

## 最適CALL教材割り当てのためのリスニング教材難易度の実証分析

与那覇 信恵・竹蓋 順子

千葉大学大学院国際学術研究院

### Optimizing Listening Material Assignments in CALL: Evaluating Listening Material Difficulty through Empirical Data

YONAHA Nobue and TAKEFUTA Junko

#### Abstract

This study investigates the difficulty levels of CALL (Computer-assisted Language Learning) materials developed and used at Chiba University through learner subjective evaluations and pre- and post- proficiency test results. Specifically, the study evaluates material difficulty levels based on learners' listening score ranges, seeking an assignment method that maximizes learning effectiveness. Results show that with beginner to intermediate-level materials, learning effectiveness significantly decreases when materials are not matched with learners' English proficiency levels. In contrast, upper-intermediate to advanced materials demonstrate stable learning effectiveness, validating the current assignment method for these levels. Additionally, it was found that learners' subjective evaluations of difficulty do not necessarily align with the proficiency gains measured by test scores. This study was unable to determine the specific difficulty level of each material. We aim to develop a more precise method for selecting materials by combining this approach with an analysis of the materials' listenability in future research.

#### *Keywords*

CALL materials, difficulty level, student assessment, TOEIC

## 1. はじめに

### 1.1 千葉大学のCALLと教材選定

千葉大学の英語カリキュラムの特色の1つとして、学内で独自開発された英語CALL (Computer-assisted Language Learning) 教材の存在がある。1994年度から使用されているこのCALL教材シリーズは、2024年現在26タイトルが使用可能となっており、普遍英語科目のCALL (1年次選択必修)、CALL 2 (2年次選択必修)、CALL 3 (選択科目)、国際教養学部の国際教養CALL英語、大学院共通科目Academic Listeningで採用され、年間約1,200名の学生に使用されている。これらの科目の中でもっとも受講生が多いのが1年次選択必修科目のCALL<sup>1</sup>である。この科目は2ターム、週2回、2単位で、全学部の学生が選択できるように年間12クラス(24コマ)設置されている。

通常の英語科目では、クラス毎に同じ教科書を使用し同じペースで学習を進める必要があるが、個別に学習を進めることができるCALLでは学生毎に別の教材を使用することが可能である。そこで、CALLでは、初回授業で実力テストを実施し、その結果に基づいて適切だと思われる聴解力養成用教材を割り当ててきた。ただし、教材を割り当てる際に必要なそれぞれの教材の難易度レベルは、CALL教材開発と指導を長年実施してきた英語教員の主観で決定してきており、割り当てられた教材が実際に最適であったかどうかの検証はこれまで行われていなかった。

学習の効果を最大限にするためには、それぞれの学習者が自身の英語習熟度レベルにあった教材を使うことは重要である。たとえば、竹蓋と水光(2005)は、土肥他(2001)を引用して教材の難易度レベルが学習者の習熟度レベルと合わない場合、効果が約半分になると指摘している。一方、英語教材の難易度と学習者の英語習熟度レベルを合わせる方法として確立したものはないため、教材割当の精度を上げるためには、データに基づいた教材難易度の決定方法を探る必要がある。そこで、本研究では千葉大学に蓄積されてきた約20年分の学習者データを用いて、教材の難易度検証を試みることにした。

### 1.2 研究の目的

本研究の目的は、CALL教材を使用した本学1年次学生の、使用教材、教材の難易度に関する主観的評価データ、および実力テストのデータから、教材の難易度を実証的に検証することである。具体的には、学習前の英語習熟度レベルと教材の難易度の組み合わせにより、実力テストの伸びに差が出るのか、教材難易度に対する学習者の主観的評価と実力テストの伸びにはどのような関係があるのか、について明らかにすることを目指す。

## 2. 先行研究の調査

### 2.1 CALL教材の効果

世界に15億人いるとも推定されている世界の英語学習者 (Knagg, 2019) の様々なニーズに合わせて多くの英語教材が開発されているが、教材の評価に関する報告は驚くほど少ないと言われる (Chapelle, 2010)。そのような状況で、本学で開発されたCALL教材を使った指導の効果は様々な方向から検証されている。たとえば、竹蓋他 (2004) は、教材を使った指導をした群は教材を使っていない群と比較して、TOEICスコアの年間上昇量が約3倍であったと報告している。与那覇他 (2015) は、CALL教材による学習時間が増えるとTOEICスコアの上昇量が増える傾向があり、35時間以上学習した群では年間100点以上の上昇がみられたことを報告している。さらに、与那覇と竹蓋 (2013) では、教材を長期にわたって使用した学習者が実用レベルまで英語力を高めることに成功したことが報告されている。

### 2.2 英語リスニング教材の難易度測定方法

教材の難易度レベルを決める方法としては、教材そのものを観察し分析するアプローチと、その教材の使用者から得られるデータを分析する方法が考えられるが、どちらにも利点と欠点がある。

前者、つまり英語リスニング教材の難易度を教材自体の観察により決定しようとする試みとして、音声言語素材の聞きやすさ (listenability) に関する研究がある。聞きやすさに影響しうる要素としては、言語素材のトピックや内容、発話速度、発音の明瞭さ、訛りの有無、語彙、文構造の複雑さなどが挙げられることが多い (Buck, 2001; Glenn et al., 1995; Saito et al., 2023; ほか)。また、素材だけではなく、教材に組み込まれている学習作業とそれを補助する情報も難易度に影響する。しかし、listenabilityを測定する方法や、学習作業の影響を測定する方法として確立されたものはない。また、素材の難易度には、訛りなど数値化しにくい要素があるだけではなく、教材のトピックと学習者の興味や既に持っている知識との適合性など一般化しにくい要素もあり、教材の分析だけで難易度を決定することは実際には難しい。

後者、つまり実際に教材を使った学習者から得られるデータを分析する方法として、学習者による教材の主観的評価データや、学習者が受けたテストの結果の活用が考えられる。Ellis (2011) は教材の使用後に得られるデータにより教材を評価することの重要性を指摘している。学習者による評価に関して、McGrath (2016) は、「学習者が使っている教材についての学習者の意見を真剣に受け止める必要がある」と指摘している。しかし、学習者による主観的評価は客観的な測定基準と一致しないことも多いという指摘 (North & Schneider, 1998) もあり、扱いには注意が必要である。

テストに関しては、学習者が教材で学んだ内容を習得しているかどうかを評価する到達

度テストと、テスト受験者のその言語での能力を測定する熟達度テストが考えられる (Brown & Abeywickrama, 2010)。本研究の学習者は異なる教材を使っているため、教材の理解度を確認する到達度テストを使って教材の難易度を測るのは、同じ基準で比較できないため難しい。そこで、熟達度テストの利用が考えられるが、その結果には学習者の英語学習への動機づけの違い、教室外での英語学習量の違い、教師の指導方法の違いなど様々な要因に影響を受けている (McGrath, 2016) ため、教材の影響のみを取り出すのは難しいという欠点がある。

### 2.3 TOEICスコアと英語習熟度レベル

標準化された言語テストの多くは、異なるテスト結果を国際的に共通の尺度で理解できるようにするために、スコアをCEFR (ヨーロッパ言語共通参照枠) に関連付けて説明している。TOEICもセクションごとのスコアが意味するCEFRレベルの検証を行っており、Listening Sectionの場合、60点までがA1レベル、その上110点までがA2レベル、275点までがB1レベル、400点までがB2レベル、490点以上がC1レベルとされている (Tannenbaum & Wylie, 2008; 2013)。このことから、TOEICスコアは学習者の英語力レベルの変化を観察するためのデータの1つとして利用できると言える。そこで、教材が適切な難易度であれば、学習者の能力向上に効果的に働いてスコア向上に影響するという仮定のもと、本研究では、使用教材別の学習前後のTOEIC形式の実力テストスコアの変化を観察することにした。

## 3. 研究方法と結果

### 3.1 使用データ

CALLでは、初回授業時に教材割当てのための実力テスト、最終授業で応用力の変化を観察するための実力テストを実施してきた。最終授業で実施するテスト結果は成績の一部とする。また、1タイトルの教材学習が終了する毎に学習者アンケートを実施し、教材に対する主観的評価データを収集してきた。実力テストとアンケート実施に関わるCALLのスケジュールは図1のようになる。

本研究開始時点で、過去にCALL科目を履修した学生の使用教材、小テスト、実力テスト、アンケートのデータは約20年間分が蓄積されていた。そこで、これらのデータを使い、現

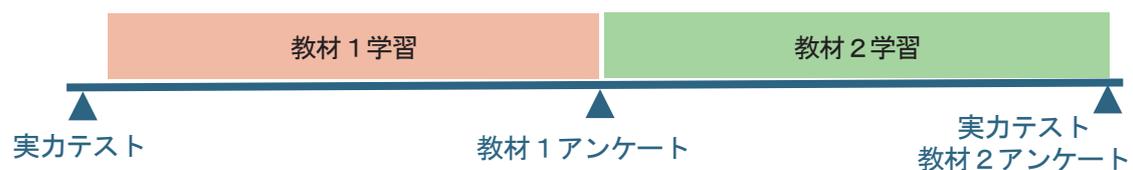


図1 CALL教材を使った1年次英語科目 (15週30回授業) のスケジュール

在使用されている聴解力養成用CALL教材の難易度を検証し、より適した教材割当を可能にすることを旨とするにしている。

分析に使用したのは、2003年度から2023年度までに、本学でCALL教材を使った週2回15週の普遍英語科目を履修した1年次学生の下の3種のデータである。

- ① 個々の学習者のListen to Me! シリーズの聴解力養成用CALL教材のタイトルデータ
- ② 学習者による教材の主観的評価データ（アンケート）のうち、教材難易度に関する質問項目への回答結果
- ③ CALL授業内で実施した学習前、学習後の英語実力テストスコア

①に関して、学習者は図1に示したように、週2回2単位の科目では前半と後半に異なる聴解力養成用教材を使用した。つまり、それぞれの学習者が使用した教材は2種類あったということである。また、2つ目に使用する教材は1つ目に使用する教材よりも難易度が高いものであった。たとえば、学期前半に初中級レベルの教材を使用した学習者は、後半には中級レベルの教材を使用した。本研究で扱ったデータに含まれていた聴解力養成用教材のタイトルとその略称を表1に示す。表中の「レベル」は教材開発者が判断したもので、5つのレベルに分けられていた。なお、本稿では以降教材名を略称で示す。

表1 聴解力養成用教材のタイトルと略称

レベル	教材タイトル	略称	レベル	教材タイトル	略称
初級	First Listening	FL	中上級	College Life	CL
	Doorway to the UK	UK		Gateway to Australia	AU
初中級	New Yoke Live	NY		Horticulture in Australia	HA
	American Daily Life	DL		English for Nursing Science	NS
中級	People at Work	PW	上級	A Bit of Britain	BB
	Canadian Ways	CW		College Life II	CT
	Introduction to College Life	IC		News from the World	NW
			World Health Issues	WH	

②に関して、CALL教材を使った科目では、教材が1タイトル終わるごとに、約30項目を5段階のリッカート尺度で評価する「学習者アンケート」を実施してきた。このアンケートは1994年度当初から実施されてきたようであるが、どの回答がどの学習者によるものかわかる形で結果が保存されていたのは2010年度以降のデータであった。そこで、本研究では2010年度以降の教材の難易度に関する質問項目への回答を分析対象とした。

③に関して、授業開始時と終了時に実施した、Listening Section 100問、Reading Section 100問からなる実力テストの結果は、2003年度から2019年度に実施したものが保存され

ていた。この実力テストの結果は、Listening 495点満点、Reading 495点満点、Total 990点満点のTOEICスコアとして記録されている。なお、2020年度はCOVID-19の流行による影響で対面授業が実施できなかったため実力テストは実施しておらず、2021～2023年度は実力テストとしてListeningのみを実施したため、本研究では分析対象からはずした。

### 3.2 データ分析方法と結果

#### 3.2.1. 主観による難易度の評価 1

教材の難易度に関するアンケート項目は、アンケート実施年度によって質問の方法が異なっていた。具体的には、2010～2011年度に実施されたアンケートでは、「難易度は易しすぎた」「難易度は難しすぎた」の2項目に分かれていたが、2012年度以降は「難易度は適切であった」という項目に変更されていた。結果はどちらの質問においても、5段階のリッカート尺度で、最も肯定的な回答が「1」否定的な回答が「5」として保存されていた。

まず、2010年度と2011年度2年間分のデータをまとめた結果を表2に示す。表2の「教材1」は学期前半に使用した教材、「教材2」は学期後半で教材1の後に使用した教材である。表中の「易し過ぎ」列は「難易度は易しすぎた」という項目に肯定を表す1または2と回答した割合、「難し過ぎ」列は「難易度は難しすぎた」という項目に肯定を表す1または2と回答した割合である。

表2 2010～2011年度実施アンケートにおける学習者による教材難易度評価

教材レベル	教材	教材 1			教材 2		
		<i>n</i>	易し過ぎ	難し過ぎ	<i>n</i>	易し過ぎ	難し過ぎ
初級	FL	103	33%	4%	0	—	—
初中級	NY	93	16%	12%	29	3%	17%
	DL	311	10%	12%	74	8%	22%
中級	PW	53	8%	21%	93	2%	48%
	CW	101	8%	22%	311	3%	49%
中上級	CL	147	11%	13%	154	6%	36%
	AU	0	—	—	62	6%	11%
上級	BB	0	—	—	85	0%	45%

まず、学期前半で使用した教材1の結果を確認すると、「易し過ぎ」と回答した割合と「難し過ぎ」と回答した割合を足した結果が22%～37%であり、大部分の学習者は難易度が適切であると感じていたことが推定される。一方、学期後半で使用した教材2については「難し過ぎ」という回答が多い教材がある。

また、表2全体を概観すると、「易し過ぎ」と評価した割合は全体的に少なく、教材間のばらつきも比較的小さいのに対して、「難し過ぎ」と評価した割合は教材間でかなり差があることに加え、教材1と教材2の違いも大きいことがわかる。特にPW、CW、CLにおいて、教材1で「難し過ぎ」と回答した割合は8%、8%、11%なのに対して、教材2で「難しい」と評価した学習者48%、49%、36%と3倍から6倍と顕著に多い。教材2は教材1よりも難易度が高い教材が割り当てられるため、学習者の英語習熟度レベルと教材レベルが合わないことが、この傾向の主な原因であるとも考えられるが、それ以外の原因もあり得るのかを探ってみることにした。

そこでまず、教材1と教材2両方で使われている教材を使った学習者のデータを、使用教材別、学習前の実力テストListeningスコア帯別に分けた。その中から、20名以上のデータがあった群のみ取り出した。その結果を表3に示す。表中のPre-Lは学習前の実力テストのListeningスコア帯を表す。

表3 使用教材と学習前のスコア帯が同じ場合の難易度評価（肯定的回答の割合）

		教材 1			教材 2		
Pre-L	教材	<i>n</i>	易し過ぎ	難し過ぎ	<i>n</i>	易し過ぎ	難し過ぎ
100-145	DL	26	0%	15%	29	10%	28%
150-195	DL	84	5%	13%	28	4%	21%
200-245	PW	19	11%	21%	54	2%	43%
200-245	CW	41	12%	24%	163	1%	49%
250-295	CW	55	5%	20%	34	6%	38%
250-295	CL	51	6%	16%	82	5%	34%
<i>Mean</i>			6%	18%		5%	36%

「易し過ぎる」と感じた割合を教材1と教材2で比較しても明確な差は見られない。しかし、「難し過ぎる」と感じた割合はすべての群で「教材2」の方が高く、平均で2倍の差がある。表3で抽出したデータの、教材1と教材2の「難易度は難し過ぎた」という項目への回答結果にはt検定で有意差も確認された ( $t=8.63, p<.001$ )。この結果から、学習前のListeningスコアに差がない場合でも2つ目の教材を「難しい」と感じる傾向があることがわかった。前述したように、教材2は教材1よりも上のレベルの教材としているため、1つ目の教材と比較して難易度を判断しているためであると考えられる。この結果により、教材2は「難し過ぎる」と感じる学習者が多いことを考慮して、アンケート結果を観察する必要があることがわかった。

### 3. 2. 2. 主観による難易度の評価 2

2012年度～2023年度の11年間のアンケートにおける難易度に関する質問項目は、「難易度は適切であった」であった。この間のアンケート回答数は4,136件であった。このデータを、学期前半で使用した教材と学期後半で使用した教材別に、学習前の実力テストListeningスコア（Pre-L）帯別に分けて、肯定的回答（5段階で1または2）の割合を出した。さらに、その中から20件以上のデータがあるもののみを取り出した。その結果を表4に示す。

表4中の「教1」はその教材を学期前半で使用した場合、「教2」は学期後半で使用した場合の評価である。表は肯定的回答の割合が高いほど濃い緑に近く、低いほど黄色に近い色で示している。

表4 教材・学習前のListeningスコア帯別「難易度が適切」と回答した割合

Pre-L	初級		初中級		中級		中上級		上級		
	UK 教1	NY 教1	NY 教2	PW 教1	PW 教2	CL 教1	CL 教2	NS 教1	NS 教2	NW 教2	WH 教2
100-145	77%	82%	76%		56%						
150-195	80%	88%	85%	64%	65%		75%				
200-245	76%	88%	88%	81%	67%		63%		68%	57%	
250-295		84%		86%	70%	82%	69%	94%	72%	58%	72%
300-345				87%		86%	69%	92%		64%	84%
350-395						82%		83%		74%	77%
400-						71%		91%		94%	96%

この表からいくつかの特徴が読み取れる。1つ目は初中級のNYについて、学習前のListeningスコアが100から295と幅広いスコア帯の学習者の多くが「難易度が適切」であると評価しており、学習前のスコアの違いによる難易度評価があまり変わらないことである。次に、中級のPWについては、プリスコアが200未満で肯定的割合が低い。PWを教材2として使った多くの学習者は1つ目の教材としてNYを使用していたため、この2教材の難易度の差が大きい可能性がある。CLは学習前のスコアが250以上の学習者が「適切」であると感じる割合が多く、またPWと同様、教材2で使用された場合は評価が低くなる傾向がみられた。

NSは250以上の学習者が「適切」であると感じる割合が高い。CLよりも肯定的回答の割合が大きいが、これはこの教材の内容が看護に関することであり、この教材の使用者が医学部、看護学、薬学部の学生であったため、学生の専門と教材の内容に関連があったこ

とが影響している可能性がある。

上級レベルのNWとWHはどちらもニュースを素材とした教材であるためか、難易度が「適切」と感じる割合が8割を超えるのは学習前のスコアが400点以上の英語習熟度レベルが高い群であった。このことから、NWとWHの難易度は他の教材と比べてかなり高い可能性がある。

以上の学習者の教材難易度に対する主観的評価の結果から、それぞれの教材の難易度を大まかに推定した結果を表5に示す。適切であると回答した割合が90%以上の場合に◎、80%台の場合は○、70%台は△、70%未満は×の印とした。

表5 学習者による難易度評価結果まとめ

Pre-L	初級	初中級	中級	中上級		上級	
	UK	NY	PW	CL	NS	NW	WH
100-145	△	○	×				
150-195	○	○	×				
200-245	△	○	○			×	
250-295		○	○	○	◎	×	△
300-345			○	○	◎	×	○
350-395				○	○	△	△
400-				△	◎	◎	◎

適切であると回答した割合 ◎：90%以上、○：80%台、△：70%台、×：70%未満

しかしながら、学習者による教材の難易度評価は以下の理由で妥当性に欠ける可能性がある。1つ目は、本研究で使用したCALL教材が「三ラウンド・システム」(竹蓋, 1997)に基づいて作られていることにある。この指導理論では、オーセンティックな音声言語素材を使用し、適切なタイミングで提示される様々な補助情報を使いながら、素材を何度も聞きながら行う複数の学習作業を、段階を踏んで行うことで、徐々に聞き取りができるようになる。この方法によると、伝統的な方法では難易度が高すぎる言語素材でも聞き取れるようになるため、学習開始時には3割程度しか理解できないレベルが適切であるという前提で教材を割り当てている。したがって、学習開始時の印象に基づいてアンケートに回答している場合、難易度が適切でないと評価する可能性がある。2つ目は、同じタイトルの教材の中でも素材の難易度がばらついている場合があることである。たとえば、発話速度はリスニングの難易度に影響を与えることが確認されているが (Griffiths, 1990; Fujita, 2017; Jones et al., 2007)、初中級レベルの教材であるAmerican Daily Life (DL) で最も発話速度が速い部分では1分間に324語のスピード、もっとも遅い部分では1分間に119語の

スピードであった。1つの教材の中で発話速度に約3倍の違いがある素材が使われていることになる。つまり、学習者がアンケート回答時に教材のどの部分を思い出したかによって難易度評価が変わってくる可能性がある。3つ目は、特に教材2の難易度評価が教材1の学習結果に影響を受けていることである。前述の通り、学習前のListeningスコア帯が同じでも、教材2の方が難しいと回答する学習者が多かった。

以上のことから、学習者アンケートの結果のみから、それぞれの教材がどのレベルの学習者に最適であるかを決定するのは難しく、別の方向からの検証を組み合わせる必要があると結論した。

### 3.2.3. 使用教材・学習前のスコア帯別実力テストの伸びの比較

次に、2003～2019年度の16年分の実力テストの伸びを、使用教材別に比較することにした。方法は以下の通りである。

1. 学期初めと学期終わりの実力テストの結果が両方揃っているデータ8,454件を抽出した。
2. 教材1と教材2の組み合わせ毎にグループ化した。使用教材の組み合わせは32通りあったが、その中から、現在使用されている教材同士の組み合わせで、100名以上の学習者データのあるもので、また教材2のレベルが教材1よりも1段階高い15通りの組み合わせを分析対象とすることにした。
3. CALLで聴解力養成用教材の割り当てのために使用したのは実力テストのListening SectionとReading Sectionの合計得点であった。しかし、英語聴解力養成用教材のレベルを検証するためにはListening Sectionスコアのみの方がふさわしいと考えられること、また受講者の実力テストTotalスコアとListeningスコアには高い相関が確認された ( $r=0.89$ ) ことから、学習前のListeningスコア (Pre-L) を学習前の英語習熟度レベルを表すスコアとすることにした。
4. 分析対象とした使用教材別の15のグループを、さらに学習前のListeningスコア帯別 (50点刻み) に分け、それぞれの群の実力テストTotalスコアの「伸び」の平均値を出した。そのなかから、データが20件以上ある群を観察対象として抽出した。その結果を表6に示す。

表6のデータを、学習前のListeningスコア帯別の7群にまとめ、学習前 (Pre-Total) と学習後 (Post-Total) のスコアとその差、またそれぞれの群で学習前のスコアと学習後のスコアの検定を行った結果を表7に示す。

学習前と学習後のスコアのt検定の結果、全ての群で有意差 ( $p<.001$ ) が確認された。また効果量 (Cohen's  $d$ ) については、100-145、150-195の2群では効果量大、他の群では中程度の効果量が観察された。この結果から、学習前のスコアに関わらず、実力テストで測定される英語力が上がっていたことがわかる。CALL教材を使った指導が、下のレベルだけではなく伸ばすのが難しいと言われる上級レベル (Tschirner, 2016) の学習者の英語力向上にも成功していたことを示していると言える。

表6 使用教材・学習前Listeningスコア帯別実力テストの伸びの平均値

使用教材	Pre-L						
	100-145	150-195	200-245	250-295	300-350	350-395	400以上
FL&NY	91.5	69.5	60.0				
<i>n</i>	220	308	76				
FL&DL	80.1	79.2	46.9				
<i>n</i>	59	72	31				
UK&NY	102.1	73.4	56.6				
<i>n</i>	68	112	28				
NY&PW	88.2	55.0	46.1	28.9			
<i>n</i>	134	802	1315	325			
NY&IC		55.2	48.5	44.0			
<i>n</i>		119	152	25			
DL&IC		63.3	38.9	32.0			
<i>n</i>		48	117	27			
DL&CW	71.6	53.9	42.0	25.3			
<i>n</i>	28	96	180	37			
PW&CL		60.9	39.8	36.1	15.3		
<i>n</i>		29	408	604	75		
IC&CL		63.1	43.3	39.1			
<i>n</i>		83	287	159			
CW&CL			55.0	36.3			
<i>n</i>			44	62			
CL&CT			58.4	46.6			
<i>n</i>			67	119			
CL&AU				44.2	44.7		
<i>n</i>				76	68		
CL&BB				22.2	29.8		
<i>n</i>				38	51		
CL&NW				32.3	34.4	36.0	33.0
<i>n</i>				175	189	61	20
NS&WH				37.2	41.2	31.5	23.8
<i>n</i>				51	61	41	28

表7 学習前Listeningセクションスコア別実力テストTotalスコア

		Pre-L						
		100-145	150-195	200-245	250-295	300-350	350-395	400以上
Pre-	Mean	282.0	363.1	433.3	503.8	601.5	679.4	772.2
Total	SD	59.0	55.7	51.7	51.0	52.8	46.6	48.3
Post-	Mean	371.6	423.8	478.7	539.0	634.7	713.6	799.8
Total	SD	91.6	80.6	77.2	77.3	79.2	67.5	51.3
上昇スコア		89.6	60.7	445.3	35.2	33.2	34.2	27.6
	<i>n</i>	509	1669	2705	1647	444	102	48
	<i>t</i>	27.78***	38.30***	38.88***	25.37***	12.86***	6.32***	5.26***
	<i>d</i>	1.16	0.88	0.69	0.54	0.49	0.59	0.55

\*\*\*有意差あり  $p < .001$

しかし、表7の「上昇スコア」を見ると、学習前のListeningスコアが高いほど上昇スコアが小さい傾向がみられる。たとえば、学習前スコアが最も低い100-145の群と最も高い400以上の上昇スコアを比較すると、89.6と27.6で約3倍の差がある。これは、学習前のスコアが高ければ高いほど伸びにくくなるという一般的な現象の影響を受けている可能性が高い。そこで、本研究ではその影響を排除した結果の観察を試みることにした。

### 3.2.4. 上昇スコア補正方法と補正後の結果

学習前の習熟度レベルが高ければさらに伸ばすことが難しいことは、多くのスキルで一般的な現象であるが、外国語学習においても当てはまることは直接的、間接的に複数の研究者に指摘されている (Anderson, 2001; Nation, 2001; Rifkin, 2005)。本研究では実力テストとしてTOEIC形式の問題を使用しているため、TOEICに関する指摘に絞ってみると、初回のTOEICスコアの差がスコア上昇に与える影響について、19,855名のデータを分析したWei & Low (2017) は、受験者の初回スコアとスコア伸びにはわずかな負の相関があり、回帰分析の結果Listening Sectionでは-0.26、Reading Sectionでは-0.15であったと報告している。ただし、この研究の調査対象は必ずしも日本人大学生ではないことから、より本研究の学習者に近い受験者のデータを使用する方が望ましい。

本研究で分析対象としている学習者は大学1年生であるため、国立大学学部1年次学生が受験するTOEICスコアにおいて、学習前のスコアの違いがどの程度上昇スコアに影響するかに関する傾向が参考になる。そこでその傾向を観察するために広島大学のデータを使用することにした。広島大学は2003年から2017年まで学内実施のTOEIC-IPの学部・コースごとの平均値を公開している (広島大学外国語教育センター)。2003年度から2015年度入学生は1年次の5月に1回目、同じ年度の11月または2月に2回目のTOEIC-IPを受験しており、その2回とも受験した学生のデータがまとめられていた。初回平均スコアはコースによって大きく異なり、コースの平均値が最も高いものと最も低いものを比較すると

350点程度の開きがあったため、初回のスコアが上昇量に与える影響を観察するのにふさわしいデータであると言える。なお、教育学部第三類（言語文化教育系）英語文化系コースは英語教員を養成するコースであるためか、他のコースとは明らかに異なるスコア変化を見せていたため、今回の分析対象からはずした。

分析対象とした8年分419件のデータから、学習前のスコアが伸びに与える影響を検証するために、従属変数として1回目と2回目Totalスコアの差（伸び）を用い、独立変数として学習前のTOEICスコアを用いた単回帰分析を行った。その結果、学習前のTOEICスコアがスコアの増加量に有意な負の影響を与えることが確認された（ $\beta_1 = -0.091$ 、 $p < .001$ ）。これは、学習前のスコアが100点高くなるごとに、TOEICスコアの増加量が9.1点減少することを示している。切片は $\beta_0 = 46.40$ であった。決定係数は $R^2 = 0.067$ であり、TOEICスコアの増加量の約6.7%が学習前のスコアによって説明されることが示された。この決定係数は高いとは言えないが、スコアの上昇はテスト慣れによる練習効果、学習者の英語学習への動機付けの強さ、英語学習量、テスト受験時の体調ややる気など多くの要因に影響を受けるため、学習前のスコアの影響度がそれほど高くないのは当然である。しかし、学習前のスコアが与える影響は有意なものであり、無視できないこともわかった。

そこで、この結果から得られた回帰直線式を用いて、表5で使用した7,175名が通常得られる「上昇スコア期待値」を出し、実際の上昇スコアと上昇スコア期待値の差を「補正上昇スコア」とした。補正上昇スコアの平均を表6と同じ群ごとにまとめた結果を表8に示す。

#### 4. 考察

これまで示した結果について、全体的な傾向と使用教材ペア毎の回帰分析、および学習者による主観的評価と実力テスト補正上昇スコアについて考察を述べる。

##### 4.1 全体的な傾向

図2は表8を図示したものである。初級と初中級の教材を使ったデータはオレンジ、初中級と中級教材の組み合わせは緑、中級と中上級教材は青、中上級と上級教材は紫で示した。

注目すべきは同じ教材ペアを使用しているも、学習前のListeningスコア帯によって補正上昇量に大きな違いがある場合とあまり違いがない場合があることである。特に、図の左半分、つまり学習前のListeningスコア帯が100点から295点の学習者群で、使用教材が同じであっても学習前のスコアによって大きく上昇量が異なっていることがわかる。学習者の伸びが学習前のListeningスコアによって大きく変わる場合とそうでない場合があるということは、教材を割り当てる際に特に学習者の英語力との関係に注意すべき学習者のレベル、または教材があるということである。そこで、注意が必要な学習者レベルと教材ペアを明らかにするために、使用教材ごとに従属変数として補正上昇スコアを用い、独立

表8 使用教材・学習前Listeningスコア帯別実力テスト補正上昇スコア

教材 レベル	使用教材	Pre-L						
		100-145	150-195	200-245	250-295	300-350	350-395	400以上
初級 & 初中級	FL&NY	68.4	51.2	44.1				
	FL&DL	57.1	60.5	30.5				
	UK&NY	80.1	55.5	40.6				
初中級 & 中級	NY&PW	71.8	43.2	37.8	22.8			
	NY&IC		44.4	39.7	36.8			
	DL&IC		52.4	31.2	25.6			
	DL&CW	56.0	41.4	32.9	19.1			
中級 & 中上級	PW&CL		55.1	38.7	36.2	16.5		
	IC&CL		55.2	38.1	37.2			
	CW&CL			53.7	36.4			
中上級 & 上級	CL&CT			58.3	47.8			
	CL&AU				50.6	54.6		
	CL&BB				28.7	39.9		
	CL&NW				38.2	43.2	50.1	55.9
	NS&WH				44.9	53.7	48.9	48.3

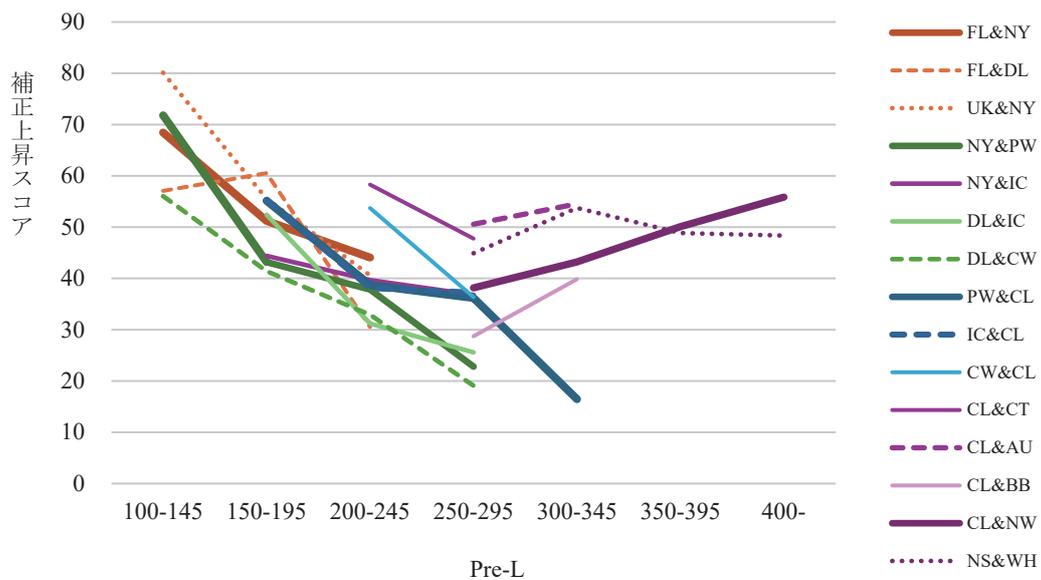


図2 使用教材・学習前Listeningスコア帯別実力テスト補正上昇スコア

表9 学習前のListeningスコアと補正上昇スコアの単回帰分析結果

教材レベル	使用教材	<i>n</i>	<i>R</i> <sup>2</sup>	<i>β</i>	<i>p</i>
初級 & 初中級	FL&NY	654	0.061	-0.465	0.000***
	FL&DL	174	0.072	-0.430	0.000***
	UK&NY	226	0.098	-0.700	0.000***
初中級 & 中級	NY&PW	2,592	0.022	-0.257	0.000***
	NY&IC	301	0.002	-0.079	0.464
	DL&IC	193	0.040	-0.378	0.005**
	DL&CW	345	0.016	-0.214	0.019*
中級 & 中上級	PW&CL	1,121	0.008	-0.171	0.004**
	IC&CL	545	0.006	-0.121	0.076
	CW&CL	111	0.030	-0.314	0.069
中上級 & 上級	CL&CT	206	0.015	-0.202	0.083
	CL&AU	174	0.005	-0.086	0.356
	CL&BB	108	0.002	0.051	0.622
	CL&NW	464	0.000	0.026	0.662
	NS&WH	191	0.000	-0.018	0.774

\**p*<.05、\*\**p*<.01、\*\*\**p*<.001

変数として学習前のTOEIC Listeningスコアを用いた単回帰分析を行った。その結果を表9にまとめる。

決定係数である*R*<sup>2</sup>の値を見ると、全体的に低い値であり、学習前のListeningスコアが補正上昇量に与えている影響は、それほど大きくないことがわかる。3.24で示したように上昇量を補正することで、学習前のスコアが上昇スコアに与える一般的な影響は排除しているために、低い結果となっていると考えられる。しかしながら、1つ目の教材レベルが初級である4群は他の教材と比べて高い傾向にある。もっとも値が大きいUK&NYでは*R*<sup>2</sup>=0.098であり、これは学習前のListeningスコアが補正上昇量に与える影響が9.8%程度であることを示している。次に、傾きである*β*と*p*値を観察する。*β*は多くの群でマイナスである。マイナスである場合、学習前のスコアが上がれば補正上昇スコアが低くなることを意味する。初級と初中級教材であるFL&NY、FL&DL、FL&IC、UK&NYで傾きが大きく、学習前のスコアが上がるごとに上昇スコアが小さくなる傾向が強いことがわかる。さらに、これらの4群ではすべて*p*<.001で有意である。学習前の実力テストが比較的低い100点から245点の学習者に割り当てた教材ペアで*R*<sup>2</sup>の値が大きく、傾きがマイナスで回帰分析の結果が有意であることから、このスコア帯で初級から中級レベルの教材で学習者の英語力とのミスマッチが起りやすいことがわかった。特に初級と初中級の教材を、英語習熟度レベルが高い学習者に割り当てないように注意する必要があると言える。

一方、傾き  $\beta$  がマイナスではあるが数値が小さい群、またマイナスではない群について、その多くは  $p$  値が有意でない。たとえば、1つ目の教材が中級のIC&CL、CW&CLおよび中上級の5つの群にこれが当てはまる。これは、学習前のスコアが補正上昇量に影響を及ぼしていないことを示しており、学習前のListeningスコア250点以上の幅広いの英語レベルの学習者に安定した効果が期待できる教材であると言える。

#### 4.2 学習者による主観的評価と実力テスト補正上昇スコア

最後に、学習者による教材難易度の主観的評価と、実力テストの補正上昇量による学習効果を合わせた結果を観察する。両方の結果がある7種の教材について、表5と表8の結果を合わせたものを表10に示す。学習者による主観的難易度評価は「難易度評価」とし、難易度評価がない部分は「-」とした。

表10 学習者の主観的評価と補正上昇スコア

Pre-L		使用教材				
		UK&NY	NY&PW	PW&CL	CL&NW	NS&WH
100-145	難易度評価	△&○	○&×			
	補正上昇量	80.1	71.8			
150-195	難易度評価	○&○	○&×	×&-		
	補正上昇量	55.5	43.2	55.1		
200-245	難易度評価	△&○	○&○	×&-		
	補正上昇量	40.6	37.8	38.7		
250-295	難易度評価		○&○	○&○	○&×	◎&△
	補正上昇量		22.8	36.2	38.2	44.9
300-345	難易度評価			○&○	○&×	◎&○
	補正上昇量			16.5	43.2	53.7
350-395	難易度評価				○&△	○&△
	補正上昇量				50.1	48.9
400-	難易度評価				△&◎	◎&◎
	補正上昇量				55.9	48.3

表10から、学習者が「難易度が適切だ」と評価した教材と実力テストの伸びの大きさは必ずしも一致していないことがわかる。たとえば、学習前のListeningスコア帯が250から295点の学習者群の80%以上がNYとPW両方の難易度を「適切だ」と回答しているが、実力テスト補正上昇量の平均は22.8点であり、PWの難易度が「適切だ」と回答した割合が

低い100点から195点の補正上昇量71.8と43.2と比べて低い。同様に、Pre-Lが300-345の群の80%以上がPWとCL適切であると評価しているが、伸びは16.5点と小さく、同じ教材を使ったListeningスコア150点から245点の学習者と比べて半分以下である。全体的に、学習者は自身の英語力レベルに対して易し過ぎる教材も「適切である」と評価する傾向があり、「難し過ぎる」と感じている場合も実力テストのスコアが伸びている場合が多い。このことから、本研究で使用されているCALL教材の場合、「難し過ぎる」と感じる教材の方が英語力を伸ばす可能性が高いことを、明示的に学習者に伝える必要があると言える。

## 5. まとめ

本研究では、学習者による教材の難易度の主観的評価と学習者が学習前後に受けた実力テストの結果から、学習者に最適なCALL教材を割り当てるための教材の難易度レベルを検証した。どの教材の組み合わせでも、どのスコア帯でも実力テストで伸びが見られていた。

しかしながら、初級から中級レベルの教材では、学習者の英語力レベルと合っていない場合に顕著に伸びが減少する傾向がみられた。このことから、学習前のListeningスコアが250から345点レベルの学習者への教材割当は特に注意する必要があると推定された。一方、中上級から上級レベルの教材では、これらの教材を割り当てていた学習前のListeningスコア250点以上のスコア帯で安定した上昇を得られていたため、従来の割当方法のままで問題がないと言える。

また、教材難易度に関する学習者の主観的評価と実力テストの伸びは必ずしも一致しておらず、難易度が適切であると学習者が感じているかどうかは、実力テストの伸びにそれほど影響しない可能性があることもわかった。しかしながら、授業担当者は、難し過ぎると感じている学習者がいることを念頭に置いた指導をする必要があると言える。

今回は、TOEIC形式のテストのListening Sectionのスコアを学習前の英語力レベルを表すスコアとして、学習前後のTotal スコアの差を、学習効果を示すスコアとして使用して最適教材について検討した。しかし、個々の教材の難易度について結果を出すことができなかった。結果を出すことが難しかった1つの大きな原因は、学習者が複数の教材を使用していたことにある。本研究では、1年次のCALLで使用された教材の組み合わせで検証したが、今後は2つの教材を組み合わせ使用させること自体を見直した上で、改めて難易度を検討することも必要かもしれない。さらに、教材の英語音声素材の客観的指標と組み合わせることで、各学習者に最適な教材を選定するための手法を確立したい。

## 注

- 1 科目名は2020年度に「CALL英語」から「CALL」に変更されたが、授業で使われる教材や授業実施方法はほぼ変わらないため、本稿では区別せずにCALLとする。

## 謝辞

本研究にあたり、「三ラウンド・システム」の理論開発をされた竹蓋幸生先生、長年にわたりCALL教材の開発と教材を使った指導を実践されてきた高橋秀夫先生、土肥充先生に深く感謝申し上げます。また、CALL補佐員の皆様には、長年にわたるデータ入力作業へのご協力に深く感謝いたします。

## 参考文献

- Anderson, L.W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.
- Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment: Principles and Classroom Practices*. Pearson Education.
- Buck, G. (2001). *Assessing Listening*. Cambridge University Press.
- Chapelle, C. A. (2010). The spread of computer-assisted language learning. *Language Teaching*, 43(1), 66-74. doi:10.1017/S0261444809005850
- Ellis, R. (2011). Macro- and micro-evaluations of task-based teaching. In B. Tomlinson (Ed.), *Materials Development in Language Teaching* (pp. 212-235). Chapter, Cambridge: Cambridge University Press.
- Fujita, R. (2017). Effects of Speech Rate and Background Noise on EFL Learners' Listening Comprehension of Different Types of Materials. *The Journal of AsiaTEFL*, 14, 638-653. 10.18823/asiatefl.2017.14.4.4.638.
- Glenn, E. C., Emmert, P., & Emmert, V. (1995). A Scale for Measuring Listenability: The Factors that Determine Listening Ease and Difficulty. *International Journal of Listening*, 9(1), 44-61. <https://doi.org/10.1080/10904018.1995.10499141>
- Griffiths, R. (1990). Speech Rate and NNS Comprehension: A Preliminary Study in Time-Benefit Analysis. *Language Learning*, 40: 311-336. <https://doi.org/10.1111/j.1467-1770.1990.tb00666.x>
- Jones, C. & Berry, L. & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech & Language*, 21, 641-651. <https://doi.org/10.1016/j.csl.2007.03.001>
- Knagg, J. (2019). *The English Effect*. British Council. [www.britishcouncil.org](http://www.britishcouncil.org).
- McGrath, I. (2013). *Teaching Materials and the Role of EFL/ESL Teachers: Practice and Theory*. London: Bloomsbury.
- McGrath, I. (2016). *Materials Evaluation and Design for Language Teaching*. Edinburgh University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge university press.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15(2), 217-263. <https://doi.org/10.1191/026553298676177132>
- Rifkin, B. (2005). A Ceiling Effect in Traditional Classroom Foreign Language Instruction: Data from Russian. *The Modern Language Journal*, 89(1), 3-18. <http://www.jstor.org/stable/3588548>
- Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (2023). Individual differences in L2 listening proficiency revisited: Roles of form, meaning, and use aspects of phonological vocabulary

- knowledge. *Studies in Second Language Acquisition*, 1-27. doi:10.1017/S027226312300044X
- Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. <https://www.ets.org/Media/Research/pdf/RR-08-34.pdf>
- Tannenbaum, R. J., & Wylie, E. C. (2013). Mapping TOEIC® and TOEIC Bridge™ test scores to the Common European Framework of Reference. Educational Testing Service. <https://www.ets.org/pdfs/toeic/toeic-mapping-cefr-reference.pdf>
- Tschirner, E. (2016), Listening and Reading Proficiency Levels of College Students. *Foreign Language Annals*, 49: 201-223. <https://doi.org/10.1111/flan.12198>
- Wei, Y. & Low, A. (2017), Monitoring Score Change Patterns to Support TOEIC® Listening and Reading Test Quality. *ETS Research Report Series*, 2017: 1-20. <https://doi.org/10.1002/ets2.12186>
- 竹蓋幸生 (1997). 『英語教育の科学』株式会社アルク.
- 竹蓋幸生・水光雅則 (編) (2005). 『これからの大学英語教育—CALLを生かした指導システムの構築』岩波書店.
- 竹蓋幸生, 草ヶ谷順子, 与那覇信恵. (2004). 外国語学部における英語教育改善の歩み(2) 平成14年度実施短期的計画での実験的指導. *文京学院大学外国語学部文京学院短期大学紀要*, (3), 1-15.
- 竹蓋幸生, 与那覇信恵, 草ヶ谷順子. (2006). e-learningの光と影に関する実証的研究. *文京学院大学外国語学部文京学院短期大学紀要*, (5), 147-162.
- 広島大学外国語教育研究センター. TOEIC® IP全学一斉実施 受験データ (PDF) (Retrieved on 2024/8/5 from <https://www.flare.hiroshima-u.ac.jp/certification/toeicip/data/>)
- 与那覇信恵, 根岸朋子, 阿佐宏一郎. (2015). 英語e-learning教材の学修実態に関する定量的分析. *文京学院大学外国語学部紀要*, (14), 37-48.
- 与那覇信恵, 竹蓋順子. (2013). 「英語教育総合システム」に基づいた英語学習の効果の検証 : LTM-CALLを継続使用した学習者群の長期的観察に基づく考察. *言語文化論叢*, (7), 43-59.