CEES Center for Entrance Examination Standardization

Benesse

GTEC スコアと CEFR レベル関連付け 調査報告

2017 年 9 月

GTEC と CEFR レベル関連付け調査メンバー

（敬称略）

根岸雅史（東京外国語大学大学院）
投野由紀夫（東京外国語大学大学院）
工藤洋路（玉川大学）
永田岳（海城中学高等学校）
高橋有加（東京外国語大学大学院生）
川本渚凡（東京外国語大学大学院生）

岡部康子（一般財団法人 進学基準研究機構）

込山智之（株式会社 ベネッセコーポレーション）
鹿島田優子（株式会社 ベネッセコーポレーション）
馬越優子（株式会社 ベネッセコーポレーション）
浜みか（株式会社 ベネッセコーポレーション）

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英語検定試験である。GTEC には、小学生・中学生向けの GTEC Junior、主に中学生・高校生向けの GTEC、そのコンピュータ版の GTEC CBT、大学生・社会人向けの GTEC Business がある。

本論文は、GTEC と GTEC CBT の CEFR レベルの関連付けを行った際の報告である。

１．調査の背景

英語検定試験のスコアと、言語運用能力を客観的に定義するために用いられている CEFR（Common European Framework of Reference for Languages）レベルとの関連付けが、スコア利用者（大学・受検者）にとって大切な指標となっている。

CEFR は 2001 年に英語版が出版され、その後も 40 の英語以外の言語でも出版された。ビザ申請のために英語の資格を証明する条件の一つとして、CEFR レベルとテストの得点を関連付ける研究を行っていることを義務付けている（中津原, 2013）。テストと CEFR を関連付ける動きは世界の各言語テストで行われており、投野（2013）では、CUP（ケンブリッジ大学出版会）の T-series という資料集と CEFR レベルとの相関や Cambridge Main Suit の諸テストと CEFR レベルの相関について言及されている。

日本でも、大学入試における英語の 4 技能外部試験の導入が検討されており、それらの外部試験は CEFR レベルとの関連付けが求められている（文部科学省, 2014）。

他の大規模英語テスト、例えば、TOEFL iBT では、2008 年に CEFR レベルとの関連付けを行っている（Tannenbaum & Wylie, 2008）。IELTS は、2011 年に IELTS テストと CEFR レベル関連付けについてレポートを出している（Cambridge ESOL, 2011）。

GTEC では 2015 年度に、フィールドテストデータ（本番試験前の予備テスト）の結果を用いて CEFR レベルとの関連付けを行った。本調査は、2016 年度に改めて GTEC および GTEC CBT の本番試験のデータに基づき、4 技能それぞれについて CEFR の閾値設定を行い、更にトータルスコアの CEFR 閾値設定について検証を行うことを目的とした。

２．GTEC および GTEC CBT の問題コンセプト、実施形式について

GTEC は、英語のコミュニケーション力を測る中高生向けの英語 4 技能（リスニング、リーディング、ライティング、スピーキング）テストである。実施形式は、リスニング、リーディング、ライティングについては紙（および CD）、スピーキングについてはタブレット PC を用いての実施となる。

リスニングとリーディングは、多肢選択式で、受検者は、配布されたマークシートに解答を記入する。ライティングは、解答用紙に手書きで解答を記入する。スピーキングは、タブレット PC 内にインストールしたアプリを使って問題を解答していく仕組みをとっている。受検結果はスコア型で提供される。教師向けと生徒向けにそれぞれフィードバックがなされ、日本国内の中学校・高等学校において指導改善や生徒への学習の動機付けのために活用されている。

GTEC CBT は、4 技能をコンピュータで受検する英語のテストである。学習指導要領から想定される「日常的な言語使用場面」におけるタスクと、大学での「アカデミックな言語使用場面」におけるタスクにより構成されている。リスニングとリーディングは、コンピュータ画面に現れる選択肢をクリックする形式で、ライティングは、キーボード入力による解答、スピーキングはヘッドセットのマイクを通して解答を吹き込む実施形式である。入学に必要な英語力の認定試験として主に米国の大学で採用されている。

３．GTEC におけるスコアについて

３．１．GTEC におけるスコア算出

GTEC および GTEC CBT では、技能ごとに項目反応理論（item response theory, IRT）に基づいてスコアが算出される。項目反応理論を用いるためには、各テスト問題の項目パラメータが、同一の能力尺度（共通尺度）上に推定されている必要がある。GTEC では、作成されたテスト問題はフィールドテストを通じてモニター受検者に対して実施され、統計的な性質に基づいて選抜される。選抜されたテスト問題は、本番試験で実際に出題されるテスト版に近い形に構成され、再度フィールドテストを通じてモニター受検者に対して実施される。2 回目のフィールドテストでは、モニター受検者は項目パラメータが推定されている過去のテスト版と、項目パラメータがまだ推定されていない新しい問題で構成されたテスト版の両方を受検することが求められる。このデータ収集法は共通受検者デザインと呼ばれ、項目パラメータを共通尺度上に推定するためのデータ収集法の 1 つである（加藤・山田・川端, 2014）。

共通尺度上に推定された項目パラメータに基づいて算出されるスコアは、出題されたテスト問題の難易度に拠らず、比較可能なものとして扱うことができる。そのため、異なる実施回に異なるテスト問題を受検した受検者間であっても、スコアを比較することができる。また、スコアの算出以外にも、テスト版の難易度および測定精度の管理において、項目反応理論が用いられている。

３．２．GTEC と GTEC CBT の関係性

リスニングとリーディングの 2 技能においては、GTEC と GTEC CBT の項目パラメータは同じ能力尺度上に推定されている。そのため、テストごとにスタンダードセッティングを行う必要はなく、2 つのテストに対して同時に行うことができる。一方、スピーキングとライティングにおいては、GTEC と GTEC CBT の項目パラメータは異なる能力尺度上に推定されているため、別々にスタンダードセッティングを行う必要がある。

４．スタンダードセッティング手法の種類について

スタンダードセッティングの実施については、CEFR マニュアル（A Manual: Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment）の第 6 章の Standard Setting Procedures に基づいた手順で行った。

スタンダードセッティングは、テストスコアをグループに分ける方法であり、テストにおいて、合格、不合格を決める際、または、クラスのプレイスメントテストによって、上級、中級、初級等に分ける場合などに、そのカットスコアを決めることを指す。

Council of Europe (2009) によると、スタンダードセッティングの方法としては、大きくテスト項目中心のもの、被験者中心のもの、IRT をベースに分類するものの３つに分かれる。

テスト項目中心のものとしては、Tucker-Angoff Method, The Yes-No Method, The Extended Tucker-Angoff Method 等がある。これらの手法は、パネルのメンバーが、それぞれのテスト項目について判断をするものであり、判断に実証的データは用いず、テスト項目のみを見て行われる。

被験者中心のものとしては、The Contrasting Groups Method, The Borderline Group Method, The Body of Work Method がある。これら手法は、被験者のことをよく知っている評価者によって判断がなされる。総合的な判断を行って、特定の受検者を閾値またはボーダーライン前後に振り分けていく。

最後に IRT をベースにするものとしては、IRT 分析を行った実証的なデータのサマリーを使って閾値設定を行う。IRT ベースの手法としては The Bookmark Method と A Cito Variation on the Bookmark Method が挙げられる（Council of Europe, 2009）。

TOEFL iBT の CEFR レベルのスタンダードセッティングでは、リスニングとリーディングは modified Angoff approach、ライティングとスピーキングでは modified examinee selection approach が用いられた（Tannenbaum & Wylie, 2008）。

５．方法

５．１．分析参加者

パネルは CEFR および英語の言語教育、教育測定に精通した研究者６名であった。また、この６名は全員、2015 年度の CEFR 関連付け調査にも参画し、CEFR には精通したメンバーで、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、英語特定の言語特徴についての知識も深いメンバーであった。2015 年の分析時には、全員が CEFTrain というツールを用いて、CEFR について理解を深めた（http://www.helsinki.fi/project/ceftrain/index.php.35.html）。

また、パネルの他に、GTEC の作問・制作に関わっているメンバーや第二言語習得の専門家が参画し、問題形式、採点基準等の説明を行ったり、質問に答えたり、議事録を取ったりする補助作業を行った。パネルの 6 名は、専門分野や教授経験年数を考慮し、2 名ずつ 3 グループに分かれ、分析を行った。

５．２．調査の方法

調査手法としては、2015 年度と同様、受容技能のリスニングとリーディングでは Bookmark Method、発表技能のライティングとスピーキングでは Contrasting-Group Method をベースとした手法（Council of Europe, 2009）を用いた。

リスニングとリーディングで Bookmark Method を用いた理由として、GTEC は IRT でスコアを算出しており、テスト項目ごとに困難度の数値が貼りついているため、それを客観的根拠として用い、それに加えてパネルの知見を加えた分析を入れることで、より適正に閾値設定ができると判断した。

ライティングとスピーキングで用いた Contrasting-Group Method は、受検者の解答パフォーマンスを能力値の順に並べて閾値を決定する方法である。マニュアルに記載されている手法に加え、リスニングとリーディング同様に、GTEC が IRT を用いてスコア算出している特徴を活かして、IRT のデータをもとにして分析を行った。IRT データに加えて、実際の受検者の算出データを詳細に分析することにより、より現実に即した閾値設定ができると判断した。

５．３．分析の流れ

調査は、事前課題と複数日にわたる集合型のワークショップ形式で行った。Council of Europe (2009) によると、スタンダードセッティングを実際に行う前に、すべての参加者が CEFR の知識をつけておくための familiarization の工程が必要だと述べられている。今回の調査においては、パネル全員が CEFR に精通したメンバーであること、加えて、2015 年での調査の際に CEFTrain でのトレーニングを終えていることから、今回の分析に先立っての familiarization の工程は省略した。

事前課題としては、リスニングとリーディングの Booklet を配布し、事前に素材やテスト項目にあらかじめ目を通したうえで、ワークショップに臨むこととした。

集合型のワークショップは、冒頭に specification と呼ばれる工程として、パネルに GTEC と GTEC CBT の各問題形式を説明し、質疑応答を行った。ライティングとスピーキングに関しては、さらに採点基準の説明も行った。

その後の流れは次の表のとおりである。

【表１．スタンダードセッティングの流れ（各技能共通)】

工程	内容	備考
①	CEFR 各レベルのディスクリプタを見ながら、Borderline Person の英語力のイメージを議論して一致させた。	全体会
②	2 名ずつ 3 グループの分科会に分かれ、資料を見ながら分析を行い、仮閾値を設定。	分科会
③	再度全体会に集合して 3 グループ合同での協議。全員で合意のもと閾値決定。	全体会

表１．の①～③の工程は、各技能共通の手順であった。最初は、各技能ともに、CEFR の最も下位レベルの A1 と A2 の閾値を行った。その後順次、A2/B1、B1/B2、B2/C1 と上位の判定に移るというプロセスで、①～③の工程を繰り返した。

①目線合わせの全体会においては、閾値にいる人（= Borderline Person）が 2 分の 1 の確率で解答できる英語力についてのイメージを全員で一致させるため、Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales. （Council of Europe, 2001b）の各技能のレベル別ディスクリプタを読み、その英語力のイメージが 6 名のパネルの間でほぼ認識が合うまで議論を行った。

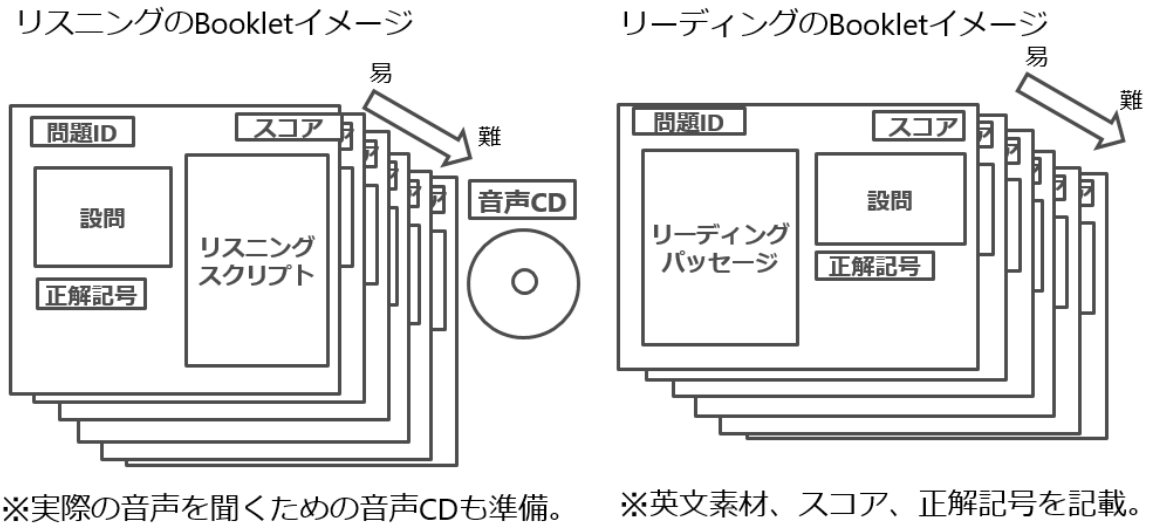
②分科会においては、2 名 1 組の 3 グループに分かれた。1 グループごとに GTEC の問題制作にかかわるメンバーが、質問に答える、機器の操作を行う、議事録を取る等の補助作業を行った。2 名のパネルは、各自が考える閾値の案件について見解を述べ合い、閾値について合意に至るまで意見交換を行った。迷う際には、その閾値前後の別の素材などもアイテムプールの中から参照し、慎重に閾値を決定した。

③全体会においては、各グループの仮閾値を発表し、仮閾値の決定に至った背景や分析手法をパネル全員と共有した。仮閾値がグループごとに異なる場合は、さらに議論を深めたり、その場で、それぞれの閾値付近の異なるアイテムを参照したりすることにより、すり合わせていった。スタンダードセッティングは基本的に判断に主観が入るが、グループごとに複数名で協議し、さらにその協議結果を全員でシェアし、確認することで、プロセスに客観的な要素を入れることを目的とした。

６．使用データ

使用した問題は、GTEC では、2016 年 7 月本番試験で出題した問題セットであり、GTEC CBT では、2016 年 7 月実施の問題セットであった。いずれも調査時期における最新の問題セットであった。

リスニングとリーディングは、GTEC CBT については、出題された項目の中から、IRT の θ 値から導かれたスコア一覧を参照し、該当のスコアのうち 0～350 の約 10 点刻みごとの英文素材を抽出し、分析用の Booklet を作成した。Booklet は、スコアの低い方から高い順にアイテムを並べ、英文素材のほかに、設問、正解記号、およびスコアを記した（図１）。困難度の高い問題については、GTEC CBT の問題を、低いものについては GTEC の問題を使用した。



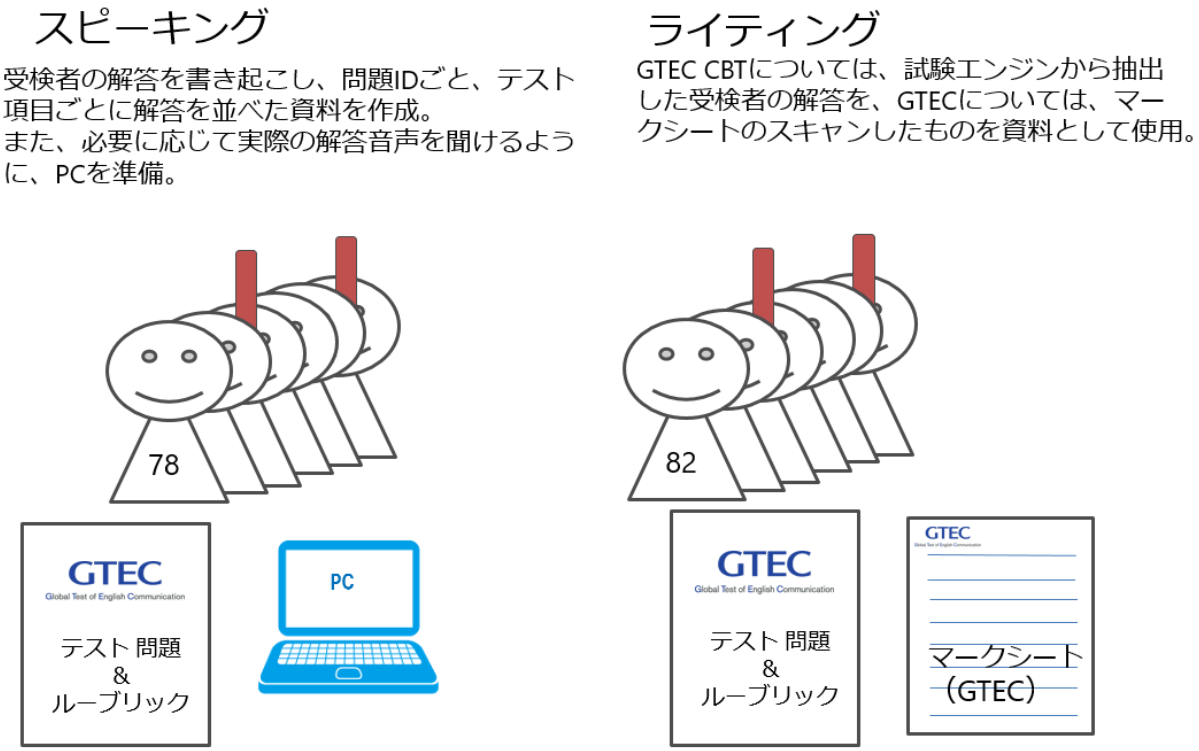
【図１：リスニングとリーディングの Booklet イメージ】

リスニングは該当のリスニング素材のみを収録した音声 CD も準備して、Booklet とともにパネルに配布した。

スピーキングとライティングに関しては、受検者の解答パフォーマンスを、スコアの低い方から高い順に並べた資料を作成した。スピーキングは録音された解答音声の書き起こしスクリプトを用意した。

スピーキングに関しては、その解答音声をスコアの低いものから高い順に並べたものについて、CEFR の各レベルの閾値に該当する受検者の解答結果を判定した。必要に応じて、解答音声も聞きながら判断を行った（図２）。

ライティングに関しては、GTEC CBT については、パソコン入力で行い、受検者の解答は試験エンジンに保管されているため、そこから抽出されたテキストデータを用いた。GTEC については、解答は手書きでマークシートに書かれるため、マークシートのスキャン画像を資料として用いた。



【図２：スピーキングとライティングの資料イメージ】

７．各技能別検討結果

この章では、技能ごと、および各 CEFR 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。

実際の検討の順番に合わせて、リスニング、リーディング、スピーキング、ライティングの順に、また閾値設定においては CEFR レベルの低い方から A1/A2、A2/B1、B1/B2、B2/C1 の順に記述する。

各スキル、各 CEFR レベル閾値の検討内容として、５．の表１にあるように、①CEFR 各レベルのディスクリプタを見ながら、Borderline Person の英語力のイメージを議論して一致させる工程、②2 名ずつ 3 グループの分科会での仮閾値設定、③全体会での議論、の流れに沿って行ったため、本章でもその流れに沿って説明する。

７．１．リスニング

閾値を決定する前段階として、全メンバーで、リスニングの各 CEFR レベルの self-assessment grid におけるディスクリプタの読み合わせを行い、それぞれのレベルの特徴を抽出し、この 2 つのレベルの差を生み出す要素、またはこの 2 つのレベルの差を示す要素となりにくいものを共有した。挙げられた要素や留意点は以下のものである。

＜リスニング　A1/A2＞

◆ディスクリプタの確認

A1
I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.

A2
I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

（Council of Europe, 2001a, pp. 26-27）

- A1 は、短い音声を聞いて単語やフレーズを聞き取る、つまり、語句レベルの処理が求められるレベルである。
- A2 では、聞いた音声の要点の理解が求められる。要点を聞き取るということは、聞くべき音声はある程度まとまったディスコースとして構成されている。まだ A レベル内ではあるので、短く簡潔な音声であることには変わりないが、要点とそうでない部分から成り立っているメッセージやアナウンスがリスニングの音声となる。その要点を把握するためには、語句レベルや文レベルのいくつかの要素を組み合わせて意味を理解するという処理が求められる。
- A1 のディスクリプタには「はっきりと話される」という記述があるが、本研究での対象はテスト項目のため、基本的にはどのレベルでもはっきりと話されていることが前提となるので、この点は A1 と A2 の差を明確に規定する要素とはなりにくい。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
145 と 150 の間	160 と 170 の間	150 と 160 の間

グループ A

A1 と A2 の閾値をスコア 145 と 150 の間と判断した理由は、今回対象としたテスト項目の中で、スコア 150 以上のレベルのテスト項目から、音声自体の難易度が A1 とは言えないものになっているというものであった。タスクが平易であれば、全体の難易度が下がることは認めながらも、スコア 150 からは音声自体の長さや複雑さを考えれば A2 以上のレベルであるとこのグループは判断した。

グループ B

スコア 160 とスコア 170 の間に閾値があると判断した。その根拠は、スコア 160 以上からは音声にまとまったディスコースが見られることから、音声全体の処理が求められ、詳細な点も聞き取る必要があるため、A2 レベルであると考えたということであった。また、スコア 160 以下のテスト項目では、語句レベルの処理が要求されており、部分的に聞いても正答が得られると判断した。

グループ C

閾値をスコア 150 とスコア 160 の間としているが、その根拠として、スコア 160 以下のテスト項目では、ディスコースがあまり見られない音声を聞いて、文以下のレベルの処理で正答が得られるテスト項目が多い点を挙げている。スコア 160 以上のテスト項目では、いくつかの文を処理していかないと正答が得られないため、A2 レベルの処理が要求されていると判断した。

全体会

これら各グループの判断とその根拠を共有した後、まず、リスニングの音声は一見難しくても、タスクが平易なものであれば、そのテスト項目の困難度は低い値となっていることが想像されることを確認した。閾値を議論する際には、各テスト項目の困難度は、音声とタスクの掛け合わせで規定されることを考慮する必要性を確かめた。

その後の議論では、この掛け合わせについて、どのように両者のバランスを調整するのかを決定することが難しいという問題点が挙げられた。あるテスト項目では、音声は長くまとまったディスコースが見られるため、音声レベルは確実に A2 レベル以上と言えるが、一方、タスクが平易なため、算出された困難度が低くなっているものが見られた。この困難度の数値から判断すると A1 であるが、音声自体を聞くと A1 と言うには抵抗があると述べたパネルもいた。そこで、スコア 155 付近にある別のテスト項目をいくつか参照して、閾値の根拠をより明確にしようと試みた。その結果、正答を得るのに部分的な理解でよい場合や、正答を得るための情報が複数回登場する場合などは A1 レベルであるだろうと判断し、また、正答を導くためには全体の理解が前提となる場合は A2 レベルであるだろうという判断を行った。したがって、スコア 155 以上であり、かつ 170 まではいかないということで、160 が閾値であると結論付けた。

＜リスニング　A2/B1＞

◆ディスクリプタの確認

A2
I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

B1
I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.

（Council of Europe, 2001a, pp. 26-27）

- A2 では、話題は個人的なものであるが、B1 レベルに上がると社会的な話題となる。つまり、ニュースや一般社会における出来事などを聞いて、理解できるレベルが B1 レベルであると言える。また、内容とタスクの予測可能性については、A2 レベルであれば、例えば、何かのイベントの紹介のアナウンスであれば開催場所を聞き取ることが予測できる。一方、B1 レベルでは社会的なニュースを扱うことにもなるため、内容の予測をすることは困難となる。従って、B1 レベルは、その場で適切に詳細も聞き取る必要が生じるレベルである。
- 聞き取るポイントについては、A2 レベルでは要点（メインポイント）であるが、B1 レベルでは詳細や手順を聞き取ることができるレベルである。さらに、部分的な理解だけではなく、部分同士のつながり、つまり話の流れの理解が B1 レベルでは可能になる。
- A2 と B1 の語彙レベルの境は、4000～5000 語レベルであるが、文字で見るリーディングとは異なり、リスニングの場合は、若干、この語彙レベルよりは低めを想定するのがよい。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
230 と 240 の間	210 と 220 の間	220 と 230 の間

グループ A
　スコア 230 とスコア 240 の間に閾値を設定したのは、スコア 230 の項目まではタスク自体が平易であることが大きな理由の一つであった。一部、タスクの難易度は高いとは思われないが、テスト項目としての困難度が高い値となっているものがあり、判断に迷った。そこで、当初の Booklet には掲載されていない別のテスト項目を新たにいくつか見てそのレベルを判定した上で、スコア 240 以上が B1 レベルであると総合的に判断した。

グループ B
　設定した閾値は、スコア 210 と 220 の間であったが、スコア 220 の項目では、タスクが明示的な 1 つの情報を聞き取るものではなく、いくつかの情報を組み合わせて聞き取るものとなっていることから B1 レベルと判断した。それより下の難易度のスコア 200 の項目は、場所を聞き取る問題であるが、音声では、場所に関わる情報が何度も読み上げられるため、B1 レベルとは言えないと判断した。さらに、スコア 220 より上のレベルのスコア 230 の項目は、時間の聞き取りではあるが、少し複雑で簡単な項目ではない。内容は、スケジュールに関する友人からのインフォーマルなメッセージではあるが、慣用的な表現も見られ、単に 1 つの時間を聞き取るだけではなく、予定の変更が理由とともに話される複雑な構造をしている。これを聞き取るために

は B1 以上の力が必要であると判断した。従って、スコア 220 以上の項目を B1 と判断した。

グループ C

A2 と B1 の閾値をスコア 220 とスコア 230 の間に設定した。スコア 220 より難易度が下のスコア 200 やスコア 210 のテスト項目については、聞き取る内容とタスクで問われる点が予測可能な範囲のものであることから、スコア 210 までの項目は A2 レベルであると、まずは判断した。スコア 220 の項目については、全体の流れを把握する必要のあるタスクが設定されているが、具体的な場所を聞き取るという点においては、難易度はそれほど高くないと判断し、スコア 220 のテスト項目は A2 レベルであると判断した。一方、スコア 230 のテスト項目については、詳細な指示の聞き取りが求められ、読み上げられるスピードが比較的速く感じられたことなどから、B1 レベルであると判断した。

全体会

これら各グループの判断とその根拠を共有した後、グループ A が参照した別の問題を全体で確認し、難易度の判定を行った。スコア 222 のテスト項目については、細かい情報処理が求められるため B1 レベルであるということで全員の意見が一致した。スコア 217 のテスト項目については、明示的な個別の情報を聞き取ることができればタスクが完了するため A2 レベルであると判断した。したがって、その間に A レベルと B レベルの分かれ目があると判断し、閾値 220 と結論付けた。

＜リスニング　B1/B2＞

◆ディスクリプタの確認

B1
I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.

B2
I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.

（Council of Europe, 2001a, pp. 26-27）

- B2 レベルからは、音声で流れる英文の内容は、より複雑になっていく。ただし、困難度はテスト項目を基準に算出されている、英文の内容がそれほど複雑ではなくても、テスト項目が難度の高いものであるため、項目難易度が高く出ているものもあることを想定しておく。
- 語彙については、B レベルではあるので社会的な話題に関する語彙が多くなると考えられる。B2 レベルが B1 レベルと異なる点は、B2 レベルではアカデミックなレベルの語彙の習熟が必要となる点である。
- B2 レベルになると、真正性が非常に高くなる。真正性を保証するために、現実のリスニングで起こり得る状況が反映されるため、英文が読まれる時間がかなり長くなることも想定する。さらに、これに伴い、speaker's mood や attitude という観点も考慮に入れるべき点となる。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
290 と 300 の間	290 と 300 の間	290 と 300 の間

このレベルの閾値は、全 3 グループがスコア 290 とスコア 300 の間で一致した判断を出した。

グループ A
スコア 270 くらいの項目から、リスニング時間も大幅に増えていく傾向が見て取れると判断した。そして、スコア 291 の項目から、内容においてアカデミックな要素が強くなっていると考えたため、スコア 290 とスコア 300 の間に閾値を引いた。

グループ B
スコア 291 の項目は音声内容は確実に B2 レベルであると言えるが、タスクがそれほど難しいものでないと考えた。ただし、これより上のレベルの問題からは確実に B2 レベルであると判断できるため、スコア 291 までを B1 と設定した。具体的には、スコア 300 の項目は、外国での大学の講義を英語で聞いている設定であるが、数分以上に渡る長さであり、配布資料や黒板での情報提示などがない中で英語を聞き続け、それを理解するのには高度な能力が必要とされる。したがって、このテスト項目は B2 以上のレベルでなければ対応できないと判断した。

グループ C
スコア 291 の項目 S について、リスニングで流れる内容が学術的であり、B2 レベルであると判断した。また、スコア 271 はタスクレベルが低いため B2 レベルとは判断できなかった。また、追加でスコア 281 のテスト項目を見たところ、内容は身近ではない話題であったが、タスクは非常に明示的で分かりやすいため、スコア 281 のテスト項目は B1 と判断した。

全体会
各グループの上記の判断理由を共有し、3 グループ間で判断が一致したスコア 290 とスコア 300 の間に閾値を引くこととした。

＜リスニング　B2/C1＞

◆ディスクリプタの確認

B2
I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.

C1
I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.

（Council of Europe, 2001a, pp. 26-27）

- C1 の音声内容は、本来は、実際の状況での会話や講義などになるべきであるが、その場合、言い直しや言いよどみ、または雑音などが入ったりすることが自然である。また、方言や訛りなども現実の言語使用場面ではリスニングの成否の大きな要因になる。ただし、テスト環境ではこのような設定はしにくいことから、これらの点でのレベル判定は実際は行うことはできないと想定される。

- B2 と C1 の違いの 1 つには、トピックの抽象度や専門性が挙げられる。トピックの抽象度や専門性は、トピックの親密度とも関わる。C のレベルは、トピックが身近なものではなく、高度な社会性のある話題であっても対応できるレベルである。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
347 と 351 の間	330 と 335 の間	335 と 351 の間

グループ A

Booklet に掲載されている問題以外のテスト項目も参照し、それらを確認した上で、スコア 347 とスコア 351 の間に閾値を引いた。スコア 335 の項目は語彙さえしっかりと分かっていれば、B2 レベルでも聞き取ることは可能であると判断した。スコア 351 の項目は、内容的な難しさよりも、設問の難しさが見て取れるため、スコア 351 は C1 と判断した。

グループ B

スコア 335 の項目について、内容がかなり込み入っており、非常に難しいと考えることができるため、閾値はスコア 330 とスコア 335 の間に引いた。トピックは高校生にとって親密度がかなり低い。また、音声の全体の長さも事前に分かるわけではなく、前半分が内容的に複雑であることから、スコア 335 は C1 レベルと判断できるとしている。

グループ C

閾値をスコア 335 とスコア 351 の間であると判断した。スコア 335 のテスト項目のリスニングパッセージは内容の流れが分かるように、繰り返し説明されている部分があったりするなど、スコア 351 に比べて分かりやすい。タスク自体もそれほど複雑ではないことから、スコア 335 までは B2 レベルの学習者が達成できるのではないかと判断した。上記の議論により、スコア 340 を閾値と設定した。

全体会

これら各グループの報告を受けて、スコア 335 の項目を B2 レベルと判断するのか、C1 レベルになってはじめてできるようになる項目かを議論した。スコア 335 は、語彙の観点ではそれほど難易度が高いというわけではないが、その一方で、1 つ目の問題の解答を導くまでに、音声の多くの部分を聞き続ける必要があるため難しいという意見に分かれた。再度スコア 335 の項目を確認したところ、文章構成はそれほど複雑なものではなく、同じ情報を別の言い方で繰り返して説明している場面もあるなど、C1 レベルほどの難易度はないのではないかという意見が大半を占めた。さらに、タスクもそれほど難しいものではないことなども、スコア 335 が B2 レベルであると判断した理由である。

＜リスニングまとめ＞

最終的にリスニングの閾値は、下記に決定した。

Listening	閾値
B2/ C1	340
B1/ B2	290
A2/ B1	220
A1/ A2	160

７．２．リーディング

＜リーディング　A1/A2＞

◆ディスクリプタの確認

リーディングにおける A1 と A2 の閾値を決定するために、全メンバーで A1 と A2 の self-assessment grid や Overall Reading Comprehension、Reading For Information & Argument などにおけるディスクリプタの読み合わせを行い、それぞれのレベル特徴を抽出し、この 2 つのレベルの差を生み出す要素、またはこの 2 つのレベルの差を示す要素となりにくいものを共有した。挙げられた要素や留意点は以下のものである。

A1
I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.

A2
I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.

（Council of Europe, 2001a, pp. 26-27）

- A1 での非常に短い素材文における限られた語句や文の理解から、A2 では短く単純ながらも、より多様な種類の素材文中においての予測可能な情報の理解へとタスクの性質が変わっている。
- A1 で想定されている素材文の種類はかなり限られているものの、検討メンバーの CEFR 準拠教材などに関する研究経験などから、ディスクリプタの文言から想定される以上に A1 には多様な言語素材が含まれる。日本における英語教科書レベルに置きかえて考えると、中学２年生程度のテキストが他者の援助なしに読めるレベル、という表現が適切だと考えられる。
- CEFR で絵や写真など視覚的情報が大きな助けとなるとされているのは A1 であり、日常的に目にするポスター等、テキストの長さが短く、理解の際にそうした視覚情報に大きく依存して解答することができるものはこのレベルだと考えられる。
- GTEC で設定されているテキストタイプは、CEFR の A1 で想定されているようなものとは必ずしも一致しないことがあり、テキストタイプのみから CEFR レベルを判断することはできない。しかしながら上で述べたような理由から、E メールなど、本来 CEFR の A1 で想定されていない種類のテキストにおいても、そのテキストの長さがとても短く、かつ語彙も単純であれば、タスクによっては A1 となりうる。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
153 と 161 の間	142 と 153 の間	171 と 182 の間

グループ A
資料中での問題の最も項目困難度の低いものが、すでに素材文のタイプとしては A2 であるものの、テストタスクについてはいくつかの語句を理解すれば正答できるような A1 の特徴が見られたとした。しかしながら、スコア 161 以降の項目については、素材文中の部分的な語句の理解や、特定の情報の位置が明確に示された図表内の単純な情報を読み取るだけでは不十分であり、素材文の内容を理解し、正答でないものを排除する力が必要であると考え、ここに両レベルの閾値があると判断した。また、これ以降の困難度の項目については、素材文の長さも伸びてきてお

り、素材文のどこに正答の根拠となる情報があるかを判断しなければならない。その文章の一部の内容を正しく読み取る必要があり、ここに A1 よりも A2 の特徴が現れていると考えた。

グループ B

グループ A と同様に、情報が素材文のどこにあるかを判断すること自体がテキストの難易度を左右していると考え、スコア 150 以降に閾値があるのではないかと考えた。しかしながら、スコア 153 の項目を検討したところ、正答の根拠となる部分さえ分かれば正答できる単純なものであるものの、素材文中の図表の構造が少し分かりにくいものになっていたため、このレベルの典型的な項目ではないと考え検討から除外した。その後、Booklet 外のスコア 150 前後の項目を参照したところ、スコア 147 の項目において A1 の特徴が見られたため、スコア 150 以降に閾値があると判断した。

グループ C

スコア 153 と 161 の項目の間に、語句レベルでの情報処理か、よりまとまりのある単位での情報処理かという点でタスクの性質に差異を感じたものの、グループ B と同じ理由からスコア 153 の項目の処理に悩んだため、それ以降の項目を確認した。その結果、スコア 170 前後まで大きな差を見出すことができず、また受検者がこれらをスキャニングの問題だと考え、関連する語句を探すことに集中すれば、正答にたどり着くことも困難ではないと判断したことから、スコア 171 とスコア 182 の間に閾値があると判断した。

全体会

これら各グループの判断とその根拠を共有した後、まず、リーディングにおいて必要な情報の位置を特定するということ自体が、A1 の学習者にとって難易度が高い可能性があることが全体で確認された。そのような A1 の特徴を検証するため、Booklet 外のスコア 150 前後の項目を確認したところ、スキャニングの対象となる語や表現自体を受検者自らが判断し、ある程度の文量を持つ文章から探すというようなタスクは、素材文の本旨をつかむといった、一般的により難易度が高いと思われるタスクと同程度の困難度を示していることが判明した。日本人学習者にとっては、こうしたスキャニング課題が教科書等で扱われないため、タスクとしての本来の難易度が項目困難度に反映されているわけではなく、素材文自体の難易度が項目の困難度に影響を与えている可能性が指摘された。こうした点を踏まえ、文章全体から関連のある箇所を探すタスクであるか、素材文全体の本旨をつかむ読解を要するタスクか、というような読みの性質そのものよりも、素材文の長さなどにより注目して判断する必要があることが全体で確認された。これらの検討内容から、スコア 140 前後の項目について、それ以降の困難度の項目と比較し、素材文の長さも短く、かつ個別の情報の位置も分りやすい傾向にあるという判断を下し、全体としてはスコア 150 が閾値であるという結論に達した。

＜リーディング　A2/B1＞

◆ディスクリプタの確認

A2

I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.

B1

I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.

（Council of Europe, 2001a, pp. 26-27）

- A2 と比較し、B1 は長めの素材文が想定されており、そのため要点の理解などもタスクとして現れてくる。
- スキャニングのような読解についても、B1 では A2 よりもさらに長い素材文から関連の深い箇所を見つけることができると考えられる。
- 素材文内の 1 箇所の理解にとどまらず、複数箇所の情報を整理して理解する必要があるなど、総合的に情報のつながりを理解する必要が出てくるのも B1 である。
- ただし、B1 の下限という限定的な意味合いにおいては、少し長めになってくる素材文において要点の認識ができるかが重要な視点となる。
- B1 では、素材文タイプも手紙などだけでなく、学習者自身の日常生活、興味関心などと関連のあるものが含まれるようになり、例えばパンフレットや単純な新聞記事、短い公的文書などもディスクリプタ内で触れられている。しかしながら素材文タイプの広がりに限界のあるテストタスク内では、こうした素材文タイプの違いを重視してレベル間の差異を見出すことは困難である。
- 受検者の生活領域や興味関心などについてもテストタスクにおいては統制できないため、こうした点を重要な要素として参考にすることもできない。非常に一般的なレベルで、より専門的、より日常的という区別を行うしかない。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
210 と 219 の間	229 と 240 の間	219 と 229 の間

グループ A
　読解時に複数箇所の情報を総合して理解する必要性に注目して判断をした。スコア 210 以降の項目について、正答にたどり着くためには、その直接的なキーワードとなるものがないなかで、いくつかの箇所を理解し、結び付ける必要があり、そのためスコア 210 を閾値と考えた。

グループ B
　グループ A と同様の視点で分析した。しかしながら、スコア 210 以降の項目にそうした複数箇所の理解が前提となる特徴は見てとれるものの、タスクの複雑性がより顕著になってくるのは、スコア 229 以降の項目であると分析した。そのため、スコア 230 前後に閾値があると判断した。

グループ C
　グループ B と同様の分析から、Booklet 外の問題も参照しつつ検討を行った。その結果、スコア 219 までの項目は、正答を導く際の情報の複雑性があまり高くないのに対し、スコア 229 の項目では日本の高校生にとってはトピックの馴染みもあまりなく、語彙の難易度も高くなる傾向が見受けられた。そのため 220 前後に A2 と B1 の閾値があると判断した。

全体会
　これら各グループの判断とその根拠を共有した後、まず、困難度が上がれば上がるほど、タスクに取り組む際の素材文中の関連する情報の複合性が重要となる傾向が顕著になっていくことが全体で確認された。その上でスコア 220 の項目については、それ以下のものよりもはっきりと B1 の特徴を示していることについて同意が取れたため、全体としてはスコア 220 が閾値であるという結論に達した。

＜リーディング　B1/B2＞

◆ディスクリプタの確認

B1
I can understand texts that consist mainly of high frequency everyday or job-related language.

I can understand the description of events, feelings and wishes in personal letters.

B2
I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.

（Council of Europe, 2001a, pp. 26-27）

- B2 は、テキストの内容に専門的な内容が見られるようになる。想定される素材文も長く複雑なものとなり、一般的な社会ニュースというよりも、よりメッセージ性のある記事や報告書、大学教養レベルの文章や文学的な内容を含んだものが含まれるようになる。
- B2 は B1 と比べ、より多様な素材文に対し、より多様な読み方が可能になるという趣旨の記述も CEFR 内にはあるが、ひとつひとつのテスト項目から閾値を判断する今回については、こうした点を参考にすることは難しい。
- テストにおける素材文がより専門的な内容を含むものであったとしても、テストタスク自体が単純で、素材文の適切な理解をあまり前提としなくても解けるようなものであれば、B1 と判断するのが妥当である。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
280 と 290 の間	260 と 270 の間	270 と 280 の間

グループ A
スコア 270 以降から B1 の特徴を持った素材文と B2 の特徴を持った素材文とが混在しているとし、Booklet 外の項目も参照した。その上で、スコア 280 前後の項目についてトピックの専門性などから B2 とするには不十分な問題も一部散見されたため、スコア 290 が閾値であると判断した。

グループ B
問題の複雑性などから、スコア 270 以降で B1 と B2 それぞれの特徴を持つタスクが混在している印象を受けた。その際、参照した問題において、タスクの複雑性の点から差異を感じたスコア 270 と 280 の項目が同一の素材文についての組問であったことから、いきなりこの間に閾値を設けることをせず、この前後の他の項目について、Booklet 外の項目を参照することとした。この結果、スコア 270 の項目においても、B2 のディスクリプタ上に見られる書き手の態度や意見を問うタスクが現れているため、スコア 270 以降を B2 とするという判断を下した。

グループ C
上２つのグループと同様に、困難度 270 以降のテスト項目において B1 の特徴と B2 の特徴が混在していると感じたが、共通資料外のスコア 270 前後の項目については素材文およびタスクの複雑性の観点から、B1 には困難すぎると考えられる問題がいくつも見られたため、スコア 270 と 280 の間に B1 と B2 の閾値があると考えた。

全体会
これら各グループの判断とその根拠を共有した後、素材文の性質以外にも、タスクの性質が B1 の上限と B2 の下限を区別する際の重要な視点となる可能性があることを全体で確認した。こうした議論から、受検者が同一の素材文であっても異なったタスクに応じて異なった読み方をするとし、そうしたタスクの複雑性に注目して全メンバーで項目の再検討を行ったところ、特に、共通資料内のスコア 270 とスコア 280 の項目は同一の素材文に関連付いた問題であるものの、その間にはタスクとしてやや質的な差が見られることが合意された。具体的にはスコア 270 の項目において、正答のキーワードを文章冒頭より読み取ることができれば課題の解決を図れる一方、ス

コア 280 の項目においては、段落のつながりから正答の根拠となる部分を見つけ、その箇所の内容を大まかにつかむ必要があった。また、スコア 280 以降の項目についてはタスクの複雑性がより増していく傾向が確認された。そのため、検討を行った 3 グループそれぞれが、問われているタスクの特徴により差異を感じたスコア 270 とスコア 280 の間に閾値があるという結論に達した。

＜リーディング　B2/C1＞

◆ディスクリプタの確認

B2
I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.

C1
I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.

（Council of Europe, 2001a, pp. 26-27）

- C1 は、B2 よりもさらに長い素材文を読むことが要求されるはずだが、テストの設計上、B2 と C1 それぞれを想定した設問で長さが変わらないため、その点では両者を区別できない。
- C1 では、自分の専門外の内容でも理解できるということから、トピックについてもかなり専門的で難解な内容についての理解が必要となる。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
313 と 317 の間	333 と 341 の間	320 と 333 の間

グループ A
　共通資料内にスコア 300 以降の項目が不足していたため、それ以外の問題を参照した。スコア 340 などで比較的易しい項目があるなど､B2 と C1 の項目が混在している様子は見受けられたが、スコア 317 の複数のテスト項目において、科学的な内容や書評などの、あまり受検者にとって馴染みがないと考えられる話題が含まれていた。そのため、この前後であるスコア 313 とスコア 317 の間に閾値があると判断した。

グループ B
　スコア 340 のテスト項目から明らかに C1 の特徴が見られると考えた。それ以前の項目については、グループ A と同様に難易度に混在が見られると考えた。

グループ C
　Booklet 外の問題を参照したところ、スコア 317 の項目の 1 つである書評の問題についても、タスク自体はやや単純な情報の比較となっており、難易度の高い項目ではないと判断した。その一方、スコア 333 の問題については、話題が多くの受検者にとって一見馴染みのありそうな教育に関した内容であるものの、実際にはその理解の鍵となる語彙や概念が難解であると考えた。そのため、それ以下のスコアのテスト項目を参照したところ、スコア 333 の項目と比較し、より一般的なトピックが扱われている印象を受けたため、この間のスコア 320 とスコア 333 の間に B2 と C1 の閾値があると判断した。

全体会
　これら各グループの判断とその根拠を共有した後、検討メンバーが専門としている分野が「教

育」に関連するものであるため、スコア 333 の項目のようにそういった分野が素材文の話題となっている場合、そのタスクの難易度を過小評価しがちであることが確認された。加えて、スコア 340 の項目については、前後の複数の項目も参照したうえで、確実に C1 であると全員の判断が一致したため、最終的にスコア 330 に閾値があるという結論に達した。

＜リーディングまとめ＞

最終的にリーディングの閾値は、下記のように決定した。

Reading	閾値
B2/ C1	330
B1/ B2	280
A2/ B1	220
A1/ A2	150

７．３．スピーキング

＜スピーキング　A1/A2＞

◆ディスクリプタの確認

Spoken Interaction
A1
I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.

A2
I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.

Spoken Production
A1
I can use simple phrases and sentences to describe where I live and people I know.

A2
I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.

（Council of Europe, 2001a, pp. 26-27）

- A レベルの学習者は、A1、A2 ともに、自身や家族などについて簡単に話すことができる。
- A2 レベルでは情報を少し加えて、身の回りのことについて簡単な説明ができるようになる。

A レベルの学習者が英語を使ってできることは、ごく身近な日常的な内容に限られている。そのため、両レベルの学習者が共通してできることの特徴として、上記の 1 つ目の内容が挙げられる。とはいえ、A2 レベルでは A1 レベルでできることよりは少し発展したことができるようになることも、全メンバーで確認した。A2 レベルでは、単に一つの事柄について短く言うだけではなく、そこに少し情報を加えて話すことができるようになる。この点が A1 と A2 レベルとを区別する判断基準となり得る。しかし、A2 レベルになると、情報を少し加えることができるとはいえ、難しい表現を使えるということではなく、まだ定型表現を用いて言えることに限られているのが A2 レベルの特徴であることを全メンバーで共有した後、各グループで話し合いを行った。

◆各グループ別判定結果の検討

各グループで話し合った結果、GTEC と GTEC CBT における A1 と A2 レベルの閾値は、それぞれ以下の表のようになった。

GTEC

グループ A	グループ B	グループ C
171 と 192 の間	166 と 171 の間	171 と 192 の間

GTEC CBT

グループ A	グループ B	グループ C
78 と 88 の間	100	97 と 110 の間

【GTEC A1/A2 の協議】

グループ A

A1 と A2 の閾値をスコア 171 とスコア 192 の間、特にスコア 188 あたりに閾値があるのではないかと判定した。スコア 154 とスコア 171 の受検者はなんとか話そうとする努力は見られるものの、意味が通るような文を作るだけの能力がまだ備わっていない印象があったと報告している。

スコア 192 の受検者となると、与えられたトピックに関する意見を話すことを求められるパート D のテスト項目においても、伝えようとすることが相手に理解されるような発話ができるようになってきていることが確認できた。また、それより低いレベルと比較すると発話数も増えてきていることがわかった。

スコア 213 の受検者となると、意味をなす文が 3、4 文ほど、かろうじて作り出すことができてきており、一つの事柄において複数の文を用いて話すことができる特徴が確認でき、A2 レベルの域に入ってきていると判断した。

また、スコア 188 あたりから、それより低い点数の受検者とは発音の質が異なってくることもグループ A のメンバーは指摘している。まだ同じ単語の繰り返しや言いよどみは目立つものの、何を言っているのかわからない単語が少なくなっている点が A2 レベルの下位群らしさ、A2 レベルに到達し始めていることを感じるものであると判断した。

これらの理由からグループ A は、GTEC における A2 レベルの閾値はスコア 171 とスコア 192 の間、特にスコア 188 あたりにあると判断した。

グループ B

A2 の閾値をグループ A よりは少し低いスコア 166 と 171 の間にあると判断した。スコア 139 の受検者は、なんとか話そうとしてはいるものの、1 文を作り出すためにでさえ、同じ語やフレーズを何度も繰り返しており、A1 レベルであるという印象を受ける。パート D のタスクにおいて、スコア 166 や 171 を取った受検者は、言いよどむ箇所はまだ複数箇所に見られるものの、意味が通じる文を作る能力が備わってきている点で、かろうじて A2 レベルに到達しそうなレベルにある印象を受けた。

スコア 192 の解答音声では、簡単な表現に留まっているものの、何を言おうとしているのかが十分に理解できるというような特徴が観察された。そのため、スコア 166 からスコア 171 の間の受検者は A2 レベルの下位群あたりの能力があり、A2 レベルの閾値はこのあたりに存在すると判断した。

グループ C

グループ A と同様のスコア 171 とスコア 192 の間に閾値があると判断した。スコア 192 の解答音声を聞いてみると、文法的な誤りが見られるものの、かなり流暢に話すことができている印象を受けた。A1 レベルのような、なんとか知っている単語を羅列しただけではなく、意味の通る文を作ることができるようになっている特徴が観察され、A2 レベルの下位群に分類されるだけの能力は備わっていると判断できることが、グループ C の閾値判断の理由であった。

全体会

3 グループ共に、スコア 171 かそれより少し上に GTEC における A2 レベルの閾値があると判断しており、A2 レベルに到達し始める下位群を意識すると、スコア 171 あたりに閾値があることは全メンバーが同意した。しかし、今回のスタンダードセッティングにおいて用いられた、ある得点を取った解答は限られており、今回は参照しなかった他の解答の出来具合によっては、A2 レベルの閾値はスコア 188 あたりに引き上げてもいいのではないかという可能性もあると全体の議論の中で意見が挙がった。また、準備していた解答が Advanced タイプのものであったため、タスクの難易度が、該当の CEFR レベルの受検者にとっては高いものであることから、詳細の分析が困難と判断された。後日改めて、より易しい難易度の Basic、Core タイプの問題より、スコア 169 から 188 の解答を準備し、その解答を一部メンバーで集まり、確認して閾値の再検討を行なった。その検討内容、検討結果を改めて参会者全員に共有し、内容を確認、承認し、最終的に A2 レベルの閾値はスコア 188 と決定した。

【GTEC CBT A1/A2 の協議】

グループ A

スコア 68 より高い得点を取った受検者は、比較的発話数が増えてきている印象である。しかし、スコア 68 の受検者は、問題設定に記載されている英語の情報を高い割合で使用しており（そのまま読み上げており）、この受検者が自ら考えて発話した部分が限られていることが解答音声からわかる。

スコア 90 の解答音声からは確実に A2 レベルであると断言できるが、A2 レベルの下位群を意識すればスコア 80 あたりに閾値を設定してもいいのではないかとも感じた。A2 レベルが A1 レベルにおいて異なる特徴は、少し情報を加えて身の回りのことを言うことができる点であったが、パート 2 のタスクにおいて、提示されたビジュアル情報から受検者が様々な情報を抽出し、その情報について話すことが求められるべきではあり、さらに加える情報を見つけ出すのが難しく、発話数が限られてしまうのも致し方ないような印象もあったため、スコア 78 と 88 の間に A2 レベルの閾値があると判断した。

グループ B

分科会において、スコア 90 とスコア 101 の受検者の解答に差があると判断した。スコア 101 の解答は A2 レベルと言えるのに対して、スコア 90 の解答はパート 2 のビジュアル情報を利用して解答する問題はきちんと答えているが、パート 1 やパート 3 の問題が不完全で、A2 とは言いがたいと判断した。したがって、スコア 90 と 101 の間に A2 レベルの閾値があるとしてスコア 100 とした。

グループ C

スコア 78 の解答音声を聞いた際に、パート 2 は言いよどみや文法的な誤りがあるとは言え、言わんとしている内容は満足に理解できるような発話であるという特徴が見られ、A2 レベルに達しているような印象は受けた。

その一方で、同じスコア 78 のパート 3 の解答音声を聞いてみると、同じ受検者であるにも関わらず、同じ単語を何度も言い返す箇所や話の途中で言いよどみ、文が途切れてしまう箇所が複数観察された。A2 レベルで言えることは身の回りのことに限られてはいるものの、比較的身近なトピックについて、このような言いよどみが複数見られる点については、A2 レベルに到達するまでには少し距離を感じることをパネルは指摘している。

さらに高い得点の解答音声を視聴した結果、スコア 110 では確実に A2 レベルに達していると言えるものであったため、スコア 90 からスコア 102 の間だと判断した。

全体会

3 グループ共に、A2 レベルの下位群に入る点数がどのあたりかを意識した閾値設定であったため、各グループが判断した閾値には多少のずれが確認された。しかし、スコア 110 の受検者は確実に A2 レベルに達しているという判断には全メンバーが一致していた。また、全体での協議の際に、スタンダードセッティングにおいて用意された解答とは異なる、スコア 90 台の受検者の解答を全体会で確認したところ、解答の内容が A2 レベルに届くか届かないか程度の出来であった。このため、スコア 100 あたりに A2 レベルの下位群が位置しており、閾値はスコア 100 に存在するという結論に達した。

＜スピーキング　A2/B1＞

◆ディスクリプタの確認

Spoken Interaction
A2
I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even

though I can't usually understand enough to keep the conversation going myself.

B1
I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).

Spoken Production
A2
I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.

B1
I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.

（Council of Europe, 2001a, pp. 26-27）

- A1 レベルと A2 レベルの当該ディスクリプタの確認の際に挙げられた内容ではあるが、A2 レベルは、比較的自分の身の回りのことについて簡単に言える程度の能力がある。
- その一方で、B1 レベルにおいては、友人間などの非公式な議論においてできること (Illustrative scales の informal discussion (with friends))を参照すると、"can compare and contrast alternatives, *discussing what to do, where to go, who or which to choose etc.*"とあり、「どこへ行く」や「何をする」などの意見のやりとりができるようになることも B1 レベルの特徴であると言える。
- B1 レベルになると、A2 レベルよりは多くのことができるようになるが、まだインフォーマルな会話ができる程度に留まっており、フォーマルな場面での発話は難しいレベルであることに留意しなければならない。

上記の 2 つのレベルを区別する特徴 3 つに加えて、これらのレベルの特徴を受検者が満たしているかどうかを判断する基準となるタスクについても全員で共有した。

2 点目の特徴として挙げた「どこへ行く」や「何をする」という意見を友人に伝えることができるかどうかを検証するタスクとして、GTEC CBT では「相手に自分の意見や要望を伝えたり、質問をしたりすること」が求められているパート 2 を中心的に見ることが B1 レベルに達しているかどうかの判断に役立つこと、一方、GTEC に関しては、自身の意見を述べることを求められるパート D の問題が B1 の閾値の判断に有益であることの 2 点が注目するタスクであることを全員で確認した。

これらの判断基準と着目すべきタスクを参考にし、それぞれのグループでの分科会を行なった。

◆各グループ別判定結果の検討

グループごとの分科会においての閾値の判定は以下の表のようになった。

GTEC

グループ A	グループ B	グループ C
258 と 277 の間	286 と 305 の間	264

GTEC CBT

グループ A	グループ B	グループ C
190	220	231 と 242 の間

【GTEC A2/B1 の協議】

グループ A

分科会において、閾値の判断材料として高得点の解答音声も参考に聞いたが、流暢さがあると言えるものは皆無に等しく、高得点を得た受検者も B1 レベルに達しているとは言い難いと感じた。

しかし、短い時間のなかで解答することが求められていること、準備時間が 1 分と比較的短いことを考慮すると、流暢さが十分ではないがスコア 264 の受検者でも B1 レベルと言えるだけの能力を備えている可能性があると考えたため、スコア 258 とスコア 277 の間に閾値があると判断した。

これまで CEFR レベルの感覚を養うために行ってきた CEFTrain や、ヨーロッパなどで行われてきたワークショップを通して身につけた B1 レベルの流暢さの感覚は、ヨーロッパの英語学習者のものであり、日本人英語学習者と流暢さという点ではかなりの差があるとされた。日本人英語学習者の話す能力がヨーロッパの学習者より低いことを考慮して、B1 レベルの下位レベルを低く見積もると、B1 レベルをスコア 264 まで引き下げられる。しかし、それよりも低い点数を B1 レベルの閾値にすることには抵抗があることがスコア 258 とスコア 277 の間、特にスコア 264 あたりに閾値があるとグループ A が判断した理由であった。

グループ B

グループ A と同様、スコア 286 を獲得した受検者であっても B1 レベルに到達しているとは言えないと判断している。CEFR のディスクリプタを確認すると、B1 レベルでは“I can deal with most situations likely to arise whilst travelling in an area where the language is spoken.”（Council of Europe, 2001b, p.6）とあり、いろいろなことがある程度高いレベルでできなければいけないことが B1 レベルの特徴として挙げられる。しかし、スコア 286 の受検者のパート C の発話の中には、同じ単語の繰り返しや言葉に詰まる場面が多く、スコア 286 の描写の程度では「B1 らしさ」を感じさせるレベルまでには届いておらず、スコア 286 とスコア 305 の間に閾値があると判断した。

グループ C

グループ B がスコア 286 の受検者はまだ B1 レベルに達していないと判断した一方で、グループ C は、スコア 286 の受検者のパート D における解答音声を聞くと、話している内容の中身は決して濃くはないが、話し続けることができるだけの能力は備わっており、流暢さは高いと判断し、スコア 286 の受検者は B1 レベルに達していると想定した。その一つ下のスコア 265 の解答音声を聞いても、自身の意見を満足に伝えられているとは決して言えないものの、意見を伝えられるだけの能力は潜在的にあるように感じた。 また、スコア 250 の解答音声では、少しまとまりがある内容で話せるようになってきていることを示唆している。B1 レベルに届きそうだが、まだ足りていないような印象であった。このことから、スコア 264 が B1 レベルの閾値であると判断した。

全体会

全体会での協議において、B1 レベルの閾値を判断する基準として議論に挙がったのは、B1 レベルの流暢さはどの程度なのかという問題であった。グループ A のメンバーが指摘したように、ヨーロッパ圏での B1 レベルとなるとかなり流暢に話せるという印象がある。その B1 レベルの流暢さのイメージを念頭に置くと、GTEC におけるスコア 170 レベルでも B1 レベルの流暢さには届かないとグループ A のメンバーは指摘している。実際、スコア 170 の受検者の 1 分間の発話数は 99 語と少なく、B1 レベルの流暢さとしては物足りなさを感じる。この比較的高得点を取っている受検者であっても、「B1 らしさ」を感じないというグループ A の指摘はグループ B の意見とも一致していた。しかし、閾値決定はある得点を取った受検者の 50%が B1 レベルの域に入っているかどうかで行う。この決定方法を考慮すれば、スコア 282 のレベルでも B1 の下位群に分類できるだけの流暢さがあると言えることから、全体会の協議ではスコア 282 を B1 レベルの閾値にしていいのではないかという結論に至った。

【GTEC CBT A2/B1 の協議】

グループ A

スコア 190 を B1 の閾値であると判定した。スコア 190 の解答が B1 レベルを代表するものではなかったが、B1 レベルに到達する入り口レベル、つまり B1 の下位群に位置する解答音声であると判断した。そのように判断した理由として、B1 レベルの特徴である、話の中で自身の意見をリニアにつなげられる能力を、スコア 191 と 200 の解答音声の中から感じ取れることが挙げられる。実際に、ロジカルコネクターの使用に着目して解答音声を聞いてみると、スコア 200 を取った受検者は first, second, because, also といったロジカルコネクターを用いることができている。これらの理由から、B1 レベルの下位群に分類できる能力が備わっていると判断し、スコア 190 を B1 レベルの閾値とした。

グループ B

グループ A と同様、スタンダードセッティングにおいて閾値判断の素材として用いられたスコア 200 の受検者は確実に B1 に達していると判断した。しかし、スコア 200 の受検者の解答音声を見ると、パート 2 タスク 2 のタスクの意味を読み間違えており、スコアが下がっていることが確認できた。そのため、この受検者は本来、もう少し高いスコアを獲得するだけの能力があると考えられる。このことは、典型的なスコア 200 レベルが必ずしも B1 レベルに達しているとは言えないことを示唆している。そこでスコア 223 の受検者の解答音声を確認してみると、文法的な誤りが少なく、簡単な表現を用いて自分の意見を比較的満足に伝えることができていることから、十分に B1 レベルに達していることが確認できた。これらの理由からグループ B は B1 レベルの閾値をスコア 220 と判断した。

グループ C

他の 2 グループより高い、スコア 231 とスコア 242 の間に閾値があると判断した。グループ C は全体会の中で、一人の解答音声を数種類聞くと、同じ受検者間でもタスクによっての出来具合が異なることが判定を難しくさせたことを指摘している。

スコア 190 の解答音声の中ではパート 3 タスク 1 の出来が良い。しかし、パート 3 のタスク 1 の 90 秒の準備時間と同じく、パート 3 のタスク 2 の 180 秒の準備時間を考慮して解答音声を聞くと、準備時間に話す内容をあらかじめ準備しているからこそ生まれる流暢さにすぎず、準備していた内容を発話しきってしまうと、即座に流暢さが低下する印象がある。そのことから、スコア 190 のレベルでは、準備をせずにある程度満足に自分の意見を話す域には達していないと判断できる。

一方で、スコア 242 の受検者の解答音声を聞くと、発話内で“in natural”と言い誤った箇所を、すぐに“in nature”と言い直す場面が確認され、タスクを遂行しながら、自分の発話をモニターすることができる特徴が現れ、それより下の得点の受検者より、より上位の英語発話レベルに達していると考えた。このことから、B1 の閾値はスコア 240 あたりにあると判断した。

全体会

グループ A は他の 2 グループに比べ、閾値を低めに判断していたが、B1 レベルの下位群を意識して判定したことがその理由にある。そのことからもグループ A のパネルも B1 の特徴が見られるのはもう少し高い点数の解答音声からであることについては同意しており、高く設定することには異論はないと述べている。グループ協議の前の CEFR レベルの目線合わせの際に、A2 レベルと B1 レベルの閾値を設定するために参考にするタスクはパート 2 であるだろうと全グループで共有した。パート 2 に道案内をするタスクがあるが、これは学校の授業の中であまり活動として多くは行われていないものであることからか、受検者にとって取り組みにくかった様子であることは全体の傾向としてうかがえた。各グループが比較的高いスコアを獲得している学習者の解答音声でないと、B1 レベルに達していると確信を持って判断できなかったことは、この日本人英語学習者のタスクへの馴染みのなさに起因すると考えられる。しかし、馴染みがないとはいえ、B1 レベルでは、様々なタスクがある程度上手にこなせる能力を備えていることを考慮すると、スコア 198 の受検者は B1 レベルに達しているとは言えない点については全員の意見が一致した。

また、グループ B が指摘したように、今回参照したスコア 200 の受検者は、パート 2 のタスクで問われていることとは異なる解答をしてしまったことで減点されている。このことから、参照したスコア 200 の解答は GTEC CBT におけるスコア 200 を代表する解答ではないと考えた。また、スコア 220 を超える受検者は、言語的な質も高くなっていることはグループ B とグループ C ともに指摘しており、確実に B1 レベルに達していると言えるだろうということで合意に至った。全体会での協議の結果、A2 レベルと B1 レベルの閾値をスコア 220 とするという結論に至った。

＜スピーキング B1/B2＞

◆ディスクリプタの確認

Spoken Interaction
B1
I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).

B2
I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.

Spoken Production
B1
I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.

B2
I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

（Council of Europe, 2001a, pp. 26-27）

- Spoken fluency の観点では、B2 レベルの学習者は、‘Can produce stretches of language with a fairly even tempo; ’（Council of Europe, 2001b, p.31）と記載があり、発話の内容の構成はまだしっかりしているとは言えないものの、テンポよく話すことができる。
- Spoken fluency では、‘there are few noticeably long pauses’（Council of Europe, 2001b, p.31)とあり、長いポーズがないことも B2 レベルの学習者の特徴である。これら 2 つの特徴から、B2 レベルに到達するためには、頭の中で話の内容を組み立てながらも、同時に途切れずに話を続けることが求められることがわかる。
- また、B1 レベルが、‘can link a series of shorter, discrete simple elements… ‘（Council of Europe, 2001b, p.31）とあるように、リニアに話をつなげられる点が特徴であることに対し、B2 レベルでは、‘can use a variety of linking words efficiently to mark clearly relationships between ideas’ （Council of Europe, 2001b, p.31）と、単にリニアというだけではなく、複数の情報を加えて話すことができることが求められる。

◆各グループ別判定結果の検討

各グループは分科会において、以下に提示する表のように GTEC と GTEC CBT における B1 と B2 レベルの閾値を判断した。

GTEC

グループ A	グループ B	グループ C
320	320	320

GTEC CBT

グループ A	グループ B	グループ C
300 と 310 の間	300 と 310 の間	290

GTEC

GTEC については、スコア 320 のパフォーマンスが B2 に達しているとは考えにくいという指摘もあったが、スコア 320 はスコアの上限値でもあり、能力値としては、かなり高い受検者もスコア 320 に集約されることから、B1/B2 の閾値として、スコア 320 ということで合意に至った。日本人の特徴として、概念化や言語化が得意でないということもあり、同様のタスクを他国の受検者が受検した場合は、もう少しアウトプット量が増えるのではないかという予測も含めての判断であった。

GTEC CBT

グループ A

B1 と B2 の閾値はスコア 300 と 310 の間にあると判断した。閾値決定のためにスコア 300 前後の 4 つの受検者の解答音声を参考にした。パネル全員でのディスクリプタの確認において、B2 レベルが B1 レベルと異なる点は「単にリニアに意見を述べるだけでなく、そこに複数の情報を加えながら意見を展開できる」能力が備わっている点であることを共有した。この点に着目すると、スコア 293 と 297 の 2 つの解答音声、スコア 308 と 312 の 2 つの解答音声との間には差が感じられた。スコア 293 と 297 の受検者もロジカルコネクターを用いながら、比較的長く発話をする能力はあるものの、単にリニアに意見を述べるに留まっていた。スコア 308 と 312 の受検者の解答音声を聞くと、スコア 308 では複数の情報を加えて自分の意見を展開しようとする B2 の特徴が垣間みられ、スコア 312 の受検者は B2 レベルに達している印象があった。このことからスコア 300 とスコア 312 の間に B2 レベルの閾値があると判断した。

グループ B

グループ A と同様に、スコア 300 とスコア 310 の間に B2 レベルの閾値があると判断した。スコア 293 の解答音声を聞くと、文法や語彙の誤りが少なく、高い言語的な質と流暢さを維持しながら、自分の意見を展開できていることから B2 レベルに達していると判断した。しかし、スコア 293 の受検者の素点を詳細に確認すると、求められているタスクの意味を取り違えており、課題に答えられていないことを理由に減点されていることがわかった。スコア 293 の解答に見られる言語的な質の高さから判断すると、この解答音声はスコア 300 レベルを代表するものであると考えられた。判断の際は、言語面およびタスクを達成しているか否か等、複数の観点に留意して判断しなければならないことを指摘している。その点を考慮した上で、他のスコア 300 台の複数の解答音声を視聴すると、十分に B2 レベルに達していると判断できたことから、グループ A と同様、スコア 300 と 310 の間に B2 レベルの閾値があるだろうという結論に至った。

グループ C

他の 2 グループより低いスコア 292 を B2 レベルの閾値と判断した。判断する際には、スコア 292 とスコア 302 の受検者の解答音声を参考材料にした。スコア 292 の受検者の解答音声とスコア 302 の受検者の解答音声を聞き比べてみると、どちらも言語的特徴や音声的特徴に大きな差異はないように感じたことを指摘している。そのことから B2 レベルの下位群を意識し、スコア 290 を B2 レベルの閾値と判断した。しかしながら、特にその点数に強いこだわりがあるわけではなく、スコア 300 を閾値としても異論はないと述べている。しかし、スコア 270 の解答音声は、発話数も少なく、自身の意見を十分に展開できるほどの能力はない。また文法的な誤りもまだ目立つことから、B2 レベルには至らないと感じた。このことからも、B2 レベルの閾値がスコア 292 から 300 に引きあがることに異論はないが、閾値がスコア 290 より下がることはないという結論

に達した。

全体会
　グループ A もグループ B と同様に、スコア 300 台を超えると、B2 レベルに達していると言えると判断した。また、先ほども述べたように、グループ C がスコア 292 に B2 レベルの閾値があるというこだわりがあるわけではなく、スコア 302 の受検者の解答パフォーマンスとの差異をあまり感じなかったことが、閾値をスコア 290 あたりと判断した理由であった。その点で、B2 レベル閾値の判断は、各グループ間で大きな差はなかったと言える。グループ B のメンバーが指摘したように、スコア 293 のオフトピックによる減点を考慮すると、スコア 300 から B2 レベルに達すると考えるのが妥当であり、全体会での協議の結果、スコア 300 が B1 と B2 レベルの閾値であるという結論に至った。

＜スピーキング　B2/C1＞

◆ディスクリプタの確認

　B2 と C1 レベルの閾値を決定するために、全メンバーで、スピーキングの B1 と B2 レベルの self-assessment grid におけるディスクリプタの読み合わせを行った結果、判断の際に留意すべき 2 つのレベルの特徴は以下の FLUENCY と COHERENCE についての 2 点が挙げられた。

Spoken Interaction
B2
I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.

C1
I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skillfully to those of other speakers.

Spoken Production
B2
I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

C1
I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.

（Council of Europe, 2001a, pp. 26-27）

	FLUENCY	COHERENCE
C1	Can express him/herself fluently and spontaneously, **almost effortlessly**. Inly a conceptually difficult subject can hinder a natural, smooth flow of language.	Can produce clear, smoothly flowing, **well-structured** speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2	Can produce stretches of language with a **fairly even tempo**; although he/she can be **hesitan**t as he/she searches for patterns and expressions. There are **few noticeably long pauses.**	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be **some 'jumpiness' in a long contribution.**

(Council of Europe, 2001b, p.31)

- B2 レベルでは、「何ができるか」に関しては C1 レベルとほぼ同等である。

4 つ目の観点に関して、C1 レベルの能力記述文を確認すると、B2 レベルと C1 レベルの差を生み出す特徴は、どのような状況においてどれくらい高い質であることをこなせるかであり、「何ができるか」という観点で 2 つのレベルをみた場合、B2 レベルと C1 レベルでの境はあまりないことも特徴として挙げられた。

- FLUENCY: "Speaking fluency"の項目に、C1 レベルでは '...almost effortlessly'に話すことができるとあることから、一生懸命話そうとする努力が解答音声から伝わる場合には C1 レベルにはまだ達しておらず、B2 レベルに留まっていると判断できる。
- COHERENCE: またスピーチの構成面に関しては、'Can produce clear, smoothly flowing, well-structured speech, ...'とあり、構造がしっかりしているか、まだしっかりとしているとは言い難いかが、B2 に留まっているか、C1 に達しているかの判断基準となる。

◆各グループ別判定結果の検討

グループ A	グループ B	グループ C
340	350	(350)

グループ A

テスト項目の中に C1 レベルを測る問題が用意されていないため、C1 レベルに達しているかどうかを判断するのが難しいことを指摘している。C1 レベルを測るタスクがテスト項目の中にはなかったものの、スコア 340 やスコア 350 レベルの解答音声を聞くと、どんなタスクにも対応できるような能力が十分に備わっていると感じさせるだけの流暢さがあった。また、スコア 333 の解答音声は、言い回しの自然さにおいては C1 と判断できる可能性もあるが、考えながら話を展開している様子がうかがわれたことで、自信を持って C1 とは判定しがたいと考えた。スコア 333 より低い受検者の解答音声を聞くと、話すという行為に慣れている様子が感じとれるものの、学術的な話題について話すことには慣れていないと認識した。C1 レベルでは、場面に応じた適切な言葉を選びながら、自身の発話を展開することが求められる。場面に応じて話す能力という観点からみると、スコア 333 が C1 レベルに達しているとは言い難いと結論付けた。このことから、B2 レベルの閾値はスコア 333 とスコア 350 の間、特にスコア 340 に閾値があると判断した。

グループ B

B2 レベルの閾値を決定するために、スコア 350 を取った 3 人の受検者の解答音声を参考に聞いた。受検者は 3 人とも、複数の視点から自分の意見を展開している点や、当該レベルの確認の際に判断基準となる項目としてあった「苦労して話しているように見えない。一生懸命努力をして話す様子が感じ取ることができない」自然な流暢さがあった点から、C1 の閾値をスコア 350 であると判断した。

グループ C

グループ B と同様、スコア 350 の解答音声を判定材料とした。スコア 350 の 2 つの解答音声を聞き比べたところ、片方は確実に C1 に達していると感じたが、もう一方は C1 レベルに達しているとは言い難かった。また、スコア 370 の受検者の解答音声からは、聞き手を考慮しながら意見を展開する能力が備わっていると言えたことからも、確実に C1 に達していると判断できた。そのことからスコア 350、またはその少し上あたりが C1 レベルの閾値だと判断した。

全体会

C1 レベルの閾値を決定する際に、満点の解答音声とスコア 350 の解答音声を全員で聞いて判断した。満点のレベルでは、言語的な質や流暢さから C1 レベルに十分に達していると言えると全員の意見が一致した。スコア 350 では、グループ C が指摘したように、スコア 350 を取った受検者が必ずしも C1 らしさを感じさせるだけの発話ができているとは言いない。しかし、このあたりから場面に応じた話し方を意識できたり、自然な発話の中でかなり高いレベルで内容のある議論を展開できたりする能力が備わってきているような印象は各グループが受けた。このことから、スコア 350 から上は C1 レベルに達していると判断でき、スコア 350 を C1 レベルの閾値とするという結論に至った。

＜スピーキングまとめ＞

各グループの分科会と全体会の結果、最終的に GTEC と GTEC CBT における各レベルのスピーキングの閾値は、下記に決定した。

Speaking	GTEC の閾値	GTEC CBT の閾値
B2/ C1	-	350
B1/ B2	320	300
A2/ B1	280	220
A1/ A2	190	100

７．４．ライティング

◆閾値設定協議の進め方

ライティングにおいては、中高生を対象としたGTECと、高校生対象のGTEC CBTの2つがあるため、GTECとGTEC CBTにおけるA1/A2、A2/B1、B1/B2の閾値設定協議を並行して行った。なお、GTECのライティングにはC1レベルに相当するテスト項目が存在しないと判断されたため、B2/C1の協議はGTEC CBTにおいてのみ協議された。それぞれのCEFRレベルの閾値設定協議の内容についてA1/A2、A2/B1、B1/B2、B2/C1の協議の順に以下に述べる。

＜ライティング A1/A2＞

◆ディスクリプタの確認

閾値設定を行うにあたり、A1 と A2 を特徴付ける要素を全メンバーで共有するため、self-assessment grid をはじめとする coherence、linguistic range、grammatical accuracy、vocabulary control、orthographic control に関するディスクリプタの読み合わせを行った。A1 と A2 を分けるのに有用な特徴として挙げられたのは以下のような点である。

A1
I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.

A2
I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something.

（Council of Europe, 2001a, pp. 26-27）

- A1 では、いつも完全な文を算出できるという段階には至らず、語句やフレーズのみの表現や、単文を主とした産出ができるレベルである。
- A2 では、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求められるため、完全な文が複数書かれていることが求められる。
- A2 は、主に覚えた文を使っており、時制などの文法的な誤りがあっても、意味が通ればよい段階である。
- 実際の GTEC CBT における A2 と予想される解答の特徴として、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受検者が多いことが挙げられる。

◆各グループ別判定結果の検討

GTEC

グループ A	グループ B	グループ C
196 と 207 の間	196 と 207 の間	168 と 196 の間

GTEC CBT

グループ A	グループ B	グループ C
90 と 101 の間	90 と 101 の間	90 の前後

【GTEC A1/A2 の協議】

グループ A
　スコア 196 とスコア 207 の間と判断した。閾値設定の理由として、スコア 207 以下にも言語的によくできているものもあるが、A2 と容易に認められる解答では受検者の課題文に対する「意見」が述べられているために、全体の文章の長さが格段に長くなる点を挙げ、スコア 207 以上を確実な A2 と判断した。

グループ B
　グループ A と同じく、スコア 196 とスコア 207 の間とし、グループ B もまた、A2 とされる解答には十分な量の課題文に対する「意見」が述べられている点に同意した。

グループ C
　スコア 168 と 196 の間とした。他のグループが A1 と判断したスコア 196 の解答を A2 と判断した理由は、作文のテーマが比較的高度な社会的な内容だったために十分なパフォーマンスが見られなかったが、より易しい身近なことに関するテーマであればもっと力を発揮できていたであろうことが言語使用の視点からうかがえることを指摘した。

全体会
　全体の閾値設定の議論として、GTEC の評価基準と CEFR の視点からの判断基準が異なる部分があるため、CEFR の視点からは A2 相当の言語的な操作が部分的に垣間見られると判断される場合でも、「意見が書かれていること」といった基準が満たされておらず、十分な作文量と判断要素が観察されないために、全体の点数としては低い点が与えられていると考えられる例がいくつかあることが挙げられた。また、スコア 196 の解答のように、実力よりも高いライティング能力が要求される社会的なテーマでの作文に挑戦しているために点数が振るわなかったが、受検者のレベルにあったテーマであれば、よりよい作文ができていたかもしれないことがうかがえる解答もあった。そのため、CEFR の視点で A2 と考えられる言語的な操作が部分的に見られるスコア 196 と 168 の間であるスコア 190 を閾値とするという結論に至った。

【GTEC CBT A1/A2 の協議】

　グループ A はスコア 90 とスコア 101 の間、グループ B はスコア 90 とスコア 101 の間、グループ C はスコア 90 の前後と判断した。

全体会
　GTEC CBT に関しては、3 グループ一致してスコア 90 周辺を閾値とした。その理由として、スコア 90 の解答の中にはいくつかのテスト項目に対して白紙の解答があることや、スコア 90 以下の解答には、問題指示文の意味が理解できていないことがうかがえる解答があることが挙げられた。一方、スコア 101 の解答から上の解答は、複数の文で構成されているものが多く、さらに複数のアイディアをつなごうとしていることがうかがえることから、確実に A2 であるという判断で一致したため、閾値をスコア 90 とスコア 101 の間であるスコア 100 に決定した。

＜ライティング A2/B1＞

◆ディスクリプタの確認

A2
I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something.

B1
I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

（Council of Europe, 2001a, pp. 26-27）

- A2 では、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レベルである。
- B1 では、A2 に比べて語彙や文法の正確さが増すため、文法の誤りやスペリングミスが目立つものは B1 ではないと考えられる。
- B1 では、文章の中でアイディアをうまくつなげて、より長いパラグラフライティングができるようになっているレベルである。

◆各グループ別判定結果の検討

GTEC

グループ A	グループ B	グループ C
250 と 269 の間	269 と 282 の間	228 と 250 の間

GTEC CBT

グループ A	グループ B	グループ C
220 と 228 の間	220 から B1	220 と 228 の間

【GTEC A2/B1 の協議】

グループ A

スコア 250 と 269 の間と判断した。スコア 250 とスコア 269 の解答を比較するとどちらも文法の誤りがあるという共通点があった。しかし、スコア 250 は複雑な文構造で英語が書けているものの、自分の主張に関する十分な理由が記述されていないのに対し、スコア 269 の解答は複雑な文構造が見られることに加え、より明確なアイディアのつながりが認められたためにスコア 250 とスコア 269 の間に閾値があり、269 は B1 レベルであると判断した。

グループ B

スコア 269 とスコア 282 の間と判断した。グループ A が B1 と判断したスコア 269 の解答は語彙的な誤りが多く見られたため A2 としたが、共通資料外のスコア 269 周辺の解答を他にもいくつか参照し、B1 の要素を多く含むものがあったため、スコア 269 が閾値でも妥当であると述べた。

グループ C

スコア 228 とスコア 250 の間と判断した。他の 2 グループよりも低いスコア 250 から B1 とした理由として、他の 2 グループが判断要素とした内容の深さは知識に関することであるため、英語を産出する力の判断要素として重要視しなかったとした。また、A2 のライティングはかなりシンプルなものであるという CEFR のディスクリプタを考慮すると、ある程度まとまりのある文章として認められるスコア 250 から B1 に区分されても良いという判断に至ったとした。

全体会

これらの要素を踏まえ、閾値の決定のための議論に移った。CEFR のディスクリプタに基づいた言語的な操作を重視しながら、共通資料外の解答もいくつか参照すると、スコア 237 の解答では量は産出できているものの、表現がかなり拙いという意見で一致した。スコア 250 になると作文量も増え、文法や構成としてもよりよい印象になるとし、スコア 237 とスコア 250 の間であるスコア 240 を閾値とするという判断に至った。

【GTEC CBT A2/B1 の協議】

グループ A

スコア 220 と 228 の間と判断した。その理由として、スコア 220 はタスクの中に出てくる Email の書き方やグラフ描写の際の数値への言及などで不十分な点が多くあることが挙げられた。スコア 228 もスコア 220 で見られたようにグラフ描写を迂回していることがうかがえるが、言語運用力や、つながりのあるパラグラフ構成ができているかという視点からみると B1 と言えると判断したと述べた。

グループ B

スコア 220 から B1 と判断した。その理由として、グループ A の「スコア 228 から B1 である」という意見には同意するが、日本の高校生は何かを客観的に描写するというタスクにあまり慣れていないであろうという背景を考慮すると、ライティングの解答に実力が十分に現れていない可能性があるとし、グループ A よりも低いスコア 220 から B1 と判断したと述べた。

グループ C

グループ A と同様に、スコア 220 とスコア 228 の間と判断した。グループ A と同様にスコア 228 から確実に B1 であると判断した。

全体会

これらの要素を考慮し、閾値の決定のための議論に移った。GTEC CBT に関しては、スコア 228 の解答は確実に B1 であるという点で 3 グループが一致した。これを踏まえ、閾値は B1 の下限にあることを考慮すると、スコア 228 よりも下のスコア 220 が閾値として妥当であるという結論に至った。

＜ライティング B1/B2＞

◆ディスクリプタの確認

B1
I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

B2
I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.

（Council of Europe, 2001a, pp. 26-27）

- B2 は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。
- B2 は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も求められるレベルである。
- B2 は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、アイディアを一貫して述べることができる。
- 実際に、B1 と B2 の閾値周辺は大学入学許可の目安とされるレベルであるということからも、アカデミックライティング相当のライティングの力が求められるレベルである。

◆各グループ別判定結果の検討

GTEC

グループ A	グループ B	グループ C
305 と 320 の間	300 が閾値	300 が閾値

GTEC CBT

グループ A	グループ B	グループ C
270 が閾値	270 と 292 の間	300 が閾値

【GTEC B1/B2 の協議】

グループ A
　スコア 305 とスコア 320 の間と判断した。その理由として、スコア 305 の解答には言語的に高度な語彙や表現の使用が認められるが、説得力のある議論ができているかという視点からは、対立する意見が書かれていない点で B2 には至らない印象である点を挙げた。

グループ B とグループ C は、グループ A よりも低いスコア 301 が閾値と判断した。

全体会
　全体の議論として、2 グループがスコア 300 を閾値と判断したのに対し、グループ A がそれよりも少し高いスコア 305 とスコア 320 の間と判断したが、閾値には B1 と B2 の人数が 50%ずついる前提であることを考えると、言語的にも誤りが少なく、議論としても構成がしっかりしてくるスコア 300 を閾値とするのが妥当であるという結論に至った。

【GTEC CBT B1/B2 の協議】

グループ A
　スコア 270 が閾値と判断した。スコア 270 の解答だと多面的に意見を述べるという点で、まだ B2 というには決め手に欠けるが、閾値としては妥当だと判断した。

グループ B
　スコア 270 とスコア 292 の間であると判断したが、スコア 280 台の解答が共通資料の中になかったため、共通資料外のスコア 280 周辺の解答を参照してから判断したいとした。

グループ C
　スコア 300 が閾値と判断した。その理由として、最後のタスクでは資料を引用するように指示があるため、どの解答でも引用はなされているものの、大学レベルで必要とされる引用の作法ができていないと B2 には至っていないであろうという点を挙げた。

全体会
　まず、グループ B が判断要素として参照する必要があるとしたスコア 280 の解答について、2 つの Booklet 外の解答を全員で確認した。その結果、これらのスコア 280 の解答は、どちらも B2 と言うには文章が平易で稚拙であることから B1 相当であるという意見で一致した。スコア 292 の解答に関しても、B2 には至っていないという印象であった。しかし、複数あるテスト項目の中で、B1 と B2 の閾値を決める際の最も有力な手がかりになるであろう最後の作文タスクは、25 分間で 250 語という制限がある中で書かれたものであることを考慮すると、CEFR のディスクリプタの視点から B1 と B2 の違いが区別できるほどの解答データが誘引できていない可能性があることが指摘された。これに関して、グループ C は、グループごとの分科会の際に B2 相当の解答を求めてスコア 340 の解答まで検討したが、上記のような制限があることを考慮し、スコア 300 周辺に戻ってきたという経緯があったと述べた。これらの意見を踏まえ、B1 と B2 の判断が難し

いスコア 292 周辺の解答を「B2 で必要とされるアカデミックなエッセイの下書きとして認められる内容で書かれているかどうか」という視点から検討した結果、スコア 290 が B1 と B2 の閾値として妥当であるという結論に至った。

＜ライティング B2/C1＞

◆ディスクリプタの確認

B2
I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.

C1
I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.

（Council of Europe, 2001a, pp. 26-27）

- C1 の作文にはほぼ誤りがなく、複雑なトピックに関しても非常によく構成された、かなり長い文章を書くことができるレベルである。

◆各グループ別判定結果の検討

GTEC CBT

グループ A	グループ B	グループ C
350 が閾値	350 が閾値	350 が閾値

【GTEC CBT　B2/C1 の協議】

グループ A、グループ B、グループ C ともにスコア 350 から C1 で一致した。

全体会

スコア 350 を閾値とするという結論に至った経緯として、まず Booklet 内にあったスコア 324 までの解答に、C1 に達する解答はないと判断されたため、Booklet 外の解答からより高い点数であるスコア 350 の解答を 2 つ参照した。その結果、1 つ目はアカデミックな表現があまり見られないという点で C1 には至らないと判断された。しかし、B1/B2 の議論内にも述べた通り、B2 と C1 とを分ける手がかりとなるであろうパート 3 のタスク 2 では、25 分間で 250 語という制限の中で書かれたものであることもあり、本来 C1 の力がある受検者においても、C1 の要素を含む内容の作文が誘引できていない可能性があることが再度指摘された。そこで、2 つ目のスコア 350 の解答を参照したところ、1 つ目の解答に比べてよりアカデミックな言語的要素や構成が見受けられ、それらが C1 相当とみなすことができるという意見で一致したため、閾値をスコア 350 と設定することに決定した。

＜ライティングまとめ＞

以上の議論を経て、GTEC と GTEC CBT における各 CEFR レベルの閾値を以下のように決定した。

Writing	GTEC 閾値	GTEC CBT 閾値
B2/ C1	-	350
B1/ B2	300	290
A2/ B1	240	220
A1/ A2	190	100

７．５．結果

各技能別、各 CEFR レベル別の検討を経て、GTEC、GTEC CBT それぞれの閾値は下記のようになった。

	GTEC					GTEC CBT				
	Reading	Listening	Writing	Speaking	Total(4)	Reading	Listening	Writing	Speaking	Total(4)
C1						330	340	350	350	1370
B2	280	290	300	320	1190	280	290	290	300	1160
B1	220	220	240	280	960	220	220	220	220	880
A2	150	160	190	190	690	150	160	100	100	510
A1	上記A2未満					上記A2未満				

※GTEC スコア　各技能 0～320、Total 1280
※GTEC CBT スコア　各技能 0～350、Total 1400

８．考察

本研究は、英語の 4 技能外部試験入試が検討されている中、これまで以上に、外部検定試験と CEFR の関連付けが重要性を増しつつある流れの中、その分析・研究を行うために計画された。

＜使用データ＞

2015 年度にも同調査を行っていたが、その際はフィールドテストのデータを用いていた。今回は、GTEC、GTEC CBT ともに、その当時の最新の本番試験のデータを使用した。本番でターゲットとしている受検者が取り組んだデータのみを使用しているため、より精緻な設定ができたと言える。

＜Borderline Person のイメージを全員で一致させる工程＞

スタンダードセッティングの一連の流れの中で、各閾値設定前の冒頭に必ず行った、CEFR ディスクリプタを読みながらの全員のメンバーによる Borderline Person 英語力のイメージを共有する工程は、ブレなく適正に閾値を設定するためには必要不可欠な工程であったと言える。各閾値の上にいる能力を持つ学習者が 50%の確率で正答できるレベルのイメージを全員のパネルが共通に持つまでやや時間を要したが、パネルの構成員は、語彙、言語テスティング、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、各パネルの中にあるイメージが徐々にすりあわされていった。

＜困難度についての議論＞

閾値設定の議論に際して、例えば、パッセージの難易度のみを考慮に入れるべきか、テスト項目の難易度も合わせて判断すべきなのか問題提起が投げかけられた場面があった。パネルで議論の末、項目困難度は純粋にパッセージの困難度ではなく、テスト項目の難易度も合わせた困難度になっていることから、パッセージのみを見て判断するのではなくテスト項目の難易度も合わせて評価する必要がある、また迷う場合には、該当の問題アイテムの困難度前後のアイテムを参照することにより、総合的に判断するということで合意した。

このようなことから、パネルの構成員として、様々な分野での有識者が参画することにより、異なった視点での問題提起が行われ、議論しあうことで、より客観的でバランスのとれた判断ができたと言える。

＜採用した分析手法の有効性＞

今回のスタンダードセッティングで用いた Bookmark Method、IRT のデータをもとにして行った Contrasting-Group Method は、IRT で算出された困難度という客観的な情報に基づき判断を行うことができた。各閾値検討にあたっては、CEFR レベルの感覚を共有するトレーニングをこなすことで、それぞれが大きくぶれずに判断できることがわかった。最初に、2 名グループでの検討、さらに次の工程で全員での議論を経て結論を導き出す方法は、様々な観点において極力漏れのないように判断することを可能にした。各グループの判断がぶれた場合でも、各グループが判断根拠を説明しあい、前後のレベルの問題や解答パフォーマンスを参照することで、全員の認識がしだいに合ってくるなど、CEFR レベルに詳しい有識者の間での話し合いでテストの Benchmarking が有効に機能することも示された。

＜あらゆる観点からの総合的な判断＞

リスニングでは、印刷された英文スクリプトやテスト項目だけでなく、パッセージを読み上げた音声を聞くことも行い、受検者がテストを受けている設定に近い状況で判断を行うことができた。同様に、スピーキングについても、書き起こしたテキストを参照するだけでなく、発音や流暢さ等の情報も合わせて解答パフォーマンスを音声で判断することができた。

＜CEFR レベルに合ったテスト項目、素材選定の必要性＞

ライティングの閾値判定の際に、A1/A2 などの、より下位レベルの閾値設定に際しては、もともと Booklet には、GTEC の Advanced タイプの解答のみ掲載していたが、タスクの困難度が高いために、解答の語数が少なく、内容的にも薄かったため、より易しいレベルの Basic、Core タイプの解答を参照し、閾値を設定した。習熟レベルに合ったタスクに対しての解答パフォーマンスを参照することが重要であるということが分かった。

スピーキングとライティングにおいては、GTEC CBT では、CEFR のディスクリプタに合わせて、複数のタスクを 3 パートに分けて課している。つまり、パート 1 からパート 3 にかけて緩やかに難易度が上がっていくような設計になっている。そのため、より低いレベルの閾値設定では、パート 1 やパート 2 前半を主に参照し、より高いレベルになるにしたがって、パート 2 後半やパート 3 の解答を参照することで、より適切に判断ができたと思われる。

今回の調査は、2016 年度に実施された本番試験の問題を用いての調査であったが、今後、何年か後のタイミングで再度調査をするなど、一定の期間をあけて見直しをかけるなどが望ましいであろう。

９．結論

本研究は、GTEC、GTEC CBT の本番データを使用し、CEFR との関連付けを行った。このため、フィールドテストのデータに基づいた 2015 年度調査に比して、より精度の高い基準の設定ができたと言える。スタンダードセッティングの経験値が高く、CEFR を熟知したメンバーが、Borderline Person の英語力のイメージを共有し、判断を行った。このことで、３つのグループが独立して行った判断に大きなブレはなく、その後の議論を経て、合意することができた。

スタンダードセッティングの手法としては、これまでに様々な手法が提案されているが、本研究では、リスニング・リーディングの受容技能のテストには Bookmark Method を、スピーキング・ライティングの発表技能のテストには IRT のデータをもとにして行った Contrasting-Group Method を行った。どちらの手法においても、IRT で算出された困難度という客観的な情報に基づき判断を行うことができた。スタンダードセッティングは本質的に主観的な判断を伴う行為であり、それ故、ある程度の判断の個人差は当然生じる。本研究で用いられた IRT に基づいたスタンダードセッティングの手法では、そのぶれの範囲に存在するテスト項目や受検者のパフォーマンスを見ることで、より精度の高い判断をすることができた。

今回の調査は、GTEC、および GTEC CBT の日本人受検者のデータに基づいている。こうしたデータに基づいたスタンダードセッティングの結果は、外国人受検者のデータの場合と異なるかどうかの検証が、今後必要となるであろう。

参考文献

加藤健太郎・山田剛史・川端一光 (2014). R による項目反応理論 オーム社

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.

中津原文代 (2013). 能力基準としての Can-do statements とテストの妥当性を検証する「社会定期・認知的枠組み」 (Socio-cognitive Framework)について. 『言語教育評価研究』第３号.pp.54-63.国際交流基金.

文部科学省 (2014). 英語教育の在り方に関する有識者会議 英語力の評価及び入試における外部試験活用に関する小委員会　審議のまとめ概要.以下より入手可能：http://www.mext.go.jp/component/b_menu/shingi/toushin/__icsFiles/afieldfile/2014/08/20/1351000_01.pdf

Cambridge ESOL (2011) . Using the CEFR: Principles of Good Practice, Bambridge: Cambridge ESOL. Available online at http://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf

CEFTrain Project. http://www.helsinki.fi/project/ceftrain/index.php.35.html

Council of Europe. (2009). A Manual: Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. 以下より入手可能：
URL http://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf

Council of Europe. (2001a). COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES:LEARNING, TEACHING, ASSESSMENT. Cambridge University Press.以下より入手可能：
http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Council of Europe. (2001b). Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales. Cambridge University Press. 以下より入手可能：https://rm.coe.int/168045b15e　(最終検索日　2017年９月１日)

Richard J. Tannenbaum & E. Caroline Wylie. (2008). Linking English-Language Test Scores Onto the Common European Framework of Reference: An Application of Standard-Setting Methodology. ETS.

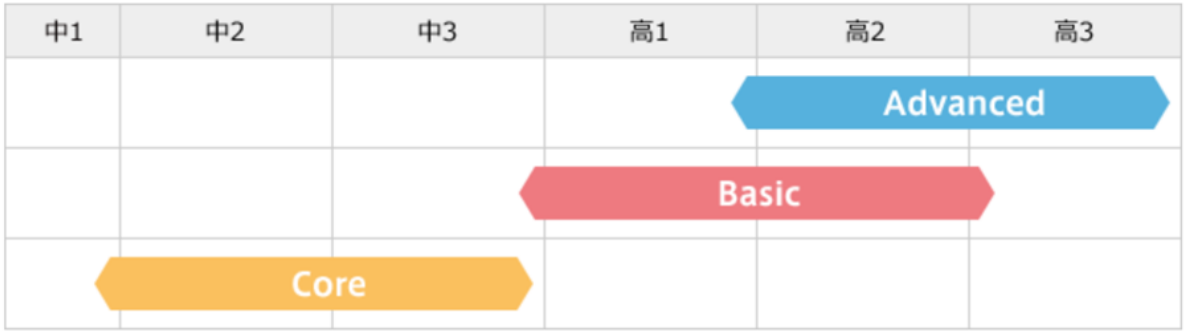
付録

１．GTEC 問題概要・実施時間
２．GTEC 問題の特徴および問題構成
３．GTEC CBT 問題概要
４．GTEC CBT 問題の特徴および問題構成

１．GTEC 問題概要・実施時間

GTEC 受検タイプについて

GTEC は Advanced、Basic、Core の３つの難易度でテストを提供している。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している。



実施時間

	Core	Basic	Advanced
Reading	32分	45分	
Listening	18分	25分	
Writing	20分	20分	
TOTAL	**70分** ※1	**90分** ※2	

※1 Coreタイプは問題冊子がReadingで1冊（32分）、Listening、Writingで1冊（38分）となっており、分割実施が可能です。

※2 Advancedタイプ、Basicタイプは問題冊子がReadingで1冊（45分）、Listening、Writingで1冊（45分）となっており、分割実施が可能です。

	Core	Basic	Advanced
Speaking	25分※3		

※3 進行の具合によって前後する可能性があります。

http://www.benesse-gtec.com/fs/about/ab_outline

２．GTEC 問題の特徴および問題構成

【リスニング】

1. 実践的な言語テストとして、リスニングの出題割合を重視

テスト全体の中でリスニングの占める所要時間は2～3割、問題数では5割前後とリスニングの出題割合を重視しています。十分な出題数の中で、高い精度でリスニング力を測ることができます。

2. 即応性、情報選択、要点理解など多角的にリスニング能力を測定

正確な聞き取りだけでなく、素早く反応して答える力や、目的に応じて必要な情報を選択して聞き分け、要点を理解する力など、多角的で実用的なリスニング能力を測定します。

3. 現実の場面や状況設定で使える課題（タスク）を出題

英語を聞いて何かをしなくてはならないような現実の場面や状況設定の中で使える課題（タスク）を出題に取り入れることで、単なるテストのための問題ではなく、受験者が問題を通して英語を使っている実感を得ることができるようになっています。

	Core		Basic		Advanced	
テスト内容	設問数	時間(分)	設問数	時間(分)	設問数	時間(分)
A 写真・イラスト説明問題	8	4	10	6	10	6
B 会話応答問題	8	4	10	5	10	5
C 課題解決問題	8	5	10	8	10	8
D 要点理解問題	8	5	10	6	10	6
TOTAL	**32**	**18**	**40**	**25**	**40**	**25**

【リーディング】

1. 英文読解の能力を多角的に測定

語い・語法レベルの読解基礎力、速読的な読解力、そして精読的な読解力などの各分野に着目して構成された多角的な出題です。

2. 速読的な読解力では、特に情報検索、概要把握する力が焦点

速読的な読みにおいては、短時間の中で必要な情報を引き出す情報検索（scanning)の力をみる問題、そして全体を大づかみして概要を理解する概要把握(skimming)の力をみる問題を出題しています。

3. 海外在住のネイティブスピーカーの執筆者による「リアル」な英語素材文

海外在住のネイティブスピーカーの執筆者が作成した問題素材文原案を、日本国内の有識者を交えてGTEC for STUDENTS編集部が吟味することで、「リアル」な英語素材でかつ日本の中高生の英語力測定に適した出題を行っています。

	Core		Basic		Advanced	
テスト内容	設問数	時間(分)	設問数	時間(分)	設問数	時間(分)
A 語い・語法問題	10	5	12	6	14	7
B 情報検索・概要把握問題	10	12	12	14	14	14
C 要点理解問題	8	15	12	25	15	24
TOTAL	**28**	**32**	**36**	**45**	**43**	**45**

【スピーキング】

Part A　Reading Aloud　サンプル問題 →

音読

対話の応答文を読み上げる形式の出題で、状況や英文を理解したうえで、正確な発音で音読ができるかどうかを診断します。

問題数	準備時間	解答時間
6問	Advanced：各5秒 Core：各10秒	各10秒

Part B　Listening and Responding　サンプル問題 →

質問を聞いて応答する

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うえで、適切に応答する力があるかどうかを診断します。

問題数	準備時間	解答時間
4問	各10秒	各15秒

Part C　Telling a Story　サンプル問題 →

ストーリーを英語で話す

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。

問題数	準備時間	解答時間
1問	Advanced：20秒 Core：30秒	Advanced：60秒 Core：40秒

Part D　Expressing Your Opinion　サンプル問題 →

自分の意見を述べる

身近なテーマに対して、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。

問題数	準備時間	解答時間
1問	60秒	Advanced：60秒 Core：40秒

【ライティング】

1. 自由記述形式の出題

与えられたテーマに対して自分の考えを表現する問題1題を、自由記述形式で出題します。限られた時間の中で自分の意見を説得力を持って表現する力を測定します。問題タイプごとの出題方針は、以下の通りです。

Core…日常生活の中から、個人が経験したことをもとに自分の意見と理由を述べます。

Basic…日常生活と社会との接点を少し持たせつつ、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。

Advanced…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。

2. 複数名の採点者による観点別評価

ライティング答案は海外の採点拠点に送られます。1枚の答案はトレーニングを積んだ複数名の外国人採点者によって、まず「意見」、「理由」、そして「語い」、「文法」、「構成・展開」の5つの観点に分けて採点されます。英語で自分のメッセージを的確に相手に伝えることができるかを測ることができます。

3. IRT計算にもとづくスコア算出

出題テーマによる書きやすさ、書きにくさの違いを調整するため、採点者が採点を行った後、IRT※といわれる統計処理を行うことで、ライティングスコアが算出されます。それにより異なる出題テーマであっても同じ指標でスコアを比べることができるようになります。

※ IRT=Item Responce Theory（項目応答理論）

	Core		Basic		Advanced	
テスト内容	設問数	時間(分)	設問数	時間(分)	設問数	時間(分)
WRITING / 自由記述式	1	20	1	20	1	20

URL:http://www.benesse-gtec.com/fs/about/ab_content

３．GTEC CBT 問題概要

測定技能	問題数	試験時間	満点スコア	解答方法
Speaking	7問	約20分	350点	マイク付きイヤホンでの音声録音による解答
Writing	6問	約65分	350点	キーボードでのタイプ入力による解答
Listening	約40問	約35分	350点	マウスクリック形式による選択
Reading	約40問	約55分	350点	マウスクリック形式による選択
Total		約175分	1400点	

※ コンピュータによる実施のため、試験時間は受験者によって異なります。
※ 試験開始～終了までに休憩時間はありません。

http://www.benesse-gtec.com/fs/about/ab_outline

４．GTEC CBT 問題の特徴および問題構成

【リスニング】

学生生活での会話や講義からの出題で、課題解決に必要な情報を「聞く力」を測定

大学の講義や学生同士の会話など、学生生活でよく出合う場面設定の出題により、実践的なコミュニケーション力を測定します。聞き取った内容をもとにタスクに取り組む、課題解決型の出題です。

場面設定	出題内容	問題数	試験時間
大学で経験するような講義や学生生活の場面	・英文を聞いて、全体の概要および要点を把握する問題 ・与えられたタスクを行うために必要な情報を聞き取る問題 ・会話や講義の内容の論点を把握する問題 ・話された内容から話し手の意図や話し手との関係性を理解する問題	約40問	約35分

【リーディング】

学生生活で遭遇する情報や、講義内容などの出題で「読む力」を測定

学生生活や講義で実際に遭遇する幅広い媒体・情報から出題し、目的に合わせて概要や要点を把握する力を測定します。アカデミックな文章において、著者の意図を理解して言外の意味を読み取る出題もあります。

場面設定	出題内容	問題数	試験時間
大学で経験するような講義や学生生活の場面	・英文を素早く読んで、全体の概要および要点を把握する問題 ・目的に合わせて、英文の主旨や詳細情報を読み取る問題 ・英文を書いたり話したりするために必要な情報を読み取る問題 ・筆者の意図を推察する問題 ・さまざまな意見の中から筆者の意見を特定する問題	約40問	約55分

【スピーキング】

会話応答力から自分の意見を述べる力まで幅広く「話す力」を測定

学生生活で英語を使用する場面を想定した、バリエーションのある出題が中心。
スピーキング力を構成するInteraction（やりとり）とProduction（発表）の両方を測定します。

	パート	出題内容	問題数	試験時間	採点基準
1	会話応答問題 Listening and responding	質問に対して即座にかつ適切に応答する問題	1題 (小問6問)	2分	・質問の意図に沿って、適切な応答ができているか
2	情報伝達および照会問題 Delivering and asking for information	ウェブサイトなどから得た情報を整理して説明する問題や、自ら質問する問題	3問	6分	・内容に基づいて、相手に伝わるような描写説明、問いかけができているか
3	意見展開問題 Expressing your opinion	・与えられたトピックに対して、自分の考えや経験に基づいて意見を述べる問題 ・他者の質問に対して即座に応答する問題	3問	12分	・自分の意見が述べられているか ・自分の意見をサポートする理由が添えられているか ・流ちょうさ

【ライティング】

バリエーション豊かな出題で多角的に「書く力」を測定

項目記入、Eメール作成、意見展開文など様々なタイプの出題で、書く力を測定します。
Eメールを書いたり、あるデータについて意見を述べるなど、学生自身が英語を使って書く実際の場面を想定した出題を行います。

	パート	出題内容	問題数	試験時間	採点基準
1	質問用紙記入問題 Completing a questionnaire	与えられた英文と状況設定を読み、条件にあった内容を書く問題	1題 (小問4問)	4分	・条件に合った内容が書けているか
2	Eメール作成問題 Writing an e-mail	与えられた状況設定を読み，条件にあったEメールを書く問題	3問	21分	・条件に合ったEメールが書けているか
3	意見展開問題 Writing an essay	・統計データなどに対して、自分の意見やその意見の背景となる理由などを書く問題 ・与えられたトピックに対して、他者の考えなどを取り入れながら意見を展開する問題	2問	40分	・自分の意見が書けているか ・自分の意見をサポートする理由が書けているか ・言語運用能力、論理の一貫性

http://www.benesse-gtec.com/cbt/about/composition

Table 2. *Common Reference Levels: self-assessment grid*

		A1	A2	B1
UNDERSTANDING	**Listening**	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
SPEAKING	**Spoken Interaction**	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.
WRITING	**Writing**	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

B2	C1	C2
I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.
I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Council of Europe, 2011, pp.26-27