

CEFR-Jに基づいたCan-Doディスクリプタの信頼性と相関関係の基礎的調査 —M大学教養英語の事例—

渡慶次正則¹⁾, Norman Fewell¹⁾, 津嘉山淳子¹⁾, 名城 義久¹⁾

A fundamental study on the reliability and correlation of CEFR-J based Can-Do Descriptors —A case of M University freshman English program—

Masanori TOKESHI¹⁾, Norman FEWELL¹⁾, Junko TSUKAYAMA¹⁾, Yoshihisa NASHIRO¹⁾

要 旨

本研究はCEFRの日本版であるCEFR-Jを大学教養英語プログラムに利用するために、CEFR-JのCan-Doディスクリプタの自己評価について信頼性と英語能力試験との相関関係を検証する事を目的とした。M大学の1年生(N=389)を対象として、CEFR-Jに基づいたアンケートで英語能力を自己評価した回答データを分析した結果、次の事を示した。5技能分野の自己評価平均値は全体的な傾向として内的信頼性は強い値(クロンバックの α :0.87)を示した。しかし、個人単位のデータを質的に詳細に分析すると、CEFR-Jレベルの順序に沿って個々の回答が対応していない自己評価が多く、ばらつきがある事を示した。質的な分析では個々の自己評価と英語能力試験順位は強い整合性を示さず、その相関関係の統計的分析(ピアソン係数:0.3程度)も同様に、自己評価と英語能力試験の相関関係が弱い事を示した。本研究の結論として、CEFR-JのCan-Doディスクリプタの自己評価の信頼性は強くなく、英語能力試験とCan-Doディスクリプタ自己評価は強い相関関係でないことを分析結果は表した。今後の日本の大学英語教育についてCEFR-Jを用いてCan-Doリストを作成したり、英語能力の絶対的な評価の基準として用いることを慎重に考慮されなければならない事が示唆される。

キーワード：カリキュラム、評価、CEFR、教養英語

Abstract

This study is intended to adopt into university freshman English programs the CEFR-J, the Japanese version of the CEFR, so that reliability and correlation with an English proficiency test regarding self-ratings of CEFR-J Can-Do Descriptors can be examined. The participants (N=389), M University freshmen answered the CEFR-based questionnaire, in which they rated their own English proficiency. Analysis of the questionnaire indicated the following. The overall tendency was that average self-ratings among five skill categories showed strong intercorrelations (Cronbach's alpha: 0.87). On the other hand, in-depth qualitative analysis of individual data showed that a large part of self-ratings did not follow the order of the CEFR-J difficulty levels and self-rating responses varied depending on the individual. Moreover, the statistical analysis between self-ratings and the English proficiency test rankings did not show strong correlation (Person's r: 0.3). This study concluded from the analysis that self-ratings of CEFR-J Can-Do Descriptors are not strongly reliable. In addition, self-ratings of Can-Do Descriptors and

¹⁾ 名桜大学国際学群 〒905-8585 沖縄県名護市為又1220-1 Faculty of International Studies, Meio University, 1220-1, Biimata, Nago, Okinawa 905-8585, Japan

the English proficiency test are not strongly correlated. The results suggest that, for future university English education in Japan, designing Can-Do list based on the CEFR-J and using the CEFR-J as the absolute standard of English proficiency should be considered with caution.

Keywords: curriculum, assessment, CEFR, freshman English

1 はじめに

本研究は、日本の英語教育の以下の2つの問題点に関する示唆を探る基礎的な調査である。

一つ目の問題点として、日本の大学英語教育は英語能力の到達度測定として、共通に認知された評価基準を持たないことが指摘される。その理由として、小学、中学、高等学校の学習指導要領の様に、共通した大学の英語教育のコアカリキュラムが定められていないことにある。Negishi, Takada & Tono (2012, p.136)によると「我々(日本の大学)は、これまで言語教育において同意した達成目標を持たなかった。その目標をどの様に達成し、どのように評価するについての合意がない (we have not had any agreed attainment target in language teaching so far. Nor do we have any consensus as to how to attain those targets or how to assess the attainment)」と指摘する。

これまで、英語検定やTOEIC, GTEC (ベネッセ社)などが英語能力の達成基準測定として用いられてきたが、これらは主にリスニングとリーディングを測定するテストで、4技能をすべて測定するテストではない。TOEFL iBT¹⁾は、世界で最も多くの国・地域(約130か国)と教育団体(約9000団体)で受験されている英語能力試験で、総合的に4技能を測定し、英語能力試験としては最も国際的に有望なテストではあるが、実際の教育現場で利用するには、実用性において高額な受験料、試験センターの確保、試験官の養成などの問題がある(渡慶次, 2013)。しかし、大学生の英語能力の到達度を測定する共通基準の導入は急務である。グローバル化の進展とともに、日本の大学生の英語能力について意思疎通ができる国際的な共通の評価基準が必要である。

もし、日本の大学生が他国の学生より英語能力が劣っているとしたら、カリキュラムにどのような教授や学習、評価を含むべきか共通の枠組みが必要である。

二つ目の問題点として、日本人英語学習者の英語能力は、国際的に通用するレベルではない事が指摘されている。その根拠としてよく用いられるのは、前述の英語能力試験であるTOEFL iBTである。同テストの作問・実施団体のETS (2014)によると、2014年実施の国・地域別の比較では、日本人獲得平均点は他国に比べてかなり低い。アジア地域で比較すると、TOEFL iBT (120点満点)の

日本人獲得平均点は70点で、30か国・地域中で27位であった。一方、日本と経済的にライバル関係にあるとされるアジア地域の国や地域の受験者の平均点は、韓国 (85点, アジア地域8位)、中国 (77点, アジア地域19位)、台湾 (80点, アジア地域13位)であり、いずれも日本語と同様に母語と英語の言語的な距離がありながら、日本人受験者より高得点を獲得している。

国際的に低い日本人の英語能力の向上や2020年開催東京オリンピックに対応するために、文部科学省 (2015b)は英語教育改革を発表し、すべての学校種に於いて、国際的に通用する英語力を持つ学習者を育成する事を提唱している。具体的施策のひとつとして、文部科学省 (2015a)は大学入学者選抜に4技能を測定するTOEFL iBT, IELTS²⁾, GTEC CBT³⁾, TEAP⁴⁾などの外部試験の導入を検討しているが、それらの測定基準の理論的枠組みとしてヨーロッパでの40年近くの研究成果であるCommon European Framework of Reference for Languages (ヨーロッパ共通参照枠, CEFRで以下省略)を活用する事を提案している。6レベルのCEFRは、点数やスケールで示す英語能力試験とは異なり、学習者が獲得した英語能力でできる事 (Can-Do) をレベル別に記述しており、英語能力が具体的に観察可能で理解しやすい利点を持つ。民間の団体である英語検定協会はすでに、2015年第1回英語検定より、CEFRを測定の枠組みとして利用する事を発表している。さらに、過去はTOEIC点数によって示してきた英語テキストのレベル表示として、現在は多くの英語テキストがCEFR (またはCEFRに準じたCEF) によって表示されており、急速にその共通枠レベルの利用が広がっている。しかし、語学関係の出版社や外部試験団体により急速に利用され始めたCEFRは、多くの現場英語教師には十分に理解されていない感を筆者は持つ。

Council of Europeより2001年に出版され、ヨーロッパでの40年近い言語教育研究の集大成であるCEFRは、ヨーロッパを中心に英語のみならず、39言語⁵⁾に活用されている。CEFRは、言語教授、学習、評価の国際基準であり (North, Ortega, & Sheehan, 2010)、最も有望な英語教育の道具セット (toolkit) である (Norton, 2014)。日本においても大学英語教育のカリキュラムや到達度測定の共通の枠組みとしてすでに多くの研究チームにより研究がなされている (例, 小池科研 (2004-2007),

投野 科 研 (2008-2011), 川 成 科 研 (2010~2012))。Fugueras (2012) は, CEFRは21世紀の言語教授, 学習, 評価において最も適切な, かつ議論を巻き起こす文書である述べている。一方, CEFRに批判的な研究者がいるにも拘わらず, 評価研究者の第一人者であるAlderson (2007)は, 「ヨーロッパの言語教育に従事している者は, 誰もCEFRの存在を無視することができない (nobody engaged in language education in Europe can ignore the existence of the CEFR)」(p.660) とCEFRの重要性を認めている。CEFRについては, 次節で詳しく述べる。

本研究で検証する12レベルのCEFR-J (投野ほか, 2012) は, 2004年から2011年まで8年間, 前述の小池生夫科研や投野由紀夫科研の英語教育研究グループによって研究開発された成果であり, 日本人英語学習者用に修正したCEFRの日本版である。開発をしたNegishiら (2012) によると, 日本の状況に最も合わせて設計されていると説明している。今後, 日本での英語教育の理論的枠組みとして利用される可能性は非常に高い⁶⁾。すでに, 文部科学省が全国の小学校, 中学校, 高等学校に作成を要請している英語Can-Doリストのモデルとして検討されている (例, 文部科学省, 2014)。CEFR-Jについても次節で詳しく述べる。

CEFR-Jの大きな可能性にも拘わらず, 日本の英語教育に応用される前に実証的な研究により検証される必要がある。長期間に渡る緻密な研究プロジェクトの成果物であるCEFR-Jは, 大学レベルで検証された調査はわずかである (例: Runnels, 2013; Runnels, 2014)。Runnels (2014) は, 「実際, 日本人 (英語) 学習者についてCEFR-Jに準じたタスク運用と自己評価, (英語) 能力の関係についてほとんど研究がなされていない (In fact, little research on the relationship between ability, self-assessment, and CEFR-aligned task performance for Japanese learners has been carried out)」(p.86) と指摘する。CEFR-Jが日本の大学英語教育のカリキュラム開発や教授や学習, 評価に導入される前に検証される必要がある。

本研究は, CEFR-Jの110のCan-Doディスクリプタを用いて, M大学1年生 (N=389) を対象にWebアンケー

トの形式で自己評価を実施し, 次のリサーチ・クエスチョンについて答えることを調査目的とする。英語能力試験として, 英語能力判定テスト (Bタイプ, 英語検定協会作成) を利用した。

- (1) CEFR-JのCan-Doディスクリプタと5技能分野の関係はどの程度あるか。
- (2) CEFR-Jの自己評価結果と英検能力判定テストとの相関関係はあるか。
- (3) 本研究から得られる大学英語プログラムへの示唆は何か。

なお, Can-Doディスクリプタは, CEFR-Jの5技能分野の12レベルにある110の英語でできる事の記述で, ディスクリプタは記述子とも訳されるが, 本稿ではディスクリプタを用いる。例として, リスニングのPreA1レベルのCan-Doディスクリプタは, 「ゆっくりはつきりと話されれば, 日常の身近な単語を聞きとることができる。」である。

2 文献研究

2.1 CEFR開発の背景と内容

第1次世界大戦と第2次世界大戦のヨーロッパでの混乱の反省から, 平和を求める活動や, European Union (EU) 圏の人々の自由な移動に対応を求められる状況下で, 2001年にCouncil of EuropeによりCEFRが40年近い研究成果として出版された。その出版は1970年代のvan-Ek (1975) やWilkins (1976) らのコミュニケーション・機能の研究に起源を遡り, Council of Europeにより出版された初期のThresholdシリーズの出版物である, “Threshold” (van Ek & Trim, 2001b) や “Waystage” (van Ek & Trim, 2001a), “Vantage” (van Ek & Trim, 2001c) をまとめた物である (表1を参照)。CEFRは1996年のDIALANGプロジェクト (Council of Europe, 2001, pp.226-243) とThe ALTE (Association of Language Testers in Europe) のCan-Doプロジェクト (ALTE, 2002) の2つの大きなプロジェクトに基づいて開発された。CEFRの基本原理はヨーロッパの複言語主義⁷⁾ と複文化主義を反映してお

表1 ThresholdシリーズとCEFR (出典: 投野, 2013, p.47)

| CEFR | | Thresholdシリーズ |
|---------------------------------|----|-------------------------------------|
| 基礎段階の言語使用者 (basic user) | A1 | Breakthrough |
| | A2 | Waystage |
| 自立した言語使用者 (independent user) | B1 | Threshold |
| | B2 | Vantage |
| 熟達した言語使用者 (proficient user) | C1 | (Effective Operational Proficiency) |
| | C2 | (Mastery) |

り、日本とは背景が大きく異なる。

CEFRは6レベル(A1-C2)のCan-Doディスクリプタを5技能分野で示している。5技能分野とは、リスニングとリーディング(理解すること)、ライティング(書くこと)に加えてスピーキングをやり取りと発表(話すこと)に分けた区分である。CEFRは言語使用、公的、私的、教育的、職業的の諸領域に分けられ、各々の領域は場所、施設、人々、物、出来事で具体化されている(吉島・大橋, 2004, pp.9-43を参照)。A1レベルとA2レベルは初級レベルであり、C1レベルとC2レベルは母語話者に近いレベルである。英語圏での第1世代の移民の多くはB1レベルで留まるとされ(North, 2014, p.50)、日本の学校教育における最終到達目標をB1レベルまたはB2レベル(文科省, 2015bによると英検2級から準1級)または、TOEFL iBT80点程度(文科省, 2015a)としている。

2.2 CEFRの強みと批判

CEFRには、特徴的な強みがある。CEFRは行動・結果中心のアプローチであり、前述したように学習者の言語運用能力をその基準(6レベル)に準じて測定する。その基準枠組みには、外国語で「できる事」を、学習者が自身の言語能力について判断できるようにCan-Doディスクリプタの形式で共通参照レベル: 自己評価表⁸⁾(Council of Europe, 2001, p.26; 吉島・大橋, 2004, pp.28-29を参照)で記述している。その中心的な考えは、言語能力よりコミュニケーションと学習者・使用者にある。CEFRは、言語教育の「明瞭な目標や内容、方法について共通の礎(a common basis for the explicit objectives, content and methods)」(Council of Europe, p.1)として開発され、コースカリキュラムやテキスト、試験の開発に役割を期待されている。CEFRの正式名称は、Common European Framework of Reference for Languages: Learning, teaching, assessmentであり、外国語の学習、教授、評価に供与するものである。つまり、CEFRは、何が評価されるか、言語運用能力がどの様に解釈され、異なるテスト間をどの様に比較するかを教師、学習者、ステークホルダー(教育政策者や保護者、企業など)が共有して理解する事に役に立つ基準である(North, 2014)。

CEFRの世界規模の利用普及にも拘わらず、その限界について一部の研究者による批判がある。Figueras(2012)はCEFRの及ぼす影響とカリキュラムや学校教育、評価への導入には不整合があると指摘する。

CEFRの枠組み基準を試験へ導入することには、強い批判がある。評価研究者ら(Weir, 2005; Alderson, 2007; Little, 2010)は、CEFRの試験に対する理論的根拠に批判的で、CEFRを試験に用いた実証的な検証を求めている。Weir(2005)の先行研究のレビューによる

と、実際にCEFRを用いて、テスト開発やテスト間の比較を試みた多くの研究が困難を経験していると分析している。

Jones(2002)の調査結果は、本研究に特に関連がある。JonesはCan-Do自己評価(アンケート方式)をケンブリッジ英語検定(KET, PET, FCE, CAE, CPE)⁹⁾と相互比較した。その調査結果によると、それぞれの英語テストの示すレベルと受験者自身の英語能力の自己評価についての整合性に大きなばらつきがあることを示した。興味深いことに、英語能力が低い回答者は自分の実際の能力より高く評価し、自分の評価に甘い傾向があり、逆に英語能力の高い回答者は、自分の英語能力に対して謙虚(実際より低く評価した)である傾向があった。Jonesは調査結果に対して年齢や英語能力などの個人的背景により「人はcan-doを異なって理解しがちである(people tend to understand 'can-do' differently)」(p.33)と結論付けている。同様に、Little(2010, pp.159-160)は、自己評価について次の懸念を指摘している。1)学習者は自分の評価の仕方を知らない場合がある、2)自分の英語能力を過剰評価してしまう危険性がある、3)自分自身のEuropean Language Portfolio⁸⁾でない言語学習資料を含むことにより、不正を行うことがある。

本研究に関連がある別の問題は、CEFRの実証的検証である。North(2014)は、CEFRディスクリプタは、第2言語学習者能力に対する教師の見解に基づいて基準付けされていると批判する。つまり、彼によるとCEFRディスクリプタは長期に渡る第2言語習得の実際の過程についての大規模な縦断的調査から産み出せられた物ではないと指摘する(p.23)。さらに、実証的検証の問題についてHulstijn(2007)はCEFRについて英語能力の質的かつ量的な側面での十分な実証的研究がなされるべきであると同様な主張をしている。

2.3 CEFR-Jプロジェクトの経緯と検証調査

CEFR-Jが開発された経緯とその後について以下で説明する。

小池科研(2004年~2007年)を引き継いで、投野を研究代表とする科研グループ(2008年~2011年)は、研究最終成果として12レベルのCEFR-J Version 1(投野ほか, 2012)をWeb上で公開した。8年間に渡る、国内最高規模のCEFR-Jプロジェクトは、本格的には東京外国語大学で2008年に始まり¹⁰⁾、英語教育に従事する18人の研究者たちによって実施されたものである。

CEFRプロジェクトは3つのCEFR-Jバージョンを作成し、時系列で次のステージ(筆者によって分けられたステージ)を経て、まだ進化中ではあるが最終成果が公開された。CEFR-Jの開発に於ける膨大な実証的プロ

セスを明らかにした学術論文が皆無なために、Negishi, Takada & Tono (2012) と投野 (2013) の報告書を中心にまとめた。

第1ステージ(2006年)：Can-DoアンケートがDIALANG 自己評価記述(Council of Europe, 2001)を基に作成された。アンケート(できるか、できないかの二者択一)は日本語に訳され360人の日本人大学生が回答した(中島・永田, 2006)。さらに同じCan-Doアンケート(4肢選択のリックカート尺度)で727人の中高校生と大学生が調査された。それらの結果により、CEFRが日本人の英語学習者に適応できることを確認した。

第2ステージ(2004年～2007年)：この時期には調査参加者の英語能力を調査するためにさまざまな調査が実施された。具体的には異なる学校種の児童・生徒の調査参加者(354人の小学生, 150人の中高校生)に加えて7,354人のビジネスパーソンが調査された。これらの結果に基づいて、日本の英語学習者の80%以上が、CEFRレベルのA1とA2のレベル(基礎段階の言語使用者)に属すると結論付けた。

第3ステージ(2008年)：それまでの調査経過を踏まえて、CEFRの従来の6レベルは、CEFRの日本版(CEFR-Jアルファ版)として、12レベルに分けられた。CEFR-Jアルファ版は、ELP⁸⁾、Can-Doディスクリプタ、GTECテスト、Super English Language high school¹¹⁾、英語検定などに関する調査結果も加えて総合的に開発された。CEFR-Jの特徴は以下である(Negishi, Takada & Tono, p.143)。1) Pre-A1が加えられた、2) A1をA1.1とA1.2, A1.3の3レベルに分けた、3) A2をA2.1とA2.2の2レベルに分けた、4) B1をB1.1とB1.2の2レベルに分けた、5) B2をB2.1とB2.2の2レベル

に分けた、6) C1とC2についての変更はない。

第4ステージ(2009年)：CEFRの専門家であるDr. Anthony Greenから専門的アドバイスを受けて、産出スキルは1)言語運用、2)規準、3)状況に細分化され、一方受動的スキルは1)タスク、2)言語文字(テキスト)、3)状況に分けられた。さらに、CEFR-Jアルファ版は、各レベルのディスクリプタの順序について206人の英語教員対象に、難易度の適切さを確認する調査結果を受けて、並び替えが行われた。レベルの並び替えは調査参加教師の70%以上が合意した時に、ディスクリプタの並び替えが行われた。この様にしてCEFR-Jアルファ版は修正され、CEFR-Jベータ版が作成された。

第5ステージ(2010年～2011年)：CEFR-Jベータ版を実証するために、1,685人の中学生と2,538人の高校生、1,245人の大学生を対象に、アンケート方式でCan-Doディスクリプタについて英語でできる事の程度について4肢選択式で回答させた。続いて、CEFR-Jベータ版に関するそれらの調査の統計的分析で明らかになった問題点を解決するためにディスクリプタ記述が修正され、ディスクリプタの順序が並び替えられた。さらにプロジェクト・グループは5技能分野のディスクリプタについて運用能力テストを実施し、受験者の自己評価と実際の運用能力の関係を分析した。

第6ステージ(2012年～2013年)：CEFR-Jベータ版の実証過程を経て、CEFR Version 1が2012年3月にネット上で公開され(<http://www.cefr-j.org/download.html>)、さらに『CEFR-Jガイドブック』(投野編)が2013年に出版され、多くの英語教育関係者が知る事となった(表2を参照)。

表2 CEFR-J Can-Doディスクリプタの例(リスニング: Pre-A1レベルからA2.2レベルまで)

| | レベル | PreA1 | A1.1 | A1.2 | A1.3 | A2.1 | A2.2 |
|--------|--------|-------------------------------------|---|--|--|---|--|
| 理 解 | リス | ゆっくりはっきりと話されれば、日常の身近な単語を聞きとることができる。 | 当人に向かって、ゆっくりはっきりと話されれば、「立て」「座れ」「止まれ」といった短い簡単な指示を理解することができる。 | 趣味やスポーツ、部活動などの身近なトピックに関する短い話を、ゆっくりはっきりと話されれば、理解することができる。 | ゆっくりはっきりと話されれば、自分自身や自分の家族・学校・地域などの身の回りの事柄に関連した句や表現を理解することができる。 | ゆっくりはっきりと放送されれば、公共の乗り物や駅や空港の短い簡潔なアナウンスを理解することができる。 | スポーツ・料理などの一連の行動を、ゆっくりはっきりと指示されれば、指示通りに行動することができる。 |
| | ン グ | 英語の文字が発音されるのを聞いて、どの文字かわかる。 | 日常生活に必要な重要な情報(数字、品物の値段、日付、曜日など)を、ゆっくりはっきりと話されれば、聞きとることができる。 | 日常生活の身近なトピックについての話を、ゆっくりはっきりと話されれば、場所や時間等の具体的な情報を聞きとることができる。 | (買い物や外食などで)簡単な用をたすのに必要な指示や説明を、ゆっくりはっきりと話されれば、理解することができる。 | 学校の宿題、旅行の日程などの明確で具体的な事実を、はっきりとなじみのあはる発音で指示されれば、要点を理解することができる。 | 視覚補助のある作業(料理、工作など)の指示を、ゆっくりはっきりと話されれば、聞いて理解することができる。 |

CEFR-Jが公開・出版された後に、開発グループ以外によってCEFR-Jの検証調査がほとんどない中で、前述したようにRunnels (2014) により590人の日本人大学生を対象に調査報告がなされている。Runnelsの調査結果によると、CEFR-JのCan-Doディスクリプタについて調査対象者の記述内容への混乱と馴染みのない事が難易度の信頼性に影響を与えたとし、さらに調査参加者の個人差が自己評価の難易度の結果に影響を与えたと報告している。調査の結論として、RunnelsはCEFR-Jの再検証の必要性を訴え、「今後の調査に於いて、仕組み・組織において統一した、共通の解釈を確かにするために、CEFR-J利用者のcan-do記述に対する反応の検証とcan-do記述の内容分析がなされるべきである (Further studies on this, the CEFR-J's target users' responses to can-do statements, and content analyses should be performed to ensure a consistent, common interpretation of the system.)」(p.86)と主張している。

3 調査方法

3.1 調査参加者

国内のM大学1年生の488人にWebアンケートに回答するように依頼し、389人が同意して回答した。調査参加者は、3学科(学群)から構成された17クラスの教養英語を受講した学生で、アンケートが実施された2014年前期にクラスは開講された。英語プレイスメント・テスト(英語能力判定テストBタイプ、英語検定協会作成)によりクラスは、受講の前に学科(学群)ごとにレベル別にクラス編成がなされた。公立大学のM大学の学生は、大学入試センター試験英語科目では全国平均点に近い点数を獲得する集団で、日本の大学生の平均的な英語力を有していると考えられた。

調査参加者は、アンケート調査に同意した者に限られた。調査参加者の個人情報が入り込まないように最大限の努力がなされた。

3.2 調査用具

本調査は、日本語で書かれたWebアンケート方式で、CEFR-J Version 1.1の5技能分野(リスニング、リーディング、スピーキングやりとり、スピーキング発表、ライティング)各々12レベル(各技能分野で22)で構成されている総数110のCan-Doディスクリプタを用いた(表3参照)。筆者は、CEFR-J開発チームにEメールで正式に許可を得て、ホームページからCEFR-J Version 1.1 Can-Doディスクリプタをダウンロードして入手した。調査参加者は、それぞれのCan-Doディスクリプタについて自分の英語能力を4肢選択(ほとんどできない、あまりできない、ある程度はできる、ほぼできる)で選択

して回答した。

本調査はM大学総合研究所一般研究に調査計画書を申請し、調査の許可を得て行われた。調査計画書で示されたように、2014年4月初旬にM大学1年生対象に実施された英語能力判定テストBタイプ(英検3級から2級の範囲を測定用、語彙・英文構成・読解50問 リスニング30問、617点満点、55分所用)の試験結果を英語能力の基準として用いた。英語能力判定テストは、2003年から2014年3月までに累計で148万人以上に利用された(英語検定ホームページより)。

3.3 データ収集方法

本調査のアンケート調査を実施する前に筆者の一人が担当している「英語科教育法I」を受講している23名の3年生に対して予備調査が2度実施された。予備調査の回答者からのフィードバックを受けてアンケートの指示文等を修正し、筆者らの一人によりWebアンケートが作成され、2014年4月下旬にウェブ上にアップロードされた。筆者らは10人の1年生英語担当教員に研究プロジェクトへの協力を依頼して、2014年4月下旬から5月下旬にかけて約1か月間、調査参加学生は授業外で携帯電話(QRコード)もしくはコンピュータ(URL)で、CEFR-Jに基づいたアンケートで自分の英語能力を自己評価した。

3.4 データ分析方法

Webアンケート結果はウェブページから収集され、EXCELに保存された。調査参加者の無効な回答データについては除かれた(例えば、一般的には、レベルが上がるにつれ、できる程度は低くなるが、すべてのCan-Doディスクリプタに同じ回答をした6人の調査参加者は削除した)。SPSS Ver.21を用いてアンケート結果と英語能力判定テスト結果の相関関係を統計処理した。さらに、アンケートと英語能力判定テストの関係を検証するために、個々のデータの詳細な質的分析も行われた。

4 結果と分析

4.1 Can-Doディスクリプタ自己評価の5技能分野間の関係

最初に、5技能分野間の関係の信頼性について検証を行った。方法として、統計的分析を行うためにアンケート回答を数値に置き換えた。アンケートの自己評価の回答選択肢は、それぞれ、1=ほとんどできない、2=あまりできない、3=ある程度は出来る、4=ほぼできる、と数値に置き換えた。5技能分野間の相関関係を確かめるために、各技能分野(例:リスニング)の各回答者のCan-Doディスクリプタを次の2段階で自己評価平均値を算出した。第1段階として、回答者ごとに、各技

能分野の平均値を計算した。例えば、英語能力判定テストで得点1位の回答者のリスニングにおける22のCan-Doディスクリプタの自己評価平均値は‘3.64’（最高値：4，最低値：1）である。第2段階として、第1段階で算出した389名の回答者の個々の自己評価平均値を総計し、技能分野ごとに回答者全員の平均値を算出した。例えば、A1.1レベルのライティング質問3に対する389名の平均値は‘3.41’である。その様にして算出された各技能分野の平均値を用いて、全体的な関係を把握するために、下の図1のグラフを作成した。グラフでは、下の表3に示されているように、アンケートの質問に対応して、横軸にCEFR-Jのレベルが示されており、縦軸は回答を数値に置き換えた値である。

図1の見方は、リスニング Q2 (PreA1) は質問2 (PreA1レベル) を指し、平均値は‘2.82’である（紙幅の都合で、個々の数値については省略）。以下、Q2は質問2を指し、省略形を用いる。

図1によると、5技能分野は最高平均値から最低平均値に右肩下がりで直線に近い関係性を形成している。つまり、Can-Doディスクリプタのレベルが高くなるに従って、高い平均値から低い平均値に変化している全体的な傾向がある。例えば、5技能分野のQ3 (A1.1) への回答平均値は4 (ほぼできる) と3 (ある程度はできる) のほぼ中間であり、Q14 (B1.1) への回答平均値は、3 (ある程度はできる) と2 (あまりできない) の中間程度にあり、Q21 (C1) への回答平均値は、2 (あまりできない) と1 (ほとんどできない) のほぼ中間にある。

図1のグラフではやや極端な値もある。例えば、リスニングのQ2 (Pre-A1) の回答平均値は‘2.82’で、この値はレベルの高いQ3回答平均値‘3.46’、Q4回答

平均値‘3.32’、Q5回答平均値‘3.06’、Q6回答平均値‘3.00’の平均値より低い。Q2のCan-Doディスクリプタは、「英語の文字が発音されるのを聞いて、どの文字かわかる」と記述されている。この質問は音声聞いてアルファベット文字が判別できるかを問う平易な能力であるが、本調査の回答者は、音声と文字の規則性を定める音声学やフォニックスの知識・技能について尋ねられたと誤解をしたのではないかと推測する。

前例とは逆の現象で、別の極端値の例はリーディングQ16 (B1.2) の回答平均値‘2.65’で、レベルの低いQ13回答平均値‘2.51’、Q14回答平均値‘2.50’、Q15回答平均値‘2.33’の平均値より高い数値となっている。Q16のディスクリプタは、「平易な英語で書かれた長めの物語の筋を理解する事ができる」と記述している。Q13ディスクリプタの「新聞や雑誌の理解」、Q14ディスクリプタの「ゲームや申込書の理解」、Q15ディスクリプタの「インターネットと参考図書の理解」と比べて、Q16ディスクリプタの「長めの物語の理解」をより易しいと認識している。この不整合が生じた要因として、Runnels (2014) が指摘した様に、回答者が経験した事がなく馴染みのないタスク（「新聞や雑誌の理解」「ゲームや申込書の理解」「インターネットと参考図書の理解」）については低い自己評価をし、学生の多読活動の対象となりがちな馴染みのあるタスク（「物語の筋を理解する」）については、高い自己評価をしたと推測できる。今後の詳細な検証が求められる。

次に、5技能分野間の自己評価における内的信頼性を統計的に処理し、クロンバックの α 係数¹²⁾は、‘0.87’で、5技能分野間の内的信頼性は強い結果を示した。つまり、本研究における5技能分野間の自己評価平均値は、相対

表3 CEFR-Jレベルと質問の対応

| レベル： | PreA1 | A1.1 | A1.2 | A1.3 | A2.1 | A2.2 | B1.1 | B1.2 | B2.1 | B2.2 | C1 | C2 |
|------|-------|------|------|------|------|-------|-------|-------|-------|-------|----|----|
| 質問： | 1&2 | 3&4 | 5&6 | 7&8 | 9&10 | 11&12 | 13&14 | 15&16 | 17&18 | 19&20 | 21 | 22 |

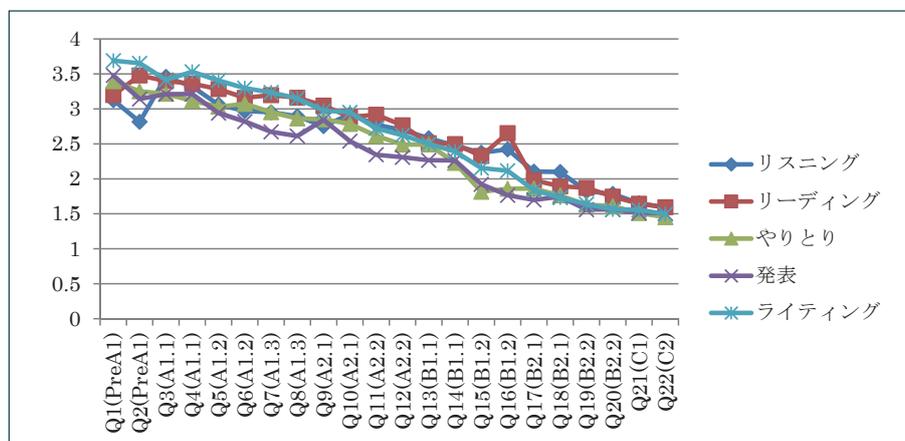


図1 5技能分野（各22問）における回答者全員の自己評価平均値の関係（N=389）

関係が強いという結果を得た。図1のグラフで示されたように、個々の回答者の自己評価平均値において5技能分野間がお互いに同様な規則性、つまり相関関係が強い事を統計的な分析も支持する結果となった。

しかし、上のグラフや統計的な分析は、調査参加者全員の平均値について5技能分野の自己評価の傾向を検証した結果であり、個人レベルでの信頼性を検証した結果ではない。以下に個人単位での自己評価の信頼性についても検証する。

4.2 Can-Doディスクリプタと英語能力判定テストの関係

この節では5技能分野Can-Doディスクリプタの自己評価と各回答者の英語能力の相互関係について分析する。自己評価回答と英語能力判定テスト点数の相関関係を統計的に処理するために、ピアソン相関係数 r （両側検定、 $\alpha=.001$ ）を算出した。表4で示されているように各技能分野のピアソン r 係数の値は‘0.3’前後で英語能力判定テストと各技能分野の相関係数は低い傾向にある。つまり、英語能力判定テストの得点と各技能分野の自己評価の相関関係は弱いという結果を示した。英語能力判定テスト（Bタイプ）はリスニング力とリーディング力を測定するテストだが、その両方の英語技能とCan-Doディスクリプタ自己評価が特に相関関係が強い訳でもない。

表4 5技能分野の自己評価と英語能力判定テストの相関係数

| | リスニング | リーディング | やりとり | 発表 | ライティング |
|------------|-------|--------|------|------|--------|
| ピアソン係数 r | .271 | .292 | .251 | .312 | .292 |

上では統計処理を行ったが、以下では個々のデータを質的に精査することで、5技能分野の自己評価と英語能力判定テストの関係について検証する。紙幅の都合で、自己評価アンケートの中からリスニングのCan-Doディスクリプタ（22問）のみについて抽出された10名の回答を分析する。抽出された10名は、英語能力判定テストの点数により順位化（調査参加者内での順位）されたローデータから各40位ごと（1位、40位.....360位）を全データから抽出した（表5を参照）。下の表では、リスニングの質問の番号が大きくなるにつれレベルが高くなるので、相関関係が正常であるならば回答はQ1では4（ほぼできる）から始まり、徐々に数字が小さくなり、Q22では1（ほとんどできない）に近い数値で終わることになる。本調査では、North（2014, p.103）の示した「関連したレベルについてチェックリストのディスクリプタを学習者が80%満たした時にそのレベルにあると考える（When a learner met 80% of the descriptors

on the checklist for the level concerned, they could be considered to ‘be’ that level.）」という基準に従い、そのレベルにある基準値を‘4’（ほぼできる）とし、‘4’の数値が回答にない場合には、‘3’で回答された最もレベルの低いディスクリプタを該当のレベルとし、太字で該当するレベルを囲った。しかし、英語能力判定テスト点数1位の回答者のように、Q17で‘4’から‘3’に明確に分かれている場合は判断しやすいが、他の抽出された回答者のデータについては、判断が困難である。例えば、80位、160位、200位の回答者については、すべてのディスクリプタについて‘4’の回答がなくレベルを決定する事が難しい。また、Q1からQ12の範囲において40位、120位、240位、280位、320位、360位の回答者の回答は、レベルの難易度に合わない形で‘4’と‘3’が規則性を持たずに交互に現れており、ディスクリプタのレベルと回答にばらつきがある。つまり、英語能力判定テスト順位1位の回答者を除いては、ディスクリプタのレベルと自己評価の回答に整合性がない事を表している。

さらに、抽出されたCan-Doディスクリプタの平均値を比較すると、英語能力テストの順位とは異なる数値がある。例えば、240位の平均値‘3.24’と320位の平均値‘2.77’は、英語能力判定テストで上位の回答者の平均値より高い数値となっており、英語能力判定テストの結果とCan-Doディスクリプタの回答に強い整合性がない事を示している。

表5 リスニングのCan-Doディスクリプタ回答ローデータ（英語能力判定テスト順位から40位ごとに抽出されたデータ。太字で囲まれた部分は各回答者のレベル）

| レベル 順位 | Pre-A1 | | A1.1 | | A1.2 | | A1.3 | | A2.1 | | A2.2 | |
|-----------|----------|----|------|----------|------|----------|------|----|------|-----|------|----------|
| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 |
| 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 40 | 3 | 2 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 80 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 |
| 120 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 |
| 160 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 3 |
| 200 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 |
| 240 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 4 |
| 280 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 |
| 320 | 3 | 3 | 3 | 4 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 3 |
| 360 | 3 | 2 | 4 | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 3 | 3 |

(表5の続き)

| レベル 順位 | B1.1 | | B1.2 | | B2.1 | | B2.2 | | C1 | C2 | 平均 |
|-----------|------|-----|------|-----|------|-----|------|-----|-----|-----|------|
| | Q13 | Q14 | Q15 | Q16 | Q17 | Q18 | Q19 | Q20 | Q21 | Q22 | |
| 1 | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 2 | 2 | 2 | 3.62 |
| 40 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2.32 |
| 80 | 3 | 2 | 3 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2.43 |
| 120 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 2.86 |
| 160 | 2 | 2 | 2 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 2.27 |
| 200 | 2 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2.5 |
| 240 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 3.24 |
| 280 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2.23 |
| 320 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2.77 |
| 360 | 2 | 1 | 3 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2.18 |

Can-Doディスクリプタと英語能力判定テストの関係について結果をまとめると、統計的分析では英語能力判定テスト結果と各技能分野の相関関係が弱い事を示し、質的な分析においても回答者のCan-Doディスクリプタの平均値が英語能力判定テストの順位と強い整合性がない事を示した。さらに、個々の回答者においてもCan-Doディスクリプタの難易度に準じて、回答の数値に強い規則性を見つけ出すことが困難であり、個々の自己評価回答にばらつきがあった。

英語能力判定テストと整合性が弱く、自己評価回答にばらつきがあった要因として、Jones (2002) が指摘した様に、英語能力や学習目標などの様々な個人の背景により英語能力の自己評価は異なるのではないかと推測される。

5 結論と示唆

本稿前半では、日本の外部試験や英語テキストに急速に利用が広がっているCEFRとその日本版であるCEFR-Jの開発の経過やその強みと批判について文献研究によって、検証した。特に文献研究から日本の大学英語教育に示唆となる点は、CEFRを試験作成の理論的枠組みにするには困難点が多い事で、今後予定されている大学入試に代わる4技能測定の外部試験の共通参照枠にするには十分な検討が必要である。さらに示唆となる点は、Can-Doディスクリプタの自己評価結果の信頼性が高くないという先行研究の結果である。今後、小学校、中学校、高等学校にCan-Doリストを作成する事が文部科学省から義務付けられており、大学においても導入が

予想されるが、英語能力の自己評価を達成度指標として用いるには慎重な対応が求められる。

本稿の後半では、M大学教養英語を受講する学生を対象として、CEFR-JのCan-Doディスクリプタについて信頼性と英語能力テストとの相関関係について検証した。Can-Doディスクリプタについて、調査者全員の自己評価平均値は5技能分野で強い内的信頼性の傾向を示した。しかし、個人単位で質的に詳細に分析した結果は、Can-Doディスクリプタについて個人の回答結果がCEFR-Jのレベルと整合性がなかったり、英語能力判定テストの順位と個人の自己評価に強い整合性が見られない結果を示した。英語能力判定テストとCan-Doディスクリプタ自己評価の相関関係については、統計的な分析も同様に弱い相関関係係数を示した。

結論として、本研究の実証的なデータは、先行研究と同様にCan-Doディスクリプタの自己評価は英語能力試験との整合性が弱く、Can-Doディスクリプタの自己評価は個人によりばらつきがあるという、先行研究を支持する結果を示した事になる。

今後の調査では英語能力判定テスト（英語検定協会の作成）以外の英語標準能力テストを用いてCEFR-Jとの整合性や相関関係について調査する必要がある。CEFR-Jが信頼性の高い共通参照枠であるかどうかは、本研究のみで断定する事はできない。仮に自己評価の信頼性について疑問があるならば、英語教師や評価専門家によるCEFR-Jを用いた他者評価も検討されるべきだが、評価者間の測定値の信頼性など多くの課題を含んでいると推察される。CEFRは元来、ヨーロッパのコンテキストで異なる言語間に利用できる、汎用性を持つ共通参照枠として作成された経緯を考慮すれば、多くの研究者が主張しているように、絶対的な枠組みではなく各言語プログラムに適応できる基礎となる参照枠として取り扱うべきかもしれない。

英語能力の国際基準の到達が喫緊の課題として求められる日本の状況下で、英語教育は確固たる評価基準やカリキュラム、学習到達目標をどのように確立するのか、実証的な検証によってCEFRおよびCEFR-Jの利用が今後進むことが期待される。

注

- 1) 米国試験団体ETSによって作成されるTOEFL iBTは4技能の英語能力をコンピュータを用いて解答し、結果は主に米国の大学入学の英語能力基準として用いられる。
- 2) IELTSは得点ではなく、バンドスケールの0（非受験者）から9（エキスパートユーザー）で評価される。IELTSはケンブリッジESOLやブリティッシュ・カウ

- ンシルなどに共同運営されている。TOEFL iBTと同様に4技能を測定する。英国やオーストラリアの大学の英語能力基準として用いられる。
- 3) GTEC CBTは、ベネッセ社によって開発され英語4技能を測定する1400点満点のテストで、コンピュータで受験し、筑波大学や九州大学は大学入試に活用している。
 - 4) TEAP は、英語検定協会と上智大学によって共同開発された4技能を測定するテストでスコアとCEFRレベルで結果は示される。上智大学の他に、多くの私立大学が大学入試判定試験として採用を予定している。
 - 5) CEFRは、多くの言語に応用されているが英語に特化した研究は、English Profile Programme (EPP) と呼ばれる。
 - 6) アルク社は、次世代Eラーニング教材「ALC Net Academy Next」で2016年2月にリリースする4コースにCEFR-Jを活用することをホームページで発表している。
 - 7) 具体的には、ヨーロッパではすべての学校で2か国以上の外国語を学習する事を義務付けており、移民らが自国の文化や言語を保持する多言語・多文化主義とは異なる。
 - 8) Council of EuropeはCEFRを実践化するために、European Language Portfolio (欧州言語ポートフォリオ, ELP) のプログラムでCEFRの自己評価表の他に、Language Portfolio (資格や能力の記録)、Dossier (課題等の記録) を利用している。CEFR-Jには、それらに相当するプログラムはない。
 - 9) ケンブリッジ英語検定 (KET, PET, FCE, CAE, CPE) とCEFRの対応は、A1レベルはKET (45-59), A2レベルはPET (45-59) / KET Pass with Merit, Pass, B1レベルはFCE (45-69) / PET Pass with Merit, Pass / KET Pass with Distinction, B2レベルはCAE (45-59) / FCE grade B or C / PET Pass with Distinction, C1レベルは、CPE (45-59) / CAE grade B or C / FCE grade A, C2レベルはCPE grade A, B or C / CAE grade A となっている。(出所: https://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages)
 - 10) Sugitani & Tomita (2012) によると、小池生夫科研 (2004年~2007年) は、「第二言語習得研究を基盤とする小、中、高、大の連携をはかる英語教育の先導的基礎研究」という研究課題で第2言語習得研究と学校英語教育の関係について研究を予定していたが、古島と大橋により2004年にCEFRの訳本 (朝日出版社) が出版されると研究の中心を第2言語習得研究からCEFR研究に変更した。小池科研を引き継いだのは投

野科研で2008年に研究プロジェクトを開始した。

- 11) Super English Language high schoolはSELhiとも呼ばれ、文部科学省により指定された高等学校で、理科や数学などの英語以外の科目を英語で教授する実践などを行った。
- 12) クロンバックの α 係数 (Cronbach's alpha) は、同一アンケートの質問や同一試験の問題の信頼性を表す値で、他のアンケートや試験との比較をしないで含まれる質問等の信頼が表せる。

謝辞

本研究の自己評価アンケートの項目となっているCEFR-Jを提供して下さったCEFR-J研究・開発チームに深く感謝する。執筆者の一人であるNorman Fewell 上級準教授がWebアンケートを作成してくれた事と木村堅一教授の統計処理への協力は研究を大きく前進させた。最後に一般研究の事務的処理や助言をして下さった名桜大学総合研究所の教職員に感謝申し上げます。

引用文献

- Alderson, J.C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91, 659-663.
- ALTE. (2002). *The ALTE CAN DO PROJECT*. Retrieved on March 30, 2015 from (http://www.alte.org/attachments/files/alte_cando.pdf)
- Council of Europe. (2001). *The Common European Framework of References for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- ETS. (2014). *Test and score data summary for TEOFL iBT tests (2013, January-2013 December tests)*. Retrieved on March 14, 2015 from (http://www.ets.org/s/toefl/pdf/94227_unlweb.pdf)
- Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66 (4), 477-485.
- Hulstijn, J.H. (2007). The shaky ground beneath the CEFR: quantitative and quantitative dimensions of language proficiency. *The Modern Language Journal*, 91, 663-667.
- Jones, N. (2002). Relating the ALTE framework to the Common European Framework of Reference. *THE ALTE CAN DO PROJECT*. Retrieved on March 27, 2015 from (http://www.alte.org/attachments/files/alte_cando.pdf)

- Little, D. (2010). The European language portfolio and self-assessment: Using "I can" checklists to plan, monitor and evaluate language learning. In Schmidt, G.S., Naganuma, N., O'Dwyer, F., Alexander, E. and Kazuni, S. (Eds.) *Can do statements in language education in Japan and beyond -Applications of the CEFR* (pp. 157-166). Tokyo: Asahi Press.
- Negishi, M., Takada, T. & Tono, Y. (2012). A progress report on the development of the CEFR-J. *Studies in Language Testing* 36: 137-157.
- North, B., Ortega, A., & Sheehan, S. (2010). *A core inventory for general English, British Council/EAQUALS*. Retrieved April 20, 2014 from (<http://www.teachingenglish.org.uk/article/british-council-equals-core-inventory-general-english-0>)
- North, B. (2014). *The CEFR in practice*. Cambridge: Cambridge University Press.
- Runnels, J. (2013). Examining the difficulty pathways of can-do statements from a localized version of the CEFR. *Journal of Applied Research on the English language*, 2 (1), 25-32.
- Runnels, J. (2014). An Exploratory Reliability and Content Analysis of the CEFR-Japan's A-Level Can-Do statements. *JALT Journal*, Vol.36, No.1, 69-89.
- Sugitani, M. & Tomita, Y. (2012). Perspectives from Japan. In Byram, M. & Parmenter, L. (Eds.) *The Common European Framework of Reference: The Globalization of Language Education and Policy* (pp. 197-211). Bristol: Multilingual Matters.
- van Ek, J.A. (1975). *The Threshold Level*. Strasbourg: Council of Europe.
- van Ek, J.A. & Trim, JLM. (2001a). *Waystage*. Cambridge: Cambridge University Press.
- van Ek, J.A. & Trim, JLM. (2001b). *The threshold level*. Cambridge: Cambridge University Press.
- van Ek, J.A. & Trim, JLM. (2001c). *Vantage*. Cambridge: Cambridge University Press.
- Weir, C.J. (2005). Limitations of the Common European Framework for developing comparative examinations and tests. *Language Testing*, 22 (3), 281-300.
- Wilkins, D.A. (1976). *National syllabuses*. Oxford: Oxford University Press.
- 川成美香科研 (2013) 『外国語コミュニケーション能力育成のための日本型CEFRの開発と妥当性の検証』 https://kaken.nii.ac.jp/pdf/2012/seika/C-19_1/32404/22320108seika.pdf (2015年9月6日取得)
- 小池生夫科研 (2008) 『第二言語習得研究を基盤とする小, 中, 高, 大の連携をはかる英語教育の先導的基礎研究』 http://www.kantei.go.jp/jp/singi/kyouiku_kondan/kaisai/dai3/2seku/2s-siryou_2.pdf#search=%E5%B0%8F%E6%B1%A0%E7%94%9F%E5%A4%AB%E7%A7%91%E7%A0%94+%E7%AC%AC2%E5%A4%96%E5%9B%BD%E8%AA%9E%E7%BF%92%E5%BE%97%E7%A0%94%E7%A9%B6 (2015年9月6日取得)
- 文部科学省 (2014) 『CEFR-Jの取組』 www.mext.go.jp/b_menu/shingi/chousa/shotou/.../1343401_03.pdf (2015年8月30日取得)
- 文部科学省 (2015a) 『英語力評価及び入学者選抜における英語の資格・検定試験の活用促進に関する作業部会基礎資料』 http://www.mext.go.jp/b_menu/shingi/chousa/shotou/106/shiryo/_icsFiles/afielddfile/2015/03/25/1356067_04.pdf (2015年8月30日取得)
- 文部科学省 (2015b) 『生徒の英語力向上推進プラン』 http://www.mext.go.jp/b_menu/shingi/chukyochukyos/053/siryochukyos/_icsFiles/afielddfile/2015/08/04/1360076_8.pdf (2015年8月30日取得)
- 中島正剛・永田真代. (2006) 「CEFRの日本人外国語学習者への適用可能性」 『外国語研究学会』 No.9, pp.8-24.
- 投野由紀夫ほか (2012) *CEFR-J Ver.1.1*. <http://www.cefr-j.org/download.html>. (2014年4月10日取得)
- 投野由紀夫科研 (2012) 『小, 中, 高, 大の一貫する英語コミュニケーション能力の到達基準の作成とその検証』 <http://www.cefr-j.org/TonoKaken2008-2011FinalReport.pdf> (2015年9月6日取得)
- 投野由紀夫編 (2013) 『CEFR-Jガイドブック』 東京: 大修館書店.
- 渡慶次正則 (2013) 「TOEFL iBTにおけるスピーキング測定とライティング測定の妥当性, 信頼性, 実現性の検証」 『名桜大学紀要』 No.19, pp. 65-76.
- 吉島茂・大島理枝訳 (2004) 『外国語の学習, 教授, 評価のためのヨーロッパ共通参照枠 (Common European Framework of Reference for Languages)』 東京: 朝日出版社.