

論文の和文要旨	
論文題目	Development and Validation of CAN-DO Descriptors for Students Majoring in Engineering in Japanese Higher Education Institutions
氏名	Mayu KASHIMURA
本論研究は、実証的なニーズ分析に基づき、工学を専攻する日本の高等教育機関で学ぶ学生達のための英語能力を評価する CAN-DO ディスクリプタを開発し、その妥当性を検証することを目的としている。工学を専攻とする学生達のニーズ分析に基づき、彼らが実際に行う頻度の高い作業に焦点を当てることで、彼らが学ぶ高等教育機関における英語の授業のカリキュラムにおける英語運用能力の枠組みや評価と、工学を専攻とする学生たちの実際のコミュニケーションや研究上必要となる英語のニーズとの格差を埋めることを目指している。 もう少し具体化すると、以下ようになる。工学を専攻する日本の高等教育機関で学ぶ学生のニーズ分析において、実施頻度の高かったタスクを洗い出し、その情報を元に、4技能5領域、つまり、聴くこと、読むこと、書くこと、話すこと（発表）、話すこと（やりとり）それぞれに分けて CAN-DO ディスクリプタを開発している。そして、それに続く妥当性の検証研究では、項目応答理論の1つで、心理計量学の手法であるラッシュモデルを用いて、開発した CAN-DO ディスクリプタが、工学を専攻とする学生達の英語熟度を区別する際の信頼性と妥当性を確認することを行っている。 本研究の第1章では、本研究の主旨と主な流れのほか、本研究の着想に至った経緯として、著者が過去に行った調査研究が紹介されている。檜村(2013)は、Gardner (2004)を用いて、海外インターンシップ及び研究留学が工学を専攻する学生たちにもたらす英語使用・英語学習への情緒面での変化として、英語を使用することへの不安が減るということを明らかにし、TOEIC®形式のリスニング、リーディング問題の事前事後の正答率の変化から、海外インターンシップ・研究留学が、リスニング、リーディングの英語運用能力を引き上げる効果を持つ可能性を国内インターンシップに行ったグループとの比較で明らかにした。また、著者独自の質問項目から、海外経験からの帰国後、英語を使いたい・学びたいという意欲は高まり、海外経験が動機付けになり、帰国後も意欲は高いままであるが、実際の英語学習や使用に結びつかないということを明らかにした。これらのことから、海外経験等の事前・事後に、自分の英語運用能力を位置づけことができる指標が必要と思われた。 第2章は、本研究テーマに関連する English for Specific Purposes (ESP)、これまでの工学専攻の学生の英語のニーズ調査、学習者による自己評価、CAN-DO ディスクリプタ及び工学専攻の学生のための CAN-DO ディスクリプタ開発の先行研究、テストや尺度の妥当性検証の研究に触れている。また、<i>The Common European Framework of Reference</i>	

for Languages (2001)の social agent としての言語使用者、複言語主義、伝統的な 4 技能で言語使用を捉えることからの脱却、European Language Portfolio といった重要な概念や言語運用能力テストやカリキュラムをどのように準拠することができるのかについてもまとめている。

第 3 章において、本研究で行ったニーズ分析について詳しく述べている。ニーズ分析は、全部で 3 つの調査から構成されている。最初の 2 つ(NAS 1, NAS 2)は、工学を専攻する学生達が参加する海外インターンシップ、研究留学および、所属機関における英語の授業以外での英語を使用する作業について実証的な調査を行うことで、工学を専攻とする学生達が海外経験の中で、どのようなタスクを何の言語を使って行っているのかを浮き彫りにすると共に、NAS 2 は、調査協力者の工学専攻の学生が行う頻度の高いタスクを可視化した。また、調査協力者のタスクの自己評価と自己評価が低かったタスクの主な原因も尋ね、タスクを実施するのに十分な言語運用能力を持ち合わせていないと、タスクの達成度が低くなるという傾向も導き出した。NAS 3 においては、工学専攻の学生が英語の授業以外の学生生活において行う英語を使った活動を明らかにした。これら 3 つのニーズ分析から、工学専攻の学生が行う頻度の高いタスクをまとめることで、CAN-DO ディスクリプタ作成のための準備を行った。

第 4 章においては、第 3 章で得た結果を元に、技能ごとに、工学専攻の学生のための CAN-DO ディスクリプタを開発した。開発した CAN-DO ディスクリプタの構成は、Listening が 36 項目、Reading が 32 項目、Writing が 36 個、Spoken Production が 31 個、Spoken Interaction が 38 個、合計 173 個である。Listening と Reading については、国際ビジネスコミュニケーション協会との共同研究の中で作成した。Writing、Spoken Production、Spoken Interaction については、Global Scale of English Learning Objectives for Academic English や Global Scale of English Learning Objectives for Professional English、国際ビジネスコミュニケーション協会の開発した CAN-DO ディスクリプタといった既存の CAN-DO ディスクリプタを参考にしながら作成した。開発した CAN-DO ディスクリプタは、海外での滞在において必要となるサバイバル・イングリッシュや現地の人と親しくなる際に起こるもの、新聞を読むことやテレビを見ること等といった日常的かつサバイバル・イングリッシュよりは必要となる英語運用能力もレベルが高くなるもの、学校生活や授業において必要となるもの、そして、工学に関連した活動(研究、海外インターンシップでの就労体験)において必要となるものというカテゴリーに分けた。しかしながら、一部の項目は、2 つ以上のカテゴリーに関連するものもあり、J. D. Brown (2016)で言及されている English for Academic Purposes と English for Science and Technology の重なりをまさに示すかたちとなった。

第 5 章は、開発した CAN-DO ディスクリプタの妥当性の検証を扱う章である。妥当性検証の調査は、日本の高等教育機関(高等専門学校、大学、大学院)で工学を専攻している学生達を対象に行った。調査協力者(総計 1,990 名)は、各技能・領域の開発した CAN-DO ディスクリプタの合計 173 項目の CAN-DO ディスクリプタに対して、英語で行うとしたらどの程度できるかについて、5 件法のリッカート形式(1=全くできない、2=ほとんどできない、3=かろうじてできる、4=ほとんどできない、5=問題なくできる)で自己評価を行った。そのほか、学年、専攻の情報、TOEIC® Program Test を自己評価調査実施時点から半年以内に受験した者については、受験した TOEIC®テストのスコアも回答

した。

収集された自己評価及び TOEIC® Program Test のスコアのデータは、CEFR のディスクリプタの妥当性検証研究である North (2000)、CEFR-J のディスクリプタの妥当性検証研究の報告もしている投野 (2013)、Pearson Public Limited Company による Global Scale of English Learning Objectives の妥当性検証研究である de Jong et. al (2016)等で用いられている計測心理学の分析方法であるラッシュモデルによる分析を行った。実際の分析には、統計ソフト WINSTEPS (Version 5.2.5.0)を用い、一次元性が担保されているかを解析するための値、各ディスクリプタの項目困難度、フィット値、調査協力者個人の能力値を求めた。

これらラッシュモデルでの分析で導き出された値は、CAN-DO ディスクリプタの妥当性の検証を行う際に、Chapelle et al. (2008) and Chapelle and Lee (2021)に示されている論証に基づく妥当性検証(argument-based validation)の手法を用いた。この論証方法では、推論(Inference)を支える論拠(Warrant)を明確にすることで、論拠を積み重ねていくことで、妥当性を高めていく。第 5 章のリサーチ・クエスチョンと直接的に関連がある Inference は Evaluation Inference, Generalization Inference, Explanation Inference, Extrapolation Inference である。これらの Inference に対して、それぞれ Warrant を設定した。

評価推論(Evaluation Inference)には、「各言語スキルに対して作成された CAN-DO ディスクリプタは信頼性が高く、英語運用能力において異なる熟達度レベルの学生を効果的に区別することができる。」を論拠として示した。この推論を支える前提条件として、以下の 3 つを設定し、これらは第 5 章のリサーチ・クエスチョンの一部と合致する。(a) 収集したデータは、ラッシュモデルの適用条件である一次元性を保持している。(b) 各言語スキルに対して作成された CAN-DO ディスクリプタは、非常に高い信頼性を示した。(c) 各言語スキルに対して作成された CAN-DO ディスクリプタは、ラッシュモデルでの分析の結果、異なる熟達度レベルの学生を効果的に区別できることが示された。(a)に関しては、各スキルの収集データにおいて、標準的な手法である Eigenvalue の値では、いずれのスキルにおいても、2.0 以上となり、一次元性が担保されない結果になったが、PTMEA 相関の値が各技能内でどの項目においても高く (Listening: 0.38-0.82, Reading: 0.57-0.85, Writing: 0.57-0.85, Spoken Production: 0.56-0.81, Spoken Interaction: 0.62-0.83)、この結果をもって一次元性が担保されていることが証明された。(b)に関しては、信頼性係数の値が、いずれの技能においても 0.97 以上で高い値を示したことから、証明された。(c)を論証する際には、person reliability と item reliability の値を用いた。Bond et al. (2021) によれば、高い person reliability は、テストの実施によって、受験者のスコアが低いものから高いものまで広く分布していることを示し、高い item reliability は、ラッシュモデルにおいて、尺度上の項目が易しいものから難しいものまで広く分布していることを示す。各技能の CAN-DO ディスクリプタ群は、いずれの値も高く、異なる熟達度レベルの学生を効果的に区別できることが示された。

一般化推論(Generalization Inference)では、「作成された CAN-DO ディスクリプタは、妥当性検証に参加した対象者および他のサンプルにも適切である。」を Warrant として設定し、以下の2つを前提条件とした。(a) 作成された CAN-DO ディスクリプタは、項目の難易度の観点から、参加者の熟達度レベルをカバーできる。(b) 作成された CAN-DO ディスクリプタは、他のサンプルに対しても適切である。(a)は先に述べた評価推論(Evaluation Inference)の論拠の前提条件(b)と同じであるため、論証されており、(b)はラッシュモデルで出力される Item-Person map で確認し、多くの CAN-DO ディスクリプタの項目困難度が調査協力者たちの個人能力値をカバーしていることが確認できたが、高い個人能力値を持つ調査協力者のための項目が一部ないことがわかった。

説明推論(Explanation Inference)では、論拠として、「作成された CAN-DO ディスクリプタの項目難易度は、大規模調査を基に作成された CAN-DO ディスクリプタの難易度と近い。」ということを設定し、作成された CAN-DO ディスクリプタの項目難易度は、CEFR における類似の CAN-DO ディスクリプタの CEFR レベルと対応しているということを前提条件とした。ラッシュモデルを用いた分析で示された各 CAN-DO ディスクリプタの項目困難度と CEFR のレベルの整合性がうまくいった項目も多かったが、うまくいかなかったものもあった。

外挿推論(Extrapolation Inference)は、「作成された CAN-DO ディスクリプタで評価される英語熟達度の構成概念は、他の英語運用能力試験で評価される英語熟達度を説明しうる。」を論拠とし、以下、2つを前提条件とした。(a) 工学系学生が自己評価で回答した CAN-DO ディスクリプタへの反応は、Rasch モデルの期待に適合していることが示された。(b) 作成された CAN-DO ディスクリプタによる自己評価結果(Rasch 分析から得られた個人能力値)は、英語運用能力試験の成績と正の関連を示した。(a)はラッシュモデルで出力されるフィット値を使用し、論証を試みた。その結果、どの技能の CAN-DO ディスクリプタ群においても、項目困難度が最も低い項目の幾つかがラッシュモデルに適合していないことがわかり、一部の項目は CAN-DO リストから除外した方が良いことが統計解析では示された。また、(b)については、TOEIC Bridge™ を受験した調査協力者の Listening 及び Reading のテストスコアと個人能力値はほぼ無相関であることが示され、TOEIC® のテストスコアと個人能力値の相関は、各技能において、中程度の相関があることがわかった(Listening: 0.46, Reading: 0.50, Writing: 0.51, Spoken Interaction: 0.56, Spoken Production: 0.50)。この結果には、TOEIC Bridge™が CEFR B1 レベルまでしか測定できないことや、TOEIC Bridge™を受験し、本調査に参加した学生の多くが、専門知識を用いて行うタスクをまだあまりやっただけで高専低学年の学生が多かったことが要因として考えられる。

工学を専攻する学生達のニーズ分析を元に CAN-DO ディスクリプタの開発と妥当性検証までの研究を行い、教育的示唆として得たもののうち、重要なものとして以下を挙げる。まず、作成した CAN-DO ディスクリプタを CEFR と整合させるためには、作成段階

において各ディスクリプタに詳細な条件や基準を明記することが重要である。ここでいう「条件」とは、タスクが実施される文脈を指し、「基準」とは期待される成果を示すものである。これらの要素をディスクリプタに明確に示すことで、教師や研究者は、そのディスクリプタが各 CEFR レベルにおけるパフォーマンス期待を正確に示していることを確認できる。

次に、CEFR レベルとの整合性を高めるためには、各熟達度レベルに対して「アンカー項目」を設定することも重要である。「アンカー項目」とは、基準点となる項目のことであり、ディスクリプタ同士が異なる言語レベル間で比較可能かつ一貫性を保つための指標となる。このように CEFR と体系的に整合させることで、ディスクリプタの妥当性と信頼性が向上し、さらにより広い言語教育の枠組みに統合しやすくなる。

さらに、ディスクリプタを妥当性検証のためだけでなく、ポートフォリオの一部として主に活用することを想定する場合、今回の妥当性検証研究で Pre-A1 レベルと判断された最も基礎的なディスクリプタも保持することが重要である。これにより、特に技術系高専の1年生など、熟達度が下位に位置する学生にとっても、自分に合った CAN-DO ディスクリプタが含まれていることが保証される。特に、CEFR A1 レベルの基本的なタスクすら習得に苦勞している学生にとっては、このような初歩レベルのディスクリプタが含まれていることで、学習から取り残されているという感覚を防ぐことができる。低いレベルであっても成功体験を積むことができれば、英語教師は学生の自信や自己効力感を高め、英語学習への動機づけを促すことができる。これにより、より前向きで持続可能な学習体験の実現につながるであろう。