

教養教育英語科目受講生のスピーキング能力の推定

Estimation of Speaking Proficiency of Students in English for Liberal Arts Courses

横内 裕一郎*

Yuichiro YOKOUCHI

要 旨

本研究の目的は、弘前大学教養教育英語科目を受講する学生のスピーキング能力を推定し、スピーキング科目の改善に必要な情報を収集することである。本研究は2つの研究で構成されており、研究1ではクラスレベルによってVersant English Testの総合得点差が生じるかどうか、各クラスレベルの受講者のスピーキングレベルがCEFR基準でどの程度のものになるのかを検証した。研究2では、スピーキング再履修科目受講者を対象にデータを収集し、再履修科目受講者のスピーキング能力の特徴を明らかにすることと、1年時に実施したVELC Testとの得点の関係を観察することでVELC[®] Testがスピーキング能力を推定する上で十分な情報を持っているかどうかを検証した。その結果、学生のスピーキング能力は上級の学生のみ抜けて能力が高く、他のレベルの学生の能力は同程度である傾向が明らかになった。また、スピーキング能力がCEFRのA1未満～A2レベルの狭い範囲に限定される可能性が高いこともわかった。

キーワード：スピーキング能力、Versant English Test

現在、弘前大学の教養教育英語科目は4技能（Listening, Reading, Speaking, Writing）に分かれた授業を行っており、基本的に入学時に実施されるプレイスメントテストの結果をもとにクラス分けがなされている。プレイスメントテストには、英語能力測定・評価研究会（VELC研究会）が開発・運営する標準テストであるVELC Testを活用していることもあり、リスニングとリーディングの能力は直接測定されているものの、スピーキングとライティングの能力は間接的に推測することしかできていない。リスニングやリーディングの能力がスピーキング能力の素地となっており、参考情報としてはプレイスメントテストの結果や前期末に実施される共通テストとしてのVELC Testの成績は尊重されるべきである。だが、それとは別に実際にスピーキングやライティング指導するために、それぞれのレベルの学生がどの程度の能力を持っているのかを把握することは重要である。そこで本研究では、弘前大学在学生のスピーキング能力を明らかにすることを目的に、Versant English Test（以下Versant）を用い、各レベル（上級・中一上・中一中・中一下）の4つのレベルの学生のスピーキング能力を測定した。さらに、再履修者を対象として同様の調査を行い、再履修者のスピーキング能力を観察した。最後に、VELC Testの結果とVersantの結果の対応関係を観察することにより、入学時に行ったクラス分けがスピーキングクラスのクラス分けに対応しているかどうかを検証した。

* 弘前大学 教育推進機構 教養教育開発実践センター

Center for Liberal Arts Development and Practices, Institute for Promotion of Higher Education, Hirosaki University

Versant English Testについて

Versant English Test

Versantはスピーキング能力を測定するテストとして、長く使われてきた。これまでに報告されている例を見てみると、留学前後でのスピーキング能力の変化を検証した研究（清水他, 2014）や発表語彙の推定のために使用された例（横川, 2006）などがVersantを使用した調査の例として挙げられる。Versantは前身のPhonePass SET-10を元に、電話やPCの専用ソフト、スマートフォンのアプリケーションやPCのブラウザを活用してスピーキング能力を測定するために開発されたテストである。Versantは英語以外にもスペイン語やアラビア語などの多数の言語に対応したテストであり、音声認識技術を応用してスピーキング能力を測定するテストになっている。このテストの妥当性や信頼性は、Pearson（2008）で詳細に報告されており、有用性についてはDowney et al.（2008）で報告されているとおり、様々な研究でテストの質の検証がなされている。このテストは、2020年現在、スマートフォンのアプリケーションもしくはPCのブラウザを使用して受験する形式がとられている。先述の通り、成績は自動採点で行われる。全体で6つのセクションがあり、(A)音読問題（8問）、(B)復唱問題（16問）、(C)簡単な質問（24問）、(D)文の構築問題（10問）、(E)再話問題（3問）、(F)自由回答問題（2問）の計63問で構成されている。試験結果は受験直後に参照でき、総合得点と文章構文、語彙、流暢さ、発音の4観点のサブスコアが表示される（Pearson, 2008）。得点は20点から80点の間で表示され、総合得点からCEFRのレベルやTOEICの相当点、TOEFL iBTのSpeaking sectionの予測得点などがフィードバックとして返される。タスクの形式は、(A)の音読問題以外は典型的なリスニングスピーキングの技能統合型の問題が複数出題される形になっている。また、問題数が多い割に20分弱で試験が終了するとおり、テンポよく問題に回答する能力が問われる形式となっている。さらに、(B)や(D)の問題では特に短期記憶の能力が必要となっており、受験者の認知能力が高くなければ十分に回答できない問題構成になっている。本研究では、この形式のテストを用いて弘前大学教養教育英語科目受講者のスピーキング能力を推定し、教養教育科目の改善のために必要な情報をまとめる。

研究1・分析結果：参加者のスピーキング能力の推定

目的

研究1の目的は、弘前大学教養教育英語科目を受講する学生のスピーキング能力を推定することと、受講しているクラスのレベル間で差があるかどうかを検証することである。これらが明らかになることによって、スピーキング科目の授業内容や使用テキスト、シラバス等の改善に繋がる情報が得られると考えられる。

研究参加者

2018年12月と2019年10月にVersant English Testを用い、教養教育英語科目のスピーキングを履修している学生を対象にデータの収集を行った。2018年12月には上級、中級一上、中級一中、中級一下、初級の5レベルのクラスの内、中級一下のクラスに所属する学生を対象にデータ収集を行い、2019年10月には上級・中級一上、中級一中のクラスに所属する学生からデータを収集した。いずれも各科目の受講生の自由意志で受験するように指示を出した。分析の対象となった学生は計120名で、上級35名、中級一上が20名、中級一中は21名、中級一下は44名となった。なお、初級を受講する学生にとってVersantの難易度は高すぎるのが想定され、床効果が現れることが予想されたため、研究・分析の対象外とした。また、採点の結果、「テストのかなりの部分が不適切であるか、もしくは、英語で理解が不可能であるため、最低スコアと判定されました」と表示され点数の判定ができなかった受講生のデータは分析

から除外した。これは、実際に能力が低く、スピーキング能力が低いと判定されたのか、機器の不具合等が理由で適切にテストを受験できなかったのかを判別できないためである。

研究・分析の流れ

研究協力者には、各自受験用の公式スマートフォンアプリケーションを用いるか、PCブラウザ版のVersantを指定された1週間以内に受験するように指示を出した。その後、得点はVersantの成績をオンラインで確認したり収録された音声を聞いたりすることのできるVersant Score Keeperで確認した。研究1の分析においては、各クラス別の平均得点・標準偏差・95%信頼区間を算出し、その値を観察した。なお、この分析結果については、紙幅の都合により、付録にデータを掲載した。また、Versant Score Keeperのスコアレポートに記載されているCEFRレベルを転記し、クラスレベルと適合しているかを表にまとめた。

RQ1: Versant総合得点は各クラスレベルによって差があるか

RQ2: Versant総合得点に基づくCEFRレベルはどの程度でクラスレベルに沿った形になっているか

分析結果（研究1）

RQ1に対応するため、クラスレベルを独立変数、Versant総合得点を従属変数とする一元配置分散分析を実施した。その結果、 $F(3, 116) = 21.36, p < .00, \eta^2 = 0.36$ となり、各クラス間で得点に差があったことが明らかとなった。事後検定の結果から、上級クラスのみが突出しており、中級の3クラスにおいては差が生じなかったことが明らかとなった。上級の学生が中級の学生より有意に成績が高かったことは想定通りであり、順当な結果であると言える。一方、事後検定の結果、中級の3クラスレベルの間に有意差はなく、平均値のみを観察すると中一中のクラスが中一上のクラスよりも平均点が高い結果となっていた。また、中一上のクラスと中一下のクラスの平均値の差が殆どないことも特筆に値する。入学時のプレースメントテストの結果から考えると、中一上のクラスと中一下のクラスのレベル差は大きく、VELC Testの結果に基づくTOEIC相当点で比較するとおよそ120点程度の差がある程度にはレベル差がある。しかしながら、スピーキングという観点で言えば大きな差がないということは今後の授業計画等に有用な情報となったと言えるだろう。上記の結果になった理由として、2019年度に実施したデータ収集は10月初頭の授業が始まっていない時期にデータを収集したのに対し、2018年度に行ったデータ収集（中一下を対象）は12月にデータ収集を行っており、授業を実際に受けていたために成績が想定よりも高かった可能性がある。このように、データ収集の時期のずれは本研究における限界点の1つだろう。実際、図1の分布を見ても中一下の受験者の分布が広く、標準偏差も他に比べて大きな値となっている（付録1参照）。更に付け加えると、中一上の分布がかなり歪んだ形になっており、上級とのレベル差が明確な形で出ていることがわかる。

RQ2はVersant総合得点に基づくCEFRレベルとクラスレベルの対応関係を検証することを目標としている。それを検証するために、表1の通り、Versant English Testの総合得点から算出されたCEFRのレベル分布をまとめた。上級の学生にA1マイナスの受講生はおらず、全員がA1レベル以上だったうえに、A2レベルが全体の7割を占めた。一方、その他の中級のレベルではA1レベルが7割以上であり、中一下のクラスではA1マイナスの学生が16%もいたことが明らかとなった。この結果からVersantの総合得点から算出されたCEFRレベルは実情のクラスレベルに沿っているものと推測できる一方、B1レベル以上がない結果となった。プレースメントテストとして実施された際のVELC Testの得点に基づくTOEICの予測得点から考えると、TOEICの各セクションで275点を超えるレベルの受験者はCEFRのレベルでB1以上（Tannenbaum & Wylie, 2008）とされていることから上級レベルの受講生にB1以上の学生がいなかったことは筆者の想定に反する結果となった。この結果になった理由は3つ考

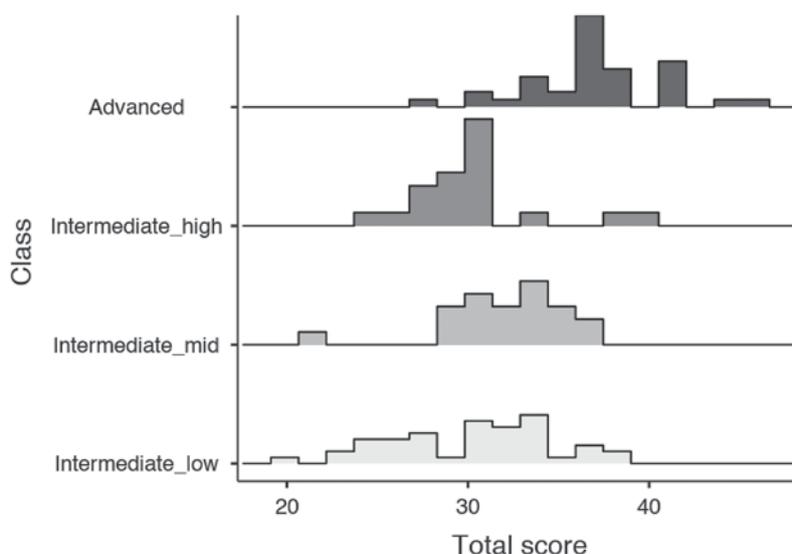


図1 各レベルの得点分布（総合得点）

えられる。1つはVersantのテストの特徴がリスニングスピーキング型の技能統合型テストであるということだ。リスニング能力が低い受講生の成績は自ずと低いものとなることが容易に想定できる。音声の指示を十分に理解できなかった場合には発話につなげることができないためである。それに関連して、2つ目は短期記憶の容量を問うような問題が出題されていることから、リスニングやスピーキングの能力以前に短期記憶の能力が総合得点に影響を及ぼした可能性は否定できない。多くの受験生は、スピーキング能力を発揮する前の段階で躓いてしまった可能性が高いと考えられる。これらの影響を最小限にするために、データ収集前に練習問題を提示するなどして問題の特性を受講生に周知していれば結果が変わっていた可能性がある。最後に、2019年に実施したテストでは、後期開始直後にデータ収集を行っていることから、多くの学生にとってスピーキングテストを受験する機会は今回が初めてであったと想定されるためである。

表1 クラスレベルごとのVersantによるCEFR相当の結果 (N = 120)

	A1マイナス	A1	A2
上級 (n = 35)	0 (0%)	10 (29%)	25 (71%)
中一上 (n = 20)	1 (5%)	17 (85%)	2 (10%)
中一中 (n = 21)	1 (5%)	18 (86%)	2 (10%)
中一下 (n = 44)	7 (16%)	32 (73%)	5 (11%)

注. カッコ内はクラスレベルごとの人数に占める各CEFRレベルの割合で、小数点以下を切り上げているため必ずしも合計が100%になるとは限らない。

以上のように、2018年度と2019年度に実施された研究の結果からは、弘前大学の教養教育英語科目を受講する学生のうち、上級の学生の多くはA2レベル以上の能力を持っており、中級レベルは大多数がA1レベルになるということが明らかになった。また、スピーキング能力の観点からは、中級と上級に明白な差が生じていることもまた明らかになった。この情報をもとに、今後教科書の選定や指導に使用するタスクの分類などを行う必要があるだろう。

研究 2 ・ 分析結果：再履修者のスピーキング能力の推定と Versant と VELC Test の相関

目的

研究 2 の目的は弘前大学において教養教育英語科目を受講する再履修者のスピーキング能力を推定することである。単位を取得することができなかった原因が、基本的に学力（スピーキング能力）の不足にあるのか、授業に参加しなかったことにあるのかについて学力の観点から調査した。また、Versant と VELC Test の相関を確認することにより、前期末の VELC Test の得点に基づいてスピーキングのクラス分けが可能であるかどうかを検討することである。なお、2 つ目の分析は本来であれば研究 1 のデータと併せて実施すべきではあるが、個人を特定せず、クラスごとに ID を管理していたため、VELC Test の得点との紐付けができなかったため、本研究で分析するデータ提供者のみを分析対象としている。

研究参加者

研究 2 の研究参加者は Speaking の再履修クラスを受講する初級～中級の学生であり、VELC Test を 1 年時の入学時と前期末に受験している学生で Versant を受験した学生を分析対象とした。そのうち初級レベルに該当する学生数は 26 名、中級は 23 名、上級は 2 名の計 51 名が分析対象となった。

研究の流れ

研究 1 同様、研究協力者には自由意志で研究に参加するよう依頼し、基本的にスマートフォンの Web アプリケーションを用いて受験するよう指示を出した。受験期間は告知から 20 日間とし、その間に受験した研究協力者のデータを分析した。研究 2 の RQ は以下の 2 つである。

RQ3: 教養教育英語科目の再履修科目を受講する学生は Versant でどの程度得点できるか

RQ4: VELC Test と Versant の相関はどの程度か

分析結果（研究 2）

RQ3 に対応するため、記述統計量を確認した。なお、RQ1 と同様に一元配置分散分析を実施しなかった理由は、上級受講者の人数が ($n=2$) であり、分析の仮定を満たすことができないと判断したためである。分析の結果（詳細は付録 2）、中級の学生については、統計的な差ではないが、再履修者の得点が高い傾向にあることがわかった。再履修者については、1 年時に中級のどのレベルの学生が受講していたかについて特定できなかったため、具体的な説明はできないが、少なくとも 2019 年度に実施した研究の参加者に比べ、わずかながら得点が高いことがわかる。また、2018・2019 年度は分析の対象とならなかった初級の学生については当初著者が想定していたよりも得点が高かった傾向がある。参加者の発話パフォーマンスが極めて低く、「テストのかなりの部分が不適切であるか、もしくは、英語で理解が不可能であるため、最低スコアと判定されました」と結果が出力された受験者は先述の理由の通り分析から除外されたため、一部の除外された受験生の能力が低い可能性が高く、スピーキング能力の全体の実態は、本研究の結果に比べてより低い可能性は否定できない。

RQ4 は VELC Test と Versant の成績の関係を検証することが目的である。VELC Test と Versant の成績の相関係数が高ければ、前期末の VELC Test の結果を用いて後期開講のスピーキングクラスのプレイスメントが可能となると考えた。分析の結果、入学時と Versant の相関係数は $r=.404$ でやや低い～中程度の相関であったのに対し、前期終了時の VELC Test の成績と Versant の相関係数は $r=.457$ とわずかながら前期終了時のほうが、相関係数が高い結果となった。しかし、この程度の相関では十分に相関が高いとは言えず、構成概念の異なる VELC Test を用いてスピーキングクラスのプレイスメントを用いる上では説得力が足りないと考えられる。したがって、現状では、前期末の VELC Test の成績を後期スピーキン

グ科目のプレイズメントテストとすることは難しいと結論づけられる。

仮にVELC Testをスピーキングクラスのプレイズメントに活用するのであれば、VELC Testの形式がリスニング・リーディング・語彙を構成概念の軸としたものであることから、スピーキングの授業の多くでリーディングスピーキング型やリスニングスピーキング型の応用的なタスク統合型の課題を授業で実施することによってプレイズメントの意味をもたせることが可能となるかもしれない。統合型のタスクを中心にスピーキングの授業を行うことにより、授業内容にリーディングやリスニングの要素を組み合わせることが可能となる。そうすることによって、後期のVELC Testの結果にも良い影響を与えられる可能性があると考えられる。このようにスピーキングのカリキュラムを改革することによってVELC Testの成績を指導にも反映しやすくなる上に、後期末開催のVELC Testでの点数の低下（横内 & 立田, 2020）を抑制することに繋がる可能性がある。

結論

本研究の全体的な限界点として言えることは、データ収集に併せてこれまでのスピーキング学習歴などの調査をしなかったことが挙げられる。これまでのスピーキング学習歴を検証することで、各学生の能力と学習経験を紐付けて分析ができなかったことが挙げられる。また、2018–2019年度実施のデータでもVELC Testとの紐付けを行った上で分析ができていればより深い考察が出来ただろう。

本研究全体を総括して言えることは、教養教育英語科目のスピーキングで使用されている教科書の難易度は今後再検討する余地があるということだ。各出版社が公表している教科書のレベルを勘案すると、上級クラスではB1～B2レベル相当の教科書を採用している場合もある。もちろん、Versantの結果が全てではなく、繰り返し述べている通り本研究に協力した学生がスピーキングテストそのものに慣れていない、そして英語で話す行為自体に慣れていない可能性を考えれば、結果的に難しめの教科書を採用する意味は十分にある。果たしてどの程度教科書の内容を理解していたか、実際のパフォーマンスを参考に15回の授業終了時にどのような能力を身につけることが出来たのかを検証し、スピーキングの授業で使用されるべきテキストの選定を行う必要があるのではないかと考える。

また、今回の分析結果をもとに、シラバスに記載の「授業としての具体的到達目標」をより具体的なものに変更する必要があると考える。現行のシラバスでは、Speaking（上級）の具体的到達目標における言語的な側面は、会話を続けるための方略を覚えることと状況に応じて適切な表現を使うことに焦点が絞られており、内容面も「身近な出来事や意見、考えなど」を扱っている。CEFRのSelf-assessment gridなどを参考にしても、話すことの能力で、自身のことや身近な内容を扱っているレベルはB1程度までとなっており、それ以上のレベルにおいては社会的な内容を話すことができる能力が求められている（Council of Europe, 2001）。したがって、今回の結果とシラバスに記載の内容を照らし合わせると、授業開始時にA2レベル相当であった上級の学生については、すでに身につけている能力を授業内で扱っている可能性が高く、シラバスの「授業としての具体的到達目標」の改訂の必要性が明らかになったと言えるだろう。中級と初級に関しても可能な限りシラバスの到達目標の具体化は行われるべきである。今回Versantを実施して収集した発話データを分析し、例えばある特定のトピックで1分間に意味のある発話を何語程度話す能力を身につけられれば各レベルとして合格なのか、という観点で検証を行う必要があるだろう。

参考文献

- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the Usefulness

- of the Versant for English Test: A Response. *Language Assessment Quarterly*, 5(2), 160–167. <https://doi.org/10.1080/15434300801934744>
- Fox, J., & Fraser, W. (2009). Test review: The Versant Spanish™ Test. *Language Testing*, 26(2), 313–322. <https://doi.org/10.1177/02655322090260020602>
- Jamieson, J., Wang, L., & Church, J. (2013). In-house or commercial speaking tests: Evaluating strengths for EAP placement. *Journal of English for Academic Purposes*, 12(4), 288–298. <https://doi.org/10.1016/j.jeap.2013.09.003>
- Pearson. (2008). *Versant English Test: Test description and validation report*. <http://www.versanttest.co.uk/pdf/ValidationReport.pdf>.
- 清水裕子, 桐村亮, & 野澤健. (2014). 経済学部英語圏短期留学プログラムにおけるスピーキング・テストの実施とその結果報告. *立命館高等教育研究*, 14, 91–102.
- 鈴木正紀, & 原田康也. (2011). 大学新入生の英語リスニング・スピーキング熟達度の定点観測. *電子情報通信学会技術研究報告*. TL, 思考と言語, 111, 37–42. <http://www.f.waseda.jp/harada/documents/2011-11-26-TL-45.pdf>
- 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. *ETS Research Report Series*, 2008(1), i-75. <https://www.ets.org/Media/Research/pdf/RR-08-34.pdf>
- 横川博一. (2006). 生起頻度・親密度の観点からみた日本人英語学習者の自発的発話における使用語彙. *神戸大学国際コミュニケーションセンター論集*, 3, 59–77.
- 横内裕一郎, & 立田夏子. (2020). 教養教育英語科目における VELC Test® の導入と結果分析. *弘前大学教養教育開発実践ジャーナル*, 4, 39–47. https://hirosaki.repo.nii.ac.jp/?action=repository_uri&item_id=6053&file_id=20&file_no=1

付録

研究 1 : Versant English Test のレベル別スコア

	Total score		Sentence mastery		Vocabulary		Fluency		Pronunciation	
	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI
上級 (<i>n</i> = 35)	36.91 (3.91)	[33.00, 40.83]	38.43 (5.21)	[33.22, 43.64]	39.14 (6.93)	[32.21, 46.07]	34.86 (6.02)	[28.84, 40.88]	35.74 (5.59)	[30.16, 41.33]
中—上 (<i>n</i> = 20)	30.20 (3.66)	[26.54, 33.86]	33.70 (5.57)	[28.13, 39.27]	29.85 (4.92)	[24.93, 34.77]	26.25 (5.75)	[20.50, 32.00]	31.30 (4.81)	[26.49, 36.11]
中—中 (<i>n</i> = 21)	32.05 (3.44)	[28.61, 35.49]	34.86 (5.24)	[29.62, 40.09]	32.00 (5.22)	[26.78, 37.22]	29.33 (6.38)	[22.95, 35.72]	32.38 (4.46)	[27.93, 36.84]
中—下 (<i>n</i> = 44)	30.18 (4.44)	[25.74, 34.62]	34.18 (5.64)	[28.54, 39.82]	28.73 (6.32)	[22.41, 35.05]	29.33 (5.96)	[20.75, 32.66]	31.68 (3.51)	[28.17, 35.19]

研究 2 : Versant English Test のレベル別スコア

	Total score		Sentence mastery		Vocabulary		Fluency		Pronunciation	
	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI
上級 (<i>n</i> = 2)	32.50 (9.19)	[23.31, 41.69]	33.00 (15.56)	[17.44, 48.56]	33.50 (19.09)	[21.75, 33.79]	31.50 (0.71)	[30.79, 32.21]	33.50 (3.54)	[29.96, 37.04]
中級 (<i>n</i> = 23)	31.87 (3.24)	[28.63, 35.11]	35.13 (4.64)	[30.49, 39.78]	31.57 (4.63)	[26.94, 36.20]	29.65 (5.21)	[24.44, 34.86]	30.96 (4.15)	[26.81, 35.11]
初級 (<i>n</i> = 26)	28.12 (4.28)	[23.83, 32.40]	32.15 (6.35)	[25.81, 38.50]	27.77 (6.02)	[21.75, 33.79]	25.42 (3.74)	[21.68, 29.17]	27.69 (4.72)	[22.97, 32.41]