研究課題	1 人 1 台の学習端末を活用した英語スピーキング指導と AI による音声自動評価の可能性に関する実証研究
副題	～最新の AI と音声工学技術を応用した英語スピーチの自動評価システムの妥当性と信頼性の検証～
キーワード	スピーキング・テスト、ESAT-J、ChatGPT、手動評価、自動評価
学校/団体 名	全国英語教育研究団体連合会研究部
所在地	〒111-0041　東京都台東区元浅草 1-6-22
ホームページ	

１．研究の背景

GIGA スクール構想により、全国の学校で１人１台の学習端末が提供されている。しかし、マイク付イヤホンと学習端末を連携することで、よりクリアな音声を取ることができることを生かした練習活動を行うことで、英語スピーキング力がどれだけ伸長するかは十分に検証されておらず、実践研究として取り組む意義がある。また、学習者音声を AI や音声認識を活用して自動採点する試みも始まっているが、自動評価得点がベテラン教員の手動評価をどれだけ反映するかは未知数であり、検証する必要がある。

２．研究の目的

スピーキング・テストの評価を、ベテラン教員の手動評価と AI・音声工学技術を使った自動評価の 2 種類で行い、その評価を比較し、AI による自動評価の信頼性・妥当性を検証する。また、ICT を活用した新しい指導法を継続して行い、英語スピーキング力の変化と英語で話すことへの不安や苦手意識の変化を、プリ・テストとポスト・テストの結果と質問紙から調査し、ICT を活用した指導法の有効性を検証する。

３．研究の経過　研究方法

3.1 手動評価と自動評価の関係

まず４月～５月にスピーキング・テストの形式についての研究から行った。東京都教育委員会が 2022 年から中学生に対して行っている「スピーキング・テスト」（ESAT-J）の形式を参考にして、以下の形式とした。

Part A　音読問題　（提示された英文を音読する）

Part B　応答問題　（提示された質問に応答する）

Part C　絵を使った視覚情報の説明問題（４コマイラストの内容を描写する）

Part D　意見発表問題（提示されたトピックで自分の意見を発表する）

この形式で、中学高校それぞれの語彙レベル・文法レベルに合わせたスピーキングのプレテスト問題・ポストテスト問題、および、アンケートを６月～８月にかけて新規に作成した。

９月～１１月にかけて、そのテスト問題の手動採点基準、自動採点基準を作成した。手動評価の基準（表 1）は教員が普段授業で使用しているような主観的な表現（「おおむねできている」

等）の残ったものになった。これ対して自動評価の採点基準（表 2）については ChatGPT が機械学習しやすいように、「文法のミス 1 か所につき-1 点」など、できていない個所を数値で表す減点方式の採点基準を作成し、ChatGPT に組み込んだ。

表 1　スピーキング・テスト評価基準表（手動評価）

評価	問題A		問題B		問題C		問題D	
	評価項目							
	発音・強勢・同化連結	スピード・イントネーション・ポーズ	音声面	意味理解・語彙・文法の選定	音声面	意味理解・語彙・文法の選定	音声面	意味理解・語彙・文法の選定
	評価基準							
3 （100～70%）	・個々の語の発音、アクセント、がほぼ正しい。 ・語の同化・連結もほぼ自然に行われている。 ・若干誤りがあっても、文の内容を十分に伝えることができる読みになっている。	・文を読む速さ、抑揚、ポーズが適切で、内容を理解した読み方になっている。 ・途中、間があっても、不自然さはない。（知らない単語の前で一瞬言いよどむなど） ・最後まで読み切れている。	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	・質問の意味が理解できており、適切な内容・自分の考え・理由などを一定程度述べるが、答え方が単語のみまたは表現方法が不十分である。 ・文法上の誤りがやや目立つ。または語彙や表現が平易なものが目立つ。 ・相手に尋ねる問題は、文法・語彙・表現方法にやや誤りがあるが、相手に質問の内容を伝えることはできる。	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	・4 つのイラスト全てについて、順を追って説明している。 ・接続詞、代名詞などを効果的に使い、文で答えている。 ・様々な文法、語彙を使用して出来事を説明している。	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	・効果的な理由や具体例を用いて明確な意見を述べている。 ・適切な語彙・文法を使っており、意味が分かりやすい。
2 （69～40%）	・個々の語の発音、アクセント、いくつか間違いがある。 ・語の同化・連結は自然とまではいかないが、ポツポツ1語1語を切った読みにはなっていない。 ・いくつか誤りがあったりややぎこちない発音でも、文の内容をおおむね伝えることができる読みになっている。	・文を読む速さ、抑揚、ポーズなどでやや不自然なところがあるが、一応内容を理解できる読み方になっている。 ・最後まで読み切れている。	・やや不自然な聞き方で答えている。 ・自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。	・質問の意味は理解できて、適切な内容・自分の考え・理由などを一応答えているが、答え方が単語のみまたは表現方法が不十分である。 ・文法上の誤りがやや目立つ。または語彙や表現が平易なものが目立つ。 ・相手に尋ねる問題は、文法・語彙・表現方法にやや誤りがあるが、相手に質問の内容を伝えることはできる。	・やや不自然な聞き方で答えている。 ・自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。	4 つのイラストを一連の出来事として説明しているが、1 つのイラストの説明が不十分である。 ・接続詞、代名詞などを使おうとしており、文で答えている。 ・文法と語彙に間違いが多少あるが、理解できる。	・やや不自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。	・理由や具体例が十分でなく説得力に欠ける。 ・ほぼ適切な語彙・文法を使っており、意味が分かりやすい。
1 （39～1%）	・個々の語の発音、アクセントに大きな間違いが多い。 ・語の同化・連結がなされず、1語1語区切った読みになっている。 ・文の内容を伝えることが困難な読みになっている。	・文を読む速さ、抑揚、ポーズなどがかなり不自然であり、内容を理解できていない読みになっている。 ・最後まで読み切れていない。	・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。	・質問の意味を理解できておらず誤った解答をしている。 ・発話として成り立っていない、または単語の羅列のみで自分の意見や質問の内容を相手に伝えることができない。 ・解答が途中で終わってしまっている。	・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。	・4 つのイラストの内、2 つか 3 つのイラストの出来事を説明している。 ・接続詞や代名詞などが適切に使われていない。 ・文法と語彙に間違いが多く、理解しづらい。	・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。	・理由や具体例を挙げて意見を述べようとしている。 ・語彙・文法に誤りがあり、意味が分かりにくい。
0 （0%）	無回答・無音	無回答・無音	無回答・無音	無回答・無音	無回答・無音	無回答・無音	無回答・無音	無回答・無音

表 2　スピーキング・テスト評価基準表（客観評価）

評価	問題A		問題B		問題C		問題D	
	評価項目							
	発音・強勢・同化連結	スピード・イントネーション・ポーズ	音声面	意味理解・語彙・文法の選定	音声面	意味理解・語彙・文法の選定	音声面	意味理解・語彙・文法の選定
	評価基準							
3	・個々の語の発音、アクセント、がほぼ正しい。 ・語の同化・連結もほぼ自然に行われている。 ・若干誤りがあっても、文の内容を十分に伝えることができる読みになっている。	・文を読む速さ、抑揚、ポーズが適切で、内容を理解した読み方になっている。 ・途中、間があっても、不自然さはない。（知らない単語の前で一瞬言いよどむなど） ・最後まで読み切れている。	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	・語彙 語彙やイディオムを正しく使っており、不適切な語はない。 2点 語彙やイディオムの使い方で1ヵ所不適切な使いがある。 1点（1語の場合は1ヶ所が不適切ならば0点） 語彙やイディオムの使い方で2ヵ所以上不適切な使いがある。 0点 ・文法 文で正しく答えている。文法的にミスがない。2点 文には なっていないが文法的ミスはない 1点 文法的にミスがある 0点 ・段階 3 4点 2 3-2点 1 1点 0 0点、または、無音・無回答	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	①イラストの説明が4枚から2枚の場合： イラスト4 枚すべて説明している。 4点 イラスト3 を説明している。 3点 イラスト2 を説明している。 2点 ・語彙 語彙やイディオムを使い方がほぼ正しく、不適切な語は総語数の 5 ％未満。 1点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上10%未満。 2点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上15%以上。 3点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上。 0点 ・文法 文法の使い方がほぼ正しく、文法ミスの箇所は9箇所1箇。 3点 文法の使い方で、不適切な使いかは2箇か3箇。 2点 文法の使い方で、不適切な使いかは4箇か5箇。 1点 文法の使い方で、不適切な使いかが6カ所以上。 0点 ・イラストとイラストの間のつなぎ言葉つなぎ語に不自然なところはない。 2点 つなぎ語に不自然なところがある。または、つなぎ言葉がない。 0点 ①イラストの説明が他の場合： を1枚説明しているが、内容が前の説明になっていない。 採点対象外のため0点 一応説明になっている場合は、3点を上限とし、語彙と文1ヵ所につき－1点、文法ミス1ヵ所につき－1点とする。 段階 3 12～10点 2 9～6点 1 5～1点 0 0点、または、無音・無回答	・自然な聞き方で答えている。 ・個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズが適切である。	・意見の表明・具体例や理由 自分の意見を述べており、その具体例や理由を述べている。 2点 自分の意見を述べているが、その具体例や理由が述べられていない。または、自分の意見ははっきりしないが具体例や理由と思われる文は述べている 1点 文章は述べているが、自分の意見や具体例・理由にあたるものはっきりしない 0点 ・語彙 語彙やイディオムを使い方がほぼ正しく、不適切な語は総語数の 5 ％未満。 3点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上10%未満。 2点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上15%未満。 1点 語彙やイディオムの使い方で、不適切な語は総語数の 5 ％以上。 0点 ・文法 文法の使い方がほぼ正しく、文法ミスの箇所は9箇か1箇。 3点 文法の使い方で、不適切な使いかは2箇か3箇。 2点 文法の使い方で、不適切な使いかは4箇か5箇。 1点 文法の使い方で、不適切な使いかが6カ所以上。 0点 段階 3 8～6点 2 5-3点 1 2～1点 0 0点、または、無音・無回答
2	・個々の語の発音、アクセント、いくつか間違いがある。 ・語の同化・連結は自然とまではいかないが、ポツポツ1語1語を切った読みにはなっていない。 ・いくつか誤りがあったりややぎこちない発音でも、文の内容をおおむね伝えることができる読みになっている。	・文を読む速さ、抑揚、ポーズなどでやや不自然なところがあるが、一応内容を理解できる読み方になっている。 ・最後まで読み切れている。	・やや不自然な聞き方で答えている。 ・自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。		・やや不自然な聞き方で答えている。 ・自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。		・やや不自然な聞き方で答えている。 ・自然な聞き方で答えている。 個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにやや難がある。	
1	・個々の語の発音、アクセントに大きな間違いが多い。 ・語の同化・連結がなされず、1語1語区切った読みになっている。 ・文の内容を伝えることが困難な読みになっている。	・文を読む速さ、抑揚、ポーズなどがかなり不自然であり、内容を理解できていない読みになっている。 ・最後まで読み切れていない。	・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。		・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。		・かなり不自然な聞かれる。 ・自然な聞き方で答えていても、個々の発音、アクセント、同化・連結、読む速さ、抑揚、ポーズにかなり難がある。	
0	無回答・無音	無回答・無音	無回答・無音		無回答・無音		無回答・無音	

１２月にプリ・テストとプリ・アンケートを行い、1 月にオンライン音声収録ソフト（SSES：Soka Speech Evaluation System）を使用したポスト・テストとポスト・アンケート（図 1）を行った。テスト結果を、教師の手動採点と、採点基準を事前情報として取り込んだ ChatGPT で自動採点し、それらの結果を比較した。

図 1　オンライン音声収録画

3．2　AI 活用個別指導・学習の効果

１２月のプレテスト（問題 B のみ）と１月のポストテストの間の約１か月間、１人１台の学習端末を利用した**AI を活用したスピーキング指導を行った。**学校によってばらばらの指導ではなく、統一した指導として STEAC Junior と STEAC というスピーキング訓練プログラム（図 2）を行った。これらは東京大学大学院工学系研究科の峯松研究室から提供を受けた英検 2 級の音声を使ったシャドーイングやオーバーラッピングなどを行うスピーキング訓練プログラムで、STEAC とは同様の形式の東京大学生用に作られたスピーキング訓練プログラムである。

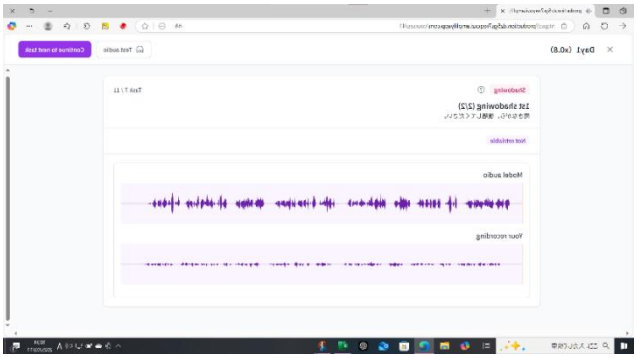

図 2　ICT 活用の訓練プログラム

準備として、全員「PC 内臓マイク」ではなく外付けのマイク付きヘッドセット等を使用すること等を参加の条件とし、録音音声ができるだけクリアに取れるように注意をした。まず、音声録音に慣れるための準備として日本語の録音等を行い、徐々に英語の録音へ移行していった。プログラムはまずは、自分のレベルにあう、英検音声の 0.8 倍速版か 0.9 倍速版かを選び、それぞれ 1 日分として指定されているトレーニングを行う。文字がない状態でのオーバーラッピング、文字がある状態でのオーバーラッピング等、段階的に発音している英文の内容理解とスピーキング能力が向上するように設計されている。

この STEAC junior は中学生・高校生でも行えるレベルのものとして設定されているので、プレテストに参加したものには全員に参加を促した。この訓練を終えたものには、高度なものへのチャレンジとして STEAC への参加を促した。

４．代表的な実践　研究の成果

4.1.1　手動評価と自動評価の関係

学習者が録音した英語音声をベテラン教員が音声面（発音、リズム、イントネーションなど）と内容面（論理的一貫性など）の評価規準（表 1）に基づいて手動評価し、同じ英語音声を自動評価基準（表 2）に基づいて自動評価得点を算出した。手動評価得点と自動評価得点の相関係数を算出した。

ChatGPT に読み込ませた減点方式の自動採点用基準をそのまま使った採点では点数が辛く出すぎるということが分かったため、生徒の言いよどみや重複などを考慮するなどの調整を施し、Q&A の問題やスピーチのような自由発話問題にも対応できる評価アルゴリズムを作成し、自動採点システムを構築することができた。この構築した自動評価システムを利用し、教員の手動評価との相関を取り、自働評価の妥当性を測った。以下がその結果である。

4.1.2　音読評価の分析結果（Part A の音読問題）

Part A の音読問題では、以下のような自動評価方法を実施した。音読問題の評価指標は大き

く「発音差異」「韻律差異」「長さ制御」の３つに分かれている。さらに、「発音差異」は以下の3つの下位基準に分けられる。① 個々の語の発音とアクセントがほぼ正しい　② 語の同化・連結がほぼ自然に行われている　③ 若干誤りがあっても、文の内容を十分に伝えることができる読みになっている。「韻律差異」の基準は、文を読む速さ・抑揚・ポーズが適切で、内容を理解した読みになっているかである。「長さ制御」の基準は「時間内で読み切ったか」である。今回の自動評価では、比較的大きな差が確認された「発音差異」と「韻律差異」を評価指標として採用した。一方、手動評価では、ベテラン教員３人の教員が各自で出した評価点を、手動採点基準を再度確認しながら、どのような音声だったか聞き直し、議論の末に決定値を出した。この手動評価による決定値と自動評価得点とのスピアマンの順位相関関数を算出したところ、有意で比較的高い相関係数が観測された($\rho = 0.68$, $p = .011$, 95% CI [.201, .894])。この結果から音読問題については、教員の手動評価の約 49%を生成 AI による自動評価で説明することができるため、手動評価する前段階として自動評価を活用するような利用法が可能であると考えられる。

4.1.3　条件つき自由発話問題の分析結果　（Part B～D の問題）

Part B～D の問題は Part A の音読問題とは種類が違い、いろいろな解答が想定される自由発話問題に適した採点方法として、評価基準（表 2）を事前情報としてプロンプトに組み込んで ChatGPT に採点させた。なお、採点法を決定するにあたり、以下の手順を取った。

1) 生徒の解答音声データを音声認識アプリを利用して書き起こした。

2) 自動採点を実施するための方法を ChatGPT を用いて 4 種類用意した。

3) 2)で用意した 4 種類の ChatGPT のうち、どれが最も精度が高いかをサンプル音声を利用して検討した。用意した 4 つは以下の 4 種類である（表 3)。

表 3　学習者音声の採点方法

	評価方法	意図推測
方法①	3 段階評価（伝わる/何となく伝わる/単語の羅列）	なし
方法②	10 点満点（発音・流暢さ:2, 文法・語彙:4, 内容の適切さ:4）	あり（自動音声認識の誤認識を補正し、自然な英語に修正）
方法③	10 点満点（発音・流暢さ:3, 文法・語彙:3.5, 内容の適切さ:3.5）	あり（自動音声認識の誤認識を補正し、自然な英語に修正）
方法④	10 点満点（文法・語彙:6, 内容の適切さ:4）	あり（自動音声認識の誤認識を補正し、自然な英語に修正）

スピーキング・テストの採点を ChatGPT で実施するにあたり、難しかったことは、非英語母語話者である学習者の音声を対象に、音声認識機能を使ってどのように捉えるかというところにあった。具体的には、言い直しや繰り返し等が書き起こし結果に含まれると、採点結果に影響を大きな影響を与えてしまう。そのため、学習者が言おうと意図していたことを推測させてサンプル音声を使った方法を含む方法②～④のうち、検証の結果、方法②の採点精度が最もよかったため、この評価アルゴリズムを採用し、ChatGPT で自動採点を行った。そして、

ベテラン教員３人の教員が議論して出した手動評価の決定値得点と ChatGPT の自動評価得点との相関係数を算出した（表 4)。

表 4　条件つき自由発話問題での手動結果と自動評価結果の相関係数

	教員 A	教員 B	教員 C	決定値
PartB1　（SDGs のうち重要と思うものとその理由を述べる）	0.65	0.83	0.66	0.66
PartB2　（学習者の好きな科目とその理由を述べる）	0.52	0.44	0.76	0.50
PartB3　（学習者の所属クラブ活動と放課後の過ごし方を述べる）	0.37	0.53	0.42	0.36
PartB4　（日本語で聞こえる内容を英訳して話す）	0.24	0.18	0.19	0.12
PartC　（4 コマ･イラストの内容を描写･説明する）	0.60	0.17	0.68	0.77
PartD　（学校の環境に関する学習者の意見を述べる）	0.74	0.52	0.74	0.74

以上から問題の性質によっても採点精度・教員の手動評価との相関係数は変わる傾向が伺える。具体的には、B1, C, D などの一般的な事柄について意見を述べる問題はほとんど 0.6 から 0.8 の数値で高い相関関係が観測された。次いで B2, B3 のような個人的なことについて答える問題はほとんど 0.3 から 0.7 という幅の広い相関係数が観測された。B4 のような日本語で指定された内容を英語にして話す問題は 0.1 から 0.2 の数値で低い相関係数しか得られなかったこれらの結果から、教員の手動評価を生成 AI による自動評価に代替させられる可能性はあると考えられる。しかし、問題の種類には一考が必要であろう。一般的な事象に対して自分の意見を答えさせる問題や自分の個人的な情報を述べる問題は、自動評価に代替できる可能生が高い一方、B4 のような英訳の問題は非常に相関が低く、自動評価にそぐはないと考えられる。この理由として３点が考えられる。①話す分量が少ないため、ちょっとしたミス（音声認識の誤認識含む）が点数に大きな影響を与えてしまう。②「話すことが決まっている」問題であるため言い直しが多く（「高校を卒業したら」と言ったが「学校を卒業したら」と言いなおしてしまったなど)、文章のつながりや流暢さの観点から必要以上に減点されてしまう。③uh などのフィラーの多い音声データや発音に癖のあるデータは書き起こし結果の時点で生徒が実際話したかった音声とは、文自体が短いために、大きく乖離してしまう。

以上の結果から ChatGPT にあらかじめ採点基準を事前情報としてプロンプトに組み込んで評価させること、問題とその解答例について、色々な種類や多くの数を用意すること、採点時に教員がある程度甘めに見ることが多いと予想される判断基準（生徒が同じ単語を繰り返し言ってしまう、すぐに正しい単語に言い直す、多少の言いよどみがある、等）については ChatGPT に事前情報として与えておく、などの準備が必要である。今回の調査結果から判断すると、生成 AI を使用した自動採点は教員の手動採点の代替となる可能性を秘めていると思われる。

4.2 ICT 活用個別指導・学習の効果

ICT を活用した学習指導の効果について、アンケート調査の数値の平均値を使って検証を試みた。期間は 1 か月ほどしかなかったが、この ICT を活用したスピーキング訓練に参加した生徒とあまり参加しなかった生徒でプレとポストのテストの得点率の変化が大きく異なった。これは ICT 活用の個別学習の効果を示す差といえる。(表 5)

表 5　ICT 活用スピーキング学習の前後で得点率が変化した学習者の割合

スピーキング訓練プログラムへの参加度	得点率アップ	得点率変化なし	得点率ダウン
参加して STEAC junior を完了した生徒	75%	12.5%	12.5%
STEAC junior を完了しなかった生徒	14.3%	0%	85.7%

ICT を活用した個別学習である STEAC Junior 参加者の 46.7%が、個々の参加者のスピーキング力の伸長をスピーキング・テストの得点率で可視化したとき、以前よりも 10%以上の伸びを示した。さらに詳しく参加率によるテスト得点率の伸びの違いは以下のようになった(表 6)。

表 6　ICT 活用スピーキング学習の前後での学習者の得点率変化

	プレテスト時	ポストテスト時
スピーキング訓練プログラム特に熱心に参加した（STEAC にも参加）	86.1% （B のみ 86.1%）	87.3% （B のみ 86.1%）
スピーキング訓練プログラムに参加した（STEAC Junior は完了した）	83.3% （B のみ 83.3%）	93.3% （B のみ 91.7%）
スピーキング訓練プログラムにあまり参加しなかった（STEAC Junior が未完了）	76.2% （B のみ 76.2%）	71.4% （B のみ 63.1%）

普段から英語の学習に前向きな上位の学習者（スピーキング訓練プログラムに特に熱心に参加した生徒）でも伸びがあった。最も伸びたのは、普段それほど英語の学習に特に熱心というわけではない中間層の生徒（STEAC Junior は完了した生徒）であった。この理由は、普段あまり英語のスピーキングをやっていない生徒が、しっかりとスピーキングの訓練に参加する機会ができたために最も大きな伸びを示したと考えられる。スピーキング訓練プログラムにあまり参加しなかった学習者は予想通り成績が落ちた。学校の授業・通常の宿題以外でも AI や ICT を利用したスピーキング訓練を利用することで、生徒のスピーキング能力を伸ばせる可能性が今回の調査で明らかになったと考えられる。

また、ポスト・テスト後の生徒のコメントでは以下のようなものがあった（表 7)。

表 7　ICT 活用スピーキング学習に対する参加者からのコメント

番号	生徒のコメント
1	何度英文の通りに発音しても、正しい音声認識が反映されない部分があって、そのフレーズは自分にとって苦手な発音だと初めて知ることができた。
2	高得点を目指して PDCA サイクルを自分一人で完結できるところがいいと思った。
3	自分ではきれいに発音したつもりでも、他の言葉として認識されてしまったり、声

	の抑揚をつけることがとても難しかった。
4	その場で音読評価をしてくれるので、改善しやすかった。
5	聞いた音をそのまま発するのが思っていたよりも難しいことが分かった。また、自分がイメージして読み上げた際のイントネーションとサンプル音声には乖離があった。シャドーイングやリピーティング、オーバーラッピングなどの様々なスピーキング勉強法に触れる事ができたのは今後の勉強に活かせると思った。
6	波形などを見てどこの抑揚や発音が違うのかが分かりやすく、修正につながりました。とても良かったです。

以上のコメントから、生徒にスピーキング訓練への意識づけにはある程度成功したといえる。

4.3 英語学習とスピーキング・テストに関するアンケート調査

回答は 4 尺度法で実施した。【1．そう思う 2．まあそう思う　3．あまりそう思わない 4．そう思わない】

表 8　スピーキング学習に対する参加者からのコメント

問	プレテスト時	ポストテスト時
「授業内でのスピーチやディスカッションなどのスピーキング活動は好きですか。」	2.0	2.2
「授業内にネイティブスピーカーや英語の先生と英語で話すことは好きですか。」	2.3	2.3

授業中英語で話すことに対する興味･関心や苦手･不安について特に変化は読み取れなかった。

表 9　スピーキング・テストに対する参加者からのコメント

問「このようなスピーキング・テストを将来また受けたいと思いますか」	プレテスト時	ポストテスト時
スピーキング訓練プログラム参加者	2.1	1.8
スピーキング訓練プログラム不参加者	2.0	2.0

本プロジェクト参加者の 73%以上が、家庭で個別スピーキング訓練プログラムを行い、今後もスピーキングに取り組みたいという意欲の向上が見られた。

表 10 スピーキング学習方法重要性の評価

・問「今後スピーキング力を向上させるためには以下の学習方法は重要だと思いますか。」

方法	A	B	方法	A	B
1 音読	1.2	1.6	6 グループでの意見交換	1.4	1.8
2 リピーティング	1.4	1.9	7 オンライン会話	1.7	2.1
3 オーバーラッピング	1.5	2.0	8 生成 AI との対話練習	2.7	2.6
4 シャドーイング	1.4	2.1	9 英語で独り言	2.0	2.1
5 ペア練習	1.2	1.6	10 英語プレゼンテーション	1.5	1.9

「授業内のスピーチやディスカッションなどのスピーキング活動は好きですか」という問いと「授業内にネイティブスピーカーや英語の先生と英語で話すことは好きですか」という問いに

対して肯定的に１．２．と答えたグループ A と、否定的に３．４と答えたグループ B に分けて、スピーキングのための学習方法に対する意識を調査した。

予想通り A グループの方がどの学習法も概して高く評価しているが、おおむね両グループとも肯定的にみており、両グループともに 2 点台をつけたのは、普段そうした機会に恵まれていないことが影響したとみられる 8「生成 AI との対話練習」と 9「英語で独り言」のみであった。

5. 今後の課題・展望

今回、いくつか課題が残った。プリ・テストとポスト・テストの間の期間があまり長く取れなかったため、1 か月を超える長期の変化を見るスケジュールを取れるようにすること、今回はプレとポストの間の指導は個別のスピーキング練習にとどまってしまったので、学校でのスピーキング活動に工夫を加えた場合どうなるかの検証、といったことは今後明らかにしていきたい。

別の課題としては、自動評価のアルゴリズムを作成するのに膨大な時間がかかってしまったことや、サンプルとする音声を大量に集めることが難しかったことである。しかし、どのような音声を、どの程度の量、どのように集めなければならないかという知見は得られた。今回作成したアルゴリズムを発展的に応用することで、今後のスピーキング・テストの採点がスムーズに行えるようになる。また、自動評価に適した、スピーキング能力を反映しやすい問題形式をあらかじめ意識したうえで、スピーキング・テストの作成を今回作成したテストフレームが柔軟に使うことで、全国の英語教師のスピーキング・テスト作成が容易になるようであれば、この研究は意味があったことになる。今後もこの研究をさらに進めていきたいと考えている。

6. おわりに

本研究は、英語教育学・教育工学の専門家である創価大学の山内豊教授、そして生成 AI や音声情報工学を専門とする東京大学の峯松信明教授のご尽力なくしては成し得ませんでした。お二方には深く感謝申し上げます。また、自動評価得点と手動評価得点の相関を統計的に比較するために協力してくださった東京大学大学院の相場真由子さん、藤原朱里さん、オンライン音声収録ソフト作成に携わっていただいた創価大学大学院生の皆様にも謝辞を申し上げます。さらに、多忙な学校業務の合間を縫い、週末も含め頻繁に会議に参加し、議論を重ねてくださった全英連研究部会の先生方に感謝の意を表します。最後に、音声録音、スピーキング・テストの実施、英語スピーキング訓練等に快く参加してくださった生徒たちに心より感謝申し上げます。

7. 参考文献

Aiba, M. et al. (2024). A ChatGPT-based oral Q&A practice system for first-time student participants in international conferences. *Proceedings of Interspeech*, 5202-5203.

藤原朱里・他（2025）.「世界諸英語話者間の相互シャドーイングに基づく聴取崩れに対する要因分析」『日本音響学会講演論文集』821-824.

Yamauchi, Y. & Nishikawa, M. (2023). Which type of speaking test predicts L2 overall proficiency most? *Proceedings of Asia TEFL 2023*, 453-462.