● 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について短いスピーチをすることができる。</li> <li>● 身近なトピックであれば、まとまりのある短い話をするすることができる。</li> </ul>

	<ul style="list-style-type: none"> <li>● ストーリーにより詳細に説明を加えることができ、聞き手の負荷が減るが、稚拙な表現がやや残る。</li> <li>● シンプルではあるが適切な語句や文を使って適切に説明することができる。</li> </ul>
--	---

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
A2.1	～219	～219
A2.1/A2.2	220～239	該当なし
A2.2	240～	220～

パネリスト A は、スコア 210 台は、基本的な語句や文を使った構文の中で与えられた課題文に関連する副詞や形容詞を用いながら主張を発話できているが、英語の誤りを含む部分も多くあり、聞き手に負荷がかかるという A2.1 の特徴が見られると述べた。スコア 220 台から、与えられた課題文についてまとまりのある内容を、意味理解を阻害する誤りが少ない状態で伝えることができているという特徴が見られることからボーダーと判断した。そして社会的な事柄に関するテーマでも、具体的な経験などを含めることができているスコア 240 台の解答が A2.2 という見解を示した。

一方、パネリスト B は、スコア 210 台は、主張したいであろうことを推測しながら聞く必要がある解答が多数をしめることから A2.1 と判断した。スコア 220 台からは、適切な語句や文を使って物事を描写できており、聞き手がストーリーの内容や展開をある程度イメージすることができることから、A2.2 と判断した。

◆閾値の検討と結果

A2.1 と A2.2 の閾値：スコア 230

これらの要素を踏まえ、閾値を決定するための議論に移った。閾値の判断をするうえで、パネリスト B が A2.2 の閾値として適切であると判断したスコア 220 台の解答音声と、パネリスト A が A2.2 の閾値として適切であると判断したスコア 230 台の解答音声を聞き比べたところ、Part C「ストーリーを英語で話す問題」において、各コマの出来事について詳しい説明を付加しながら話せるようになることがスコア 230 台の解答の特徴であることが確認できた。

パネリスト A は、与えられた課題文に関連した定型表現、副詞や形容詞を用いて、まとまりのある意見と理由を伝えることができている解答が見られるという理由からスコア 230 台が A2.2 の閾値として適切であると判断した。パネリスト B は、社会的な事柄に関する課題文でも適切な語句や文を使って説明でき、主張が伝わりやすくなることで聞き手の負荷が減るという理由からスコア 230 台を A2.2 の閾値とすることで合意した。

<スピーキング B1.1/B1.2>

◆B1.1 と B1.2 の特徴

	各 CEFR-J の特徴
B1.1	<ul style="list-style-type: none"> <li>● 使える語句や表現をつないで、自分の経験や夢、希望を順序立て、話を広げながら、ある程度詳しく語るができる。</li> <li>● B1.2 に比べて流ちょうさが下がる。</li> <li>● 短い具体例を入れるなど、ある程度詳しく意見や理由を言うことはできるが、内容は深められておらず、発話している情報量が (B1.2 と比較すると) 少ない。</li> <li>● 抽象的な意見を言う場面で適切な語が選べていないなど、自分の言いたいことを的確に表現できていない部分がある。</li> <li>● ある程度順序立てて情報を整理して話すことができる。</li> <li>● 身近なトピックであれば、適切な語を使ってわかりやすく説明できるが、社会的なトピックに対する意見や理由をメモなしで話す場合には、適切な語や文法が選べていないなど、自分の言いたいことを的確に表現できていない部分がある。</li> <li>● 使用している語や文法の幅が広がる。</li> </ul>
B1.2	<ul style="list-style-type: none"> <li>● 短い読み物か短い新聞記事であれば、ある程度の流ちょうさをもって、自分の感想や考えを加えながらあらすじや要点を順序立てて伝えることができる。</li> <li>● 自分の関心事であれば、社会の状況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、聴衆から質問ができれば相手に理解できるように答えることができる。</li> <li>● ある程度の流ちょうさで自分の意見を伝えることができる。</li> <li>● 効果的に情報を付加しながら説明することができる。</li> <li>● 語彙や文法の使い方が適切であるため、理解に影響が出る誤りはほぼ見られない。</li> <li>● ストーリーを描写する際、出来事だけでなく、当事者の感情も入れながら、効果的に伝えることができる。</li> </ul>

◆各パネリストによるグルーピングの結果

	パネリスト A	パネリスト B
B1.1	～279	～279
B1.1/B1.2	280～299	280～289
B1.2	300～	290～

パネリスト A は、スコア 270 台は、言いたいことは伝わるものの的確に表現しきれてい

ない部分があるという特徴から B1.1 だと述べた。スコア 280～290 台では、B2 と比べると短い発話ではあるものの、基礎的な語句や表現を適切に用いて、ある程度流ちょうに自分の意見を伝えることができているが、内容を深められていないという B1.1 の特徴を持つ解答と、4 コマのストーリーを分かりやすく丁寧に描写でき、語彙や文法も正しく使用できているという B1.2 の特徴を持つ解答が混在しているという理由で、ボーダーと判断した。スコア 300 台から、使用している語句や表現が豊かになり、また、複雑な文構造を使用しながら、効果的に情報を付加して説明できているという特徴が安定して見られることから B1.2 と判断した。

一方、パネリスト B は、スコア 270 台は、ある程度詳しく意見と理由を伝えることができるが、スコア 280 台と比較すると発話量が少なく、結果として含まれる情報量も少なくなっていると分析した。また、同じ内容を繰り返したり、たどたどしい伝え方をしたりするなどの特徴が見られることから、B1.1 の解答と判断した。スコア 280 台は、ある程度の流ちょうさをもって、自分の感想や考えを加えながらあらすじや要点を順序立てて伝えることができている解答が散見されると判断し、ボーダーと述べた。スコア 290 台からは語彙や文法を適切に使いながら要点を伝えることができ、さらに内容理解を阻害する誤りはほぼ見られないという特徴から B1.2 と判断した。

#### ◆ 閾値の検討と結果

##### B1.1 と B1.2 の閾値：スコア 290

これらの要素を踏まえ、閾値を決定するための議論に移った。両パネリストは、スコア 280 台の解答音声とスコア 290 台の解答音声を聞き比べた。スコア 280 台は、基礎的な文構造を用いて流ちょうな英語で自分の意見と理由を述べているが、内容を深められておらず、発話量も少ない解答が目立つという理由で B1.1 と判断した。スコア 290 台は、ある程度の流ちょうさを保ちながら、自分の主張に複数の情報を付加したり、学習してきた語句や表現の中から、各タスクに応じて最も効果的な語句や表現を選択して発話したりことができ、それによって発話内容の質を高めることに成功しているという理由で B1.2 と言えることを改めて確認した。以上の検討を踏まえ、スコア 290 台を B1.2 とすることで合意した。

## 9. 結果

各 CEFR・CEFR-J レベル別の検討を経て、「GTEC」のライティング・スピーキングそれぞれの閾値は下記のようになった。

【表 6. 調査前・調査後のCEFR-Jの閾値】

CEFR	CEFR-J	調査前		調査後	
		ライティング	スピーキング	ライティング	スピーキング
B2	B2	300	320	290	320
B1	B1.2	260	300	260	290
	B1.1	240	280	230	260
A2	A2.2	220	220	210	230
	A2.1	190	190	190	180
A1	A1.3	140	140	140	130
	A1.2	100	100	100	90
	A1.1	60	80	70	60
Pre-A1	Pre-A1	～59	～79		

## 10. 考察

本研究は、2016年度から2017年度に行った「GTEC」スコアとCEFRレベルとの関連付け調査、及び2018年度に行ったCEFR-Jレベルとの関連付け調査から3年以上が経過していること、また、2019年度に問題構成と採点基準が改訂されたことを受けて計画された。2019年度から現在までで、検証するに十分な受検者のデータが蓄積されたと考え、閾値の見直しを行った。

### <採用した分析手法>

今回のスタンダードセッティングでは The Bookmark Method と The Body of Work Method を組み合わせ、IRT で算出された能力値という客観的な情報に基づいて閾値の判断を行うことができた。パネリストに CEFR や CEFR-J の見識が深い有識者を選定したこと、また有識者が過去に CEFR のスタンダードセッティングを行った経験を有していたことが、円滑な閾値の設定に寄与した。スタンダードセッティングの前には、Familiarization で CEFR、及び CEFR-J のレベルを確認した。2018年度に行ったスタンダードセッティングで作成した、「GTEC」の解答を利用した CEFR-J 別の特徴をまとめた資料を使って目線合わせをしたことで、パネリストや参加者の間でそれぞれのレベルの認識に大きな齟齬が生まれることなく、閾値を決定することができた。

#### <当日の進め方>

最初に、パネリストがスコアの付いた受検者の解答を確認しながら、閾値の下のレベルのグループ、閾値となるボーダーラインのグループ、閾値の上のレベルのグループ、のように3つのグループに分ける作業を行った。次の工程でそれぞれのパネリストが3つのグループに分けた結果とその判断に至った理由を共有し、閾値となるボーダーラインのスコア帯が異なった場合に両者で議論をし、合意する閾値を導き出した。最初の3つのグループに分ける工程では、パネリストの個人ワークとしたことで、他者の意見に誘導されず、パネリスト一人ひとりの知見に基づいた意見を引き出すことができた。協議の工程では、両者の活発な議論の末、参加者が納得したスコアを閾値に定めることができた。パネリスト間でボーダーラインのグループのスコア帯に差異があった場合、判断根拠を説明し合い、前後のレベルの解答を参照したり、受検者の音声を確認したりすることで、次第に意見が合致していく様子が見られた。パネリストは、語彙、言語テスト、第二言語習得等、幅広い専門分野の見識、および教授経験を持っているため、それぞれの専門分野の知見に基づく意見を交換し合うことで、各パネリストの中にあるイメージが徐々にすりあわされていった。

#### <使用データ>

2018年度に調査を行った際は、「GTEC」の本番試験1回分の問題、解答データを使用した。今回は複数の本番試験の問題、解答データを調査に活用した。実受検者である中高生が取り組んだデータのみを使用したうえ、複数の問題版の解答やデータを用いることで、特定の問題起因による解答の偏りを防ぎ、より精緻な閾値の設定ができたと言える。

#### <選定した素材の有効性>

前回の調査では、スピーキングは「発表」のタスクである Part C と Part D のみを検証素材としていたが、今回は「やりとり」も含めてスピーキングの力を判断することを目的として、Part B のパフォーマンスもブックレットに表示させて検討を行った。ただし、短い時間でイラストの情報を読み取り、すぐに解答することが求められる Part B は、特に Advanced/Basic タイプではタスクが少し複雑になる。その結果、Part C や Part D が良く解答できている受検者でも Part B で解答を誤るケースが見られたため、Part B のパフォーマンスから CEFR・CEFR-J レベルを判定することはやや難しかった。一方で、自分自身について解答するタスクが中心であり、イラストの情報量も少ない Core タイプでは、Part B の解答を CEFR 下位レベルの検討に含めることができた。

ライティングにおいては、前回の調査では、2019年度の出題改訂前だったということもあり、意見展開問題のみで検証していたことから、Pre-A1 レベルや A1 レベルの解答は語数が少なく内容がほとんど書かれていない解答が多かった。しかし、本調査では E メール

問題も含めたことで、Pre-A1 レベルや A1 レベルであってもアウトプットに明確な差が見られ、より妥当な閾値の判断につながったと考える。

なお、本調査で決定した新閾値による結果返却は、2023 年度の「GTEC」アセスメント版・検定版からを予定している。

11. 参考文献

Adam E. Wyse, Michael B. Bunch, Craig Deville, Steven G. Viger. (2013). *A Body of Work Standard-Setting Method With Construct Maps*.  
<https://journals.sagepub.com/doi/10.1177/0013164413502037>

Council of Europe. (2020). *COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT, Companion volume*. Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>

Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*. Council of Europe Publishing. <https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>

Edited by RelEX team. José Noijons, Jana Bérešová, Gilles Breton, Gábor Szabó. (2011). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): Highlights from the manual*. Council of Europe Publishing. [https://www.ecml.at/Portals/1/documents/ECML-resources/2011\\_10\\_10\\_relex\\_E\\_web.pdf?ver=2018-03-21-100940-823](https://www.ecml.at/Portals/1/documents/ECML-resources/2011_10_10_relex_E_web.pdf?ver=2018-03-21-100940-823)

Michael Zieky & Marianne Parie. (2006). *A Primer on Setting Cut Scores on Tests of Educational Achievement*. ETS

ベネッセコーポレーション. (2017). *GTEC スコアと CEFR レベル関連付け調査報告*.  
<https://www.benesse.co.jp/gtec/assets/pdf/doc-02-03.pdf>

ベネッセコーポレーション. (2018). *GTEC スコアと CEFR-J レベル関連付け調査報告*.  
<https://www.benesse.co.jp/gtec/schoolofficials/research/pdf/doc-2018-01.pdf>

投野由紀夫(編著). (2013). *英語到達度指標 CEFR-J ガイドブック*. 大修館書店.

東京外国語大学 投野由紀夫 研究室. “CEFR-J Resources”・“CEFR-J 本体: 日本語版、英語版”・“CEFR-J CAN-DO テスト”. <https://www.cefr-j.org/download>



Downing, S. M., & Haladyna, T. M (Eds.). (2008). 『テスト作成ハンドブック (Handbook of test development) : 発達した最新技術と考え方による公平妥当なテスト作成・実施・利用のすべて』 (池田央 [日本語版監訳], Trans.). 教育測定研究所.

大友賢二 (監修) (2009). 『言語テスト : 目標の到達と未到達 [Setting Performance Standards on Language Tests]』 . NPO 法人 英語運用能力評価協会 (ELPA).

大友賢二・渡部良典・伊東祐郎・法月健・藤田智子 (2012). 「言語テストの規準設定報告書 (2)」 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

## 12. 付録

### ◆ 「GTEC」の問題概要・実施時間

#### ・「GTEC」受検タイプについて

「GTEC」では、レベルによって Core、Basic、Advanced、CBT の4つの問題タイプを提供している。学年による受検タイプの制限はないが、下記に受検に適した目安の時期を記載する。

中1	中2	中3	高1	高2	高3
					GTEC CBT
				GTEC Advanced	
			GTEC Basic		
	GTEC Core				

#### ・各問題タイプの試験概要

今回の CEFR・CEFR-J の関連付け調査で使用了問題タイプ (Core、Basic、Advanced) の試験概要は下記のとおりである。

	Core	Basic	Advanced
測定技能	Reading・Listening・Writing・Speaking		
測定方法	紙：(マーク式) Reading・Listening,(記述式) Writing タブレット：Speaking		
解答時間	約92分 Reading 約32分/Listening 約20分 Writing 約25分/Speaking 約15分	約110分 Reading 約45分/Listening 約25分 Writing 約25分/Speaking 約15分	
4技能上限スコア (測定可能CEFRレベル)	840 (~A2)	1080 (~B1)	1280 (~B2)
語いレベル	中学の学習指導要領の 範囲の語い	中学～高校1年の 学習指導要領の範囲の語い	高校の学習指導要領の 範囲の語い

◆ 「GTEC」のライティング・スピーキング問題の出題構成  
・ライティング

	Core		Basic		Advanced	
	設問数	時間(分)	設問数	時間(分)	設問数	時間(分)
合計	2	約25	2	約25	2	約25
A Eメール問題	1	5	1	5	1	5
B 意見展開問題	1	20	1	20	1	20

※パート別の時間は、秒を省略して記載しているため、合計時間は約〇〇分と表示しています。

■出題の特長

1 実践的な出題形式

Eメールに返信する形で記述する問題、与えられたテーマに対して自分の考えを自由記述形式で表現する問題を出題します。いずれも海外での実生活を想定した実践的な出題が特長です。

2 読み手に解答内容や主張がどれだけ伝わるかを徹底して測定

Writingでは「読み手に解答内容や主張が伝わるか」が最も重要な評価ポイントとなります。次に、「いかにわかりやすく、効果的に伝えられているか」を語い・文法などの正しきや、構成の仕方などで評価します。

■出題例 ● Advanced

Part A Eメール問題

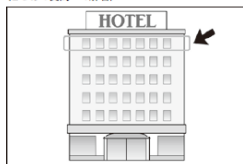
相手からのEメールに返信する実践的な形式の出題で、イラストに沿った内容の返信メールを作成します。

あなたはカナダに留学中です。あなたは、部活の合宿で宿泊する予定のホテルからメールを受け取りました。返信メールを、単語のみではなくできるだけ文で書きなさい。

(1つめの質問への解答)



(2つめの質問への解答)



※●Basic ●Core ●Advancedと同様の形式で出題します。

From	The Laughing Cat Hotel
To	Kaoru
Subject	Questions regarding your stay

Hello,

Thank you again for choosing the Laughing Cat Hotel. We just have two additional questions about your stay. First, does anyone in your group have any special meal requirements? Second, do you have any requests about your rooms?

We look forward to hearing from you.

Best regards,  
John Cave

From	Kaoru
To	The Laughing Cat Hotel
Subject	RE: Questions regarding your stay

Hello,

Sincerely,  
Kaoru

Part B 意見展開問題

Advancedタイプでは、社会との接点を通して、個人の経験や他の事例をもとに自分の意見と理由を述べます。

あなたは留学先の授業でエッセーを書くことになり、以下のテーマを選びました。このテーマを読んでいない人にも伝わるようにエッセーを書きなさい。

エッセーのテーマ

若者の活字離れが問題となっていることについて、あなたはどのように思いますか。あなたの意見とその理由を書きなさい。

(参考)



※●Basic ●Core ●Advanced同様、参考となるイラストが2枚つきます。

・スピーキング

	Core		Advanced	
	設問数	時間(分)	設問数	時間(分)
<b>合計</b>	<b>8</b>	約15	<b>8</b>	約15
A Reading Aloud ~音読~	2		2	
B Listening and Responding ~質問を聞いて応答する~	4		4	
C Telling a Story ~ストーリーを英語で話す~	1		1	
D Expressing Your Opinion ~自分の意見を述べる~	1		1	

※Speakingについては、AdvancedタイプとBasicタイプが同一の問題です。  
 ※機材の準備時間は含まれておりません。試験開始前に別途準備時間が必要となります。

■出題の特長

**1** リアルな場面設定で、音読から意見展開まで幅広い力を測定  
 音読を含めた全てのパートで、日常生活や学校、留学場面などの設定による出題となっています。日本の中高生に適した出題内容で、話す力を多角的に測定します。

**2** 相手に、自身の主張・情報がどれだけ伝わるかを徹底して測定  
 Speakingの採点では、自身の意見、主張、または情報をいかに相手に伝えられているかという点を大事にしています。また、海外の採点拠点で英語話者の視点による採点を行っているため、英語話者に対して、与えられた情報を元に、自身の意見・主張を述べることができるかを確認することができます。

■出題例 ● Advanced ● Basic

**Part A** 音読 準備時間 各30秒 解答時間 各40秒

問題数 2問 | Reading Aloud

英文を読み上げる形式の出題で、状況や英文を理解した上で、正確な発音と適切な流ちょうさで音読ができるかどうかを診断します。

**Part A 出題画面イメージ**

あなたはインターナショナルスクールの中生です。あなたは、朝の校内放送でクラス大会について案内することになりました。次の英文を読み出し、読んでください。  
 (準備時間30秒/解答時間40秒)

**問題文**

Good morning students. Are you looking for a challenge? Why not sign up for Team Quiz Night, to be held on May 10th at 6 pm in the school hall? Gather your friends, and get ready for some strange and interesting questions. Bring food to share and try to beat the teachers' team. Good luck!

**採点観点**  
 発音・流ちょうさ (0, 1, 2, 3の4段階)

**Part B** 質問を聞いて応答する 準備時間 各10秒 解答時間 各15秒

問題数 4問 | Listening and Responding

図示された情報を読み取り、それに関する質問を聞き取った上で、適切に応答する力があるかどうかを診断します。  
 読まれる文: What are you going to do this Saturday?

**Part B 出題画面イメージ**

図表上の情報をもとに、質問に英語で答えてください。  
 (準備時間10秒/解答時間15秒)

**図表**

	SAT	SUN
9:00		
12:00		12:00-13:00 dinner
15:00	grandparents Picnic	
18:00	gym	18:00- concert

**採点観点**  
 Goal Achievement (質問に対して明確に回答できているかの観点で2段階評価)

**Part C** ストーリーを英語で話す 準備時間 30秒 解答時間 60秒

問題数 1問 | Telling a Story

日常的な出来事について、話の流れを踏まえて相手に伝わるように状況を説明する力を診断します。

**Part C 出題画面イメージ**

あなたは、先日ある少年が経験したことを、留学生の友だちに話すことになりました。相手に伝わるように英語で話してください。  
 (準備時間30秒/解答時間60秒)

**図表**

**採点観点**  
 Goal Achievement (1コマめ~4コマめのそれぞれの内容を明確に説明できているかの観点で、2段階評価)。加えて、語い・文法 (0, 1, 2, 3, 4の5段階)、発音・流ちょうさ (0, 1, 2, 3の4段階) の2観点

**Part D** 自分の意見を述べる 準備時間 60秒 解答時間 60秒

問題数 1問 | Expressing Your Opinion

身近で社会的なテーマに対して、自分の意見とその意見をサポートする理由が言えているかを診断します。

**Part D 出題画面イメージ**

あなたは英語の授業で、次のテーマについて発表することになりました。自分の考えを述べ、その理由を詳しく具体的に説明してください。聞いてる人に伝わるように話してください。  
 (準備時間1分/解答時間1分)

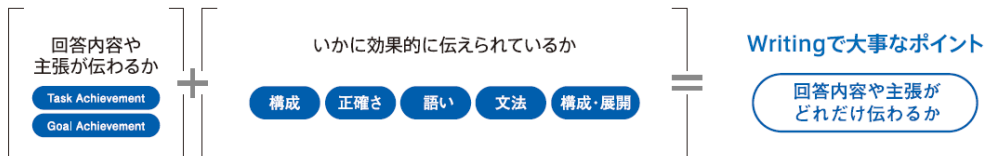
Schools shouldn't allow students to use their cell phones while at school. What do you think about this? State your opinion and give at least one reason with an example or explanation.

**採点観点**  
 Goal Achievement (与えられたテーマに対して、「意見」「理由」を明確に説明できているかの観点で、「意見」は2段階評価、「理由」は3段階評価)。加えて、語い・文法 (0, 1, 2, 3, 4の5段階)、発音・流ちょうさ (0, 1, 2, 3の4段階) の2観点

## ◆「GTEC」の採点基準（要約版）

### ・ライティング

#### ■Writing観点別の採点基準



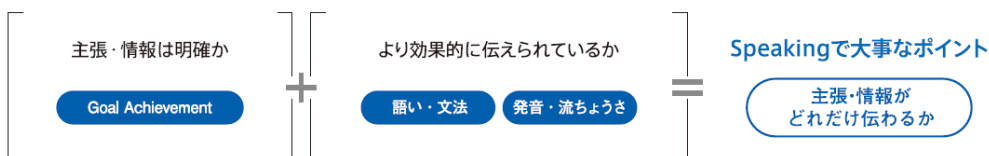
Task Achievement Part A	タスク	Writing			
		0点	1点	2点	3点
Eメール内の質問に対して、イラストに沿った内容を適切に回答できている。	構成	英文が書かれていない。または、質問された内容に文で十分に回答できていない部分がある。	質問された内容に文で十分に回答できていない部分がある。	質問された内容に文で十分に回答できている。	質問された内容に文で十分に回答できている。メールとしても自然な文面を構成できている。
	正確さ	英文が書かれていない。または、語いや文法の誤りにより、意図した内容が伝わらない。	語いや文法の誤りにより、意図した内容の理解が困難な部分がある。	語いや文法の誤りがあっても、意図した内容を伝えることができている。	語いや文法を適切に用いて、意図した内容を伝えることができる。

Goal Achievement Part B	意見	理由	Writing					
			0点	0.5点～1.5点	2点～3.5点	4点～5.5点	6点～8点	
課題に対する自分の意見を伝えることができている。 自分の意見をサポートする理由や具体例などを伝えることができている。	意見	理由	語い	英文が書かれていない。または、全体を通して出題のテーマから外れたことが書かれている。	使える語いが基本的なものに限られている。または、語いの誤りが多く、意図した内容が伝わらない。	使える語いが限られている。または、意味理解を妨げるような誤りが見られ、意図した内容が伝わりにくい部分が多くある。	必要な語いを使って、簡単な内容を表現することができている。誤りがあっても、意図した内容を伝えることができている。	さまざまな語いを使って、自分の考えや物事を詳しく説明することができる。一部誤りがあっても、意図した内容を十分に伝えることができている。
			文法	英文が書かれていない。または、全体を通して出題のテーマから外れたことが書かれている。	定型表現を使うことや、いくつかの単語を組み合わせることはできる。	使える文法が限られている。または、意味理解を妨げるような誤りが見られ、意図した内容が伝わりにくい部分が多くある。	基本的な文法を使うことができる。誤りがあっても、意図した内容を伝えることができている。	基本的な文法に加えて、複雑な文法を使うことができる。一部誤りがあっても、意図した内容を十分に伝えることができている。
			構成・展開	英文が書かれていない。または、全体を通して出題のテーマから外れたことが書かれている。	単語や文は書かれているが、内容のつながりが見られない。	いくつかのアイデアが書かれているが、内容のつながりが見えにくい。	アイデアのつながりが見えにくい部分があるが、書かれている内容にある程度まとまりが見られる。	書かれている内容のつながりがわかりやすく、全体としてよく構成されている。

それぞれの評価を、IRTを用いた統計処理にてスコアを算出します。それぞれの評価項目に配点がある訳ではありません。  
※例えば、正答率の高い評価観点で誤ってしまうと、スコアへのマイナス影響は大きくなります。

### ・スピーキング

#### ■Speaking観点別の採点基準



Goal Achievement Part C	意見	理由	Speaking					
			0点	1点	2点	3点	4点	
各設問の問いかけに適切に内容を伝えることができている。 各コマの内容を伝えることができている。 意見を伝えることができている。	意見	理由	語い・文法	出題意図から明らかに外れている。／英語ではない、あるいは、英語として通じない。／力を測るために十分な量の発話がない。	使用する語いや文法が非常に限られているが、簡単に具体的なことであれば、語や句をつなげながら表現することができている。	限られた語いや文法で、簡単な描写を羅列したり、簡単にアイデアをつなげたりすることができている。	複雑な内容を説明するときに誤りが見られることもあるが、語いや文法を適切に使用し、アイデアを順序立ててつなげることができている。	豊富で幅広い語いや文法を、柔軟に使用することができている。アイデア間の関係性を整理して示すことができ、聞き手に効果的に内容を伝えられている。
			発音・流ちょうさ	出題意図から明らかに外れている。／英語ではない、あるいは、英語として通じない。／力を測るために十分な量の発話がない。	全体を通して、発音やイントネーション、リズムの誤り、また、不自然な沈黙や言いどまりが多いため、聞き手に大きく負担がかかるが、簡単な単語や語句の強調は正しい。	部分的な発音やイントネーション、リズムの誤り、また、不自然な沈黙や言いどまりのため、聞き手にやや負担がかかるが、理解できる発音である。	全体を通して正しい発音、イントネーション、リズムで話さることができている。不自然な沈黙がなく、聞き手に負担がかからない。	

それぞれの評価を、IRTを用いた統計処理にてスコアを算出します。それぞれの評価項目に配点がある訳ではありません。  
※例えば、正答率の高い評価観点で誤ってしまうと、スコアへのマイナス影響は大きくなります。