

項目分析と古典的テスト理論に基づく 「OIDAI 英語力診断テスト」の研究開発

松 宮 新 吾

Validation and Implementation of the OIDAI Diagnostic Test of English Through Classical Test Theory

Shingo MATSUMIYA, Faculty of International Liberal Arts

要 旨

This research aimed to analyze multiple-choice test items of a researcher-made test, the OIDAI Diagnostic Test of English Proficiency. Analysis is based on Classical Test Theory, and it involves item difficulty and item discrimination as well as reliability and its validity. The researcher developed and designed a Beta-version and implemented the test to fifty students in the Department of International Liberal Arts in March 2021. Item analysis of collected data indicated significant reliability and validity. By eliminating items with lower levels of discrimination power and item difficulty, a new diagnostic test was edited. The revised version of the diagnostic test had 67 question items and was given to 1,685 first-year students at OGU in April. Item analysis of date with the revised version was conducted. The findings show that 61 items fell within an acceptable and within a satisfactory range of reliability and validity level, while 6 items were rejected due to their poor discrimination index. This research concludes with some limitations of Classical Test Theory and its practical application to a researcher-made diagnostic test.

【キーワード】 古典的テスト理論、項目分析、カリキュラム評価、英語運用能力診断テスト

1. はじめに

2021年度から、追手門学院大学外国語教育室が所管する英語科目が、新カリキュラムの下、従来の General Reading と Oral English に替わり、「総合英語」及び「Online English Seminar」とい

う新科目として展開されることとなった。

これに伴い、「総合英語」は、取り扱う題材とインプットとアウトプットに関わる言語活動とを有機的にリンクさせ統合させた「コンテンツ科目」と、「スキル科目」の両方の特性を併せ持つ新科目としてデザインされた。そのために、学生が所属する学部・学科の基礎的な専門科目内容と、総合英語で取り扱う題材や言語材料等とを関連付けしながら、今日の社会問題や人類が取り組むべき重要課題として国連が提唱しているSDGsに関わるトピックスを取り扱うこととした。また、CBI（Content-based Teaching/Instruction：内容重視型教授）やCLIL（Content and Language Integrated Learning：内容言語統合型学習）に関わる教授法と、タスク中心教授法（TBLT: Task-based Language Teaching/Learning）とを採用することにより、コンテンツ科目としての認知・理解学習と、言語スキルに焦点を当てた四技能（聞くこと、話すこと、読むこと、書くこと）の習得・習熟を図る言語活動とを統合させることとした。

新カリキュラムによる英語教育効果を最適化し教育成果を最大化するためには、新カリキュラムによる各科目の履修前の学生の英語運用能力を的確に把握するとともに、一定の履修期間を経た後の教育効果を測定し分析することが求められる。

そこで、本研究においては、各授業での教育効果を高めるため、履修前（4月）の学生の英語運用能力を測定し、学生の適正配置を可能とするクラス編成を行う「診断テスト1」と、春学期末（7月）と秋学期末（12月）における学生の英語学習の目標到達度を測定するための「診断テスト2、3」を研究開発し実施することとした。

なお、本論文においては、4月期に実施したクラス編成を目的としたプレースメントテストと、入学時の学生の英語運用能力を把握し、学生自らが英語学習到達目標を設定するための診断テストの2つの機能を有する「入学前OIDAI英語力診断テスト（診断テスト1）」と、春学期末における学生の英語学習目標の到達度を測定し、秋学期末の英語学習の到達目標を設定するための「春学期末OIDAI英語力診断テスト（診断テスト2）」の分析結果に基づき考察を行うこととする。

2. 研究の背景

2.1 新カリキュラムについて

これまで外国語教育室では、履修生の英語学習に関する背景情報を得るとともに、教授学習方略を最適化するための基礎データを収集する手段として、入学時に「英語学習実態調査」を実施してきた。また、履修生の英語学習に対する意識の変容を確認するために、各学期末に同様の学習実態調査を実施し分析してきた。

その結果、学習者要因としての適性や経験に関わる因子を抽出することができ、英語学習有

能因子や能力得点との因果関係を明らかにすることができている。しかし、大学での初年次の英語学習を通じて、英語学習有能因子得点やグローバル志向因子得点が期待される以上に有意に高まる傾向は確認できず、2年次の選択履修科目として設定されている「Advanced English」の履修者人数は7%程度にとどまっている。一方、1年次の必修英語科目の単位習得ができず、再履修となる初年次生は20%にも上っている。

そこで、大学の初年次教育における英語学習に対するモチベーションを高め、それを2年次以降へ継続的に発展させることができ、就職活動等でアピールすることができるTOEIC等のグローバルスタンダードに準拠した英語運用能力の指標を得ることができるよう、初年次英語科目のカリキュラムをリデザインすることとした。

特に、「英語ざらい」と形容されている英語有能感やグローバル意識、異文化経験値が低い学生が多数を占める学習者集団において、英語学習の有用性と自己有能感を継続的・系統的に高めるための方策として、内容面と教授学習法の両面から、新カリキュラムの英語科目を以下のとおり設定した。

(1) 「総合英語」の学習内容及び題材について

「総合英語」では、「1. はじめに」で記載したとおり、SDGsに関わる教材を取り扱うこととした。SDGsとは、持続可能な開発目標（Sustainable Development Goals）のことで、2001年に策定されたMDGs（ミレニアム開発目標：Millennium Development Goals）の後継として、2015年9月の国連サミットで採択されたものである。SDGsには、2030年までに持続可能でよりよい世界を実現するために必要な17のゴールと169のターゲットを含む国際目標が設定されている。特に、SDGsでは、地球上の「誰一人取り残さない（leave no one behind）」という信念が貫かれ、発展途上国のみならず、先進国自身が取り組むべき「普遍的な」課題が設定されており、大学生の日常生活や社会生活全てに関連する豊かな題材が提供されている。

そのため、従来のリメディアル的な英語学習の「焼き直し」や「再学習」ではなく、学生が所属する学部・学科の専門基礎学習内容と有機的にリンクさせることができる題材として、取り扱うことが容易で、コンテンツ科目のテキストとして、その教育目的を効果的に達成することが期待できるものである。

(2) 「総合英語」で用いる教授学習方略と評価テストについて

新カリキュラムの特性と学習効果を最大化するための教授学習方略として、CBI（Content-based Teaching/Instruction：内容重視型教授）と、TBLT（Task-based Language Teaching/Learning：タスク中心教授学習法）を採用した。

また、新たな枠組みによる新科目の創設に伴い、①学習者に最適化された教育環境を提供す

るとともに、②その教育的効果を検証することが求められる。そこで、入学時と各学期末における学生の英語運用能力を測定・診断するための英語運用能力診断テストと、英語学習実態調査を追跡的に実施することとした。

2.2 英語運用能力診断テスト開発のための古典的テスト理論と項目分析について

教育における評価の課題は、教育測定（Educational measurement）として幅広く研究が行われている。特に、今日の英語運用能力テスト開発の動向は、TOEIC®やIELTS®を代表とするテスト理論に即したものであり、また、CEFR等の国際的な基準に適合したものとなっている。しかし、日常の英語の授業における大学・学部単位の、あるいは、クラス単位の英語教育における語学テストの研究は、決して十分とはいえない状況が続いている。その原因の一つは、英語担当教員が言語テスト研究の専門的な知識や経験を有していないことにある。すなわち、英語のできる人（英語の教員）は、誰でも英語のテスト問題作成ができ、テスト実施ができ、また、テスト結果を科学的に分析し学習者へ適切にフィードバックを行い、授業改善を適宜行うことができると思われているところにある。

実際に、学生の英語運用能力を正しく評価するためには、正確に測定できる質問紙やテストなどの尺度を用いなければならない。そこでは、測定したデータが正確に英語運用能力を反映しているかどうかを、統計分析などの手法を用いて検証する必要がある。古典的テスト理論（Classical Test Theory）では、そのために用いる概念が妥当性と信頼性と、開発した問題の難易度である。

言語テストの開発作成においては、通常、問題素案作成を行った上で、問題構成（Test assemble）を行い、構成された言語テストの検証と校正（Test edit & Review）を実施し、一つの問題に対して正しい答えが一つであるか、使われている語句が適切か、曖昧でないかなど、さまざまな角度から検証をするプロセスが求められる。特に、昨今の言語テストにおいては、人権やジェンダー等に関するセンシティブレビューが必須となり、問題文の不適切な表現などを検証することが厳しく求められている。

開発された言語テストについては、プレテストを実施し、その結果を項目分析（Item analysis）し、正答率や難易度の分析、信頼性の確認等を行い、不適切問題等の確認・除外を行うなど、多岐に及ぶ実証と検証のプロセスが求められている。

3. 研究の目的：古典的テスト理論の限界と本研究の位置づけ

診断テストを実施することの目的の一つは、学生の英語運用能力を向上させるために必要な関連情報を収集するとともに、学生への適時フィードバックを行うことである。英語運用能力

を向上させるために必要な関連情報としては、カリキュラム評価や授業改善を行うための情報や、学習者自らが学習状況や英語学習の目標設定を行うために必要な情報等が含まれ、「点数」といった形式情報だけではなく、それぞれの目標を達成するための意味情報へ容易に転換させることができるものとして提供されなければならない。

そのために重要なことは、英語運用能力測定用に開発され用いられたツールであるテストが、本当にテストの備えるべき条件を満たした十分に優れたものであったかということを、科学的、統計的に検証することである。

開発した診断テストは、物差しのように、いつ誰が測定しても同じような結果を示すものか、また、語彙力やリスニング能力など、本当に測定したい英語運用能力を正しく測定しうるものかということを、客観的に示す根拠が求められる。また、そもそも英語の授業を通じて、コミュニケーション能力を育成することが目標とされているが、それを診断テストを用いて、どのように測定することができるのかという根拠や基準を明示することが求められている。

本研究により開発する英語運用能力テスト（英語力診断テスト）は、学習者の学習状況を評価するとともに、カリキュラム評価を行うことがその主たる目的であるため、目標基準準拠測定（*criterion-referenced measurement*）を用いる。そして、その目標設定にはいくつかの観点を定め、観点毎に測定可能な、しかも、妥当な到達目標を設定し測定することが求められる。また、設定した目標に達していることを、どのように決定するのかということも重要な要因であり、分割点（*cut-off score*）の適切な設定も課題となる。

3.1 到達目標の設定について

新カリキュラム及び本研究で開発する英語力診断テストにおける到達目標の枠組みとして、CEFR（*Common European Framework of Reference for Languages*）を用いた。

入学時に実施した診断テスト1においては、CEFRの「基礎段階の言語使用者（A1、A2レベル）」を到達目標とした質問項目を配列した。その基準となる能力記述文は以下の通りである。

（1） A1 レベルの能力記述文

具体的な欲求を満足させるための、よく使われる日常的表現と基本的な言い回しは理解し、用いることができる。自分や他人を紹介することができ、住んでいるところや、誰と知り合いであるか、持ち物などの個人的情報について、質問をしたり、答えたりすることができる。もし、相手がゆっくり、はっきりと話して、助けが得られるならば、簡単なやり取りをすることができる。

(表1：Common European Framework of Reference for Languages, CEFR)

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

(2) A2 レベルの能力記述文

ごく基本的な個人情報や家族情報、買い物、地元の地理、仕事など、直接的関係がある領域に関しては、文やよく使われる表現が理解できる。簡単で日常的な範囲なら、身近で日常の事柄について、単純で直接的な情報交換に応じることができる。

一方、春学期末に実施した英語力診断テスト2においては、CEFRが規定している「自立した言語使用者 (B1、B2 レベル)」を具体的な到達目標とした質問項目を開発し配列した。その基準となる能力記述文を以下に示す。

(3) B1 レベルの能力記述文

仕事、学校、娯楽などで普段出会うような身近な話題について、標準的な話し方であれば、主要な点を理解できる。その言葉が話されている地域にいるときに起こりそうな、たいていの事態に対処することができる。身近な話題や個人的に関心のある話題について、筋の通った簡単な文章を作ることができる。

(4) B2 レベルの能力記述文

自分の専門分野の技術的な議論も含めて、抽象的な話題でも具体的な話題でも、複雑な文章の主要な内容を理解できる。母語話者とはお互いに緊張しないで、普通にやり取りができるくらい流暢かつ自然である。幅広い話題について、明確で詳細な文章を作ることができる。

3.2 古典的テスト理論の限界

古典的テスト理論では、誰がいつ測定してもその測定結果が変わらないという、テスト結果の安定性を示すテストの信頼性 (reliability) と、開発したテスト項目により何が測定されているのか、また、測定したいものが測定されているかという、テストの妥当性 (validity) が問題となる。

特に、古典的テスト理論の基盤となっているものは、テスト得点の求め方にある。すなわち、正解した数 (素点 raw score) を受験生の成績・能力とする方法で、正答に基づく得点 (number-right score) とも呼ばれている。

(1) 項目困難度 (item difficulty) について

テスト項目の難易度は、テストの対象となる被験者を仮定している表現であり、そのため、受験者集団によって決定されるということになる。すなわち、テスト項目の難易度は標本依存 (sample-dependent) となる。逆に、受験するテスト項目により被験者の能力が定義されるということは、能力の決定はテスト依存 (test-dependent) ということになる。

そこで、英語のテスト得点が0点ということは、時間や温度や長さなどにおける絶対的な0とは異なり、当該のテストで示される0点から満点というレンジでの0を示すものである。すなわち、ある特定の被験者能力と、ある特定のテスト項目の難易度によって、任意に定められた0となる。また、テスト得点における測定単位の精度にも問題が生じてくる。0点が任意に設定されるため、0よりも低い能力の被験者や、同様に、満点よりも高い能力の被験者のスコアは、それぞれ0点、満点というレンジ内に圧縮されてしまうことになる。このように、0点付近でゆがみが生じる床面効果 (floor effect) と、満点付近でのゆがみとなる天井効果 (ceiling effect) を

受けることがなく、テスト得点のどの位置においても同一の測定単位と精度を保つことが課題となる。

(2) 項目弁別力 (item discrimination power) について

古典的テスト理論で用いられている項目弁別力指数も、項目困難度と同様に、被験者を構成する受験者集団の特性に依存したものである。従って、テスト項目の持つ絶対的な弁別力指数を求めることは困難である。

このように、受験者、すなわち、被験者集団の特性により、項目困難度や項目弁別力指数が相対的に決定されるという構造を持っているのが、古典的テスト理論の課題であり限界でもある。

(3) 信頼性 (reliability) について

古典的テスト理論の信頼性、あるいは、信頼性係数は、最も重要な項目の一つである。これは、テスト得点の安定性を示す係数で、同一のテストを使って同一の被験者に対して誰が測定しても、いつどこで測定しても、同じようなテスト得点を得ることができれば、そのテストの信頼性は高いものであるといえる。すなわち、能力の測定に誤差があまり生じないテストは、信頼性が高いテストであると評価される。

(4) 妥当性 (validity) について

テストの妥当性の検討ということは、測定しようとしている能力を、本当にそのテスト項目が測定しているかどうかを検討することである。一般的に古典的テスト理論で検討される妥当性には、内容的妥当性 (content validity)、基準関連妥当性 (criterion-related validity)、構成概念妥当性 (construct validity)、表面的妥当性 (face validity) 等がある。本研究では、妥当性の検討で最も一般的に用いられている、外部基準との関連性を求めるという基準関連妥当性を用いることとする。開発した診断テストの妥当性の度合いを、そのテスト以外の外部基準 (external criterion) と照合し検討するものである。そのためには、外部基準として、測定しようとする英語運用能力に最も近く、より妥当性の高い外部基準を参照することとした。

このように本研究では、診断テストという限定的なテストの実施環境ということを踏まえ、以下の古典的テスト理論の限界を理解した上で、古典的テスト理論の重要な構成概念や要因である項目困難度、項目弁別力、信頼性と妥当性を用いて、開発した診断テストの項目分析を行った。

(5) 古典的テキスト理論の限界

テスト項目自体の持っている困難度と被験者の能力が組み合わさり、初めて、信頼性、妥当性、項目困難度、弁別力等のテスト項目の特性が決定されるのが、古典的テスト理論の限界である。換言すれば、古典的テスト理論で求められたテスト項目の特性に関する値は、かなり限定された状況でしか意味をなさないといえる。すなわち、古典的テスト理論に基づく分析で求めたデータは、そのデータを作成したものと同一のテスト項目が、データ収集のためにそのテストを受験した集団と全く同じ集団に、もう一度与えられる場合にだけ、意味のある情報を与えてくれることになる。

また、古典的テスト理論では、被験者集団毎に、その項目の困難度や弁別力が異なったものとなる。つまり、項目特性の値は、被験者集団の数だけ存在することになる。そして、正解したテスト項目の数で表されるテスト得点というかたちで、被験者の能力を単に推定することになってしまう。すなわち、古典的テスト理論のテスト得点という概念では、被験者能力と項目困難度という2つの要因を切り離すことができなくなる。また、項目弁別力も同様に、被験者の能力との関係で、相対的に決定されてしまうことになる。

このような限界を理解した上で、「OIDAI英語力診断テスト」という、同一被験者を対象とし、CEFRの能力記述文で示される到達目標を基準とした診断テストの開発と、テスト項目分析による教育的有用性を検証することが、本研究の主たる目的である。

4. OIDAI英語力診断テストの開発と研究の方法

英語教育の現場で実施されている様々なテストや評価（診断的評価、形成的評価、総括的評価等）の究極の目的は、児童・生徒や学生の英語運用能力を向上させることにある。そのために、入学試験をはじめとした選抜的評価を行うテストを含め、多種多様なテストが開発され実施されている。そして、テスト開発者や授業担当者は、テスト結果から得られる各種情報に基づいて、学習者の英語運用能力を把握し、具体的な学習指導やカリキュラムや授業評価等の教育的評価を実施している。しかし、開発されたテスト項目や構成については、各項目の項目困難度や項目弁別力を十分に検証することなく実施され、後日改変されることは希である。また、テスト全体としても、信頼性や妥当性の検証や検討をすることなく、一つのテストとして実施され、被験者に対するフィードバックも総得点が示される程度で、その教育的な効果は限定的なものとなっている。テスト開発から実施・評価、フィードバックというプロセスに膨大な時間とエネルギーを費やすものの、その多くは再利用されずそのまま破棄されてしまう傾向がある。その原因の一つが、テスト項目やテスト結果の科学的な分析とその分析結果の積み上げがなされていないことにある。

そこで本研究においては、古典的テスト理論に基づき、項目困難度、項目弁別力、信頼性と妥当性を参照し、開発した診断テストを構成する各テスト項目の分析を行った。テスト開発においては、評価の目的や出題の意図に基づいて、また、これまでの教育やテスト作成の経験や実践に基づいて、テスト項目の難易度や信頼性や妥当性等を一定レベルにおいて想定しながら、テスト項目の作成を行った。そこでは、語彙や表現のレベル、構文、文型、テーマなど、テスト項目の難易度や弁別力に関わる要因等に配慮しながら、また、被験者の反応や応答を予測しながら、一つ一つのテスト項目の質問文や選択肢を作成したり編纂したりしている。しかし、どれほど細心の注意を払いテスト項目を作成したとしても、テスト項目作成時の想定や予測や作問の意図は、実際のテスト場面やテスト結果には十分反映されていないことがある。それどころか、実際に試験をすることにより、被験者が誤答選択肢により、想定以上の混乱や困難さを感じていたりするという実態が、はじめて顕在化したりするケースが出現する。そこで、作成したテスト項目の良否を見極め、それらを管理するための項目分析を継続的に実施し、良質項目と欠陥項目を特定し、欠陥項目については、その根拠を詳細に分析し、項目の改廃や改善を行うこととした。

そのために、本研究では、約2,000名の新1年生を対象として実施する大規模診断テストの前に、 β 版の予備診断テストを作成・実施し、項目分析を行うこととした。これにより、各テスト項目の特性を把握し、不良テスト項目を削除するなど、本来測定したい英語運用能力を、適切に測定することができる良質なテスト項目群からなる英語力診断テストを構成・編纂し、診断テストとしての妥当性を高めることとした。

5. 予備診断テスト（ β 版）の項目分析結果

「OIDAI英語力診断テスト」のプロトタイプとして開発した β 版の実施概要を表2に、また、 β 版の得点分布を示すヒストグラムを図1に示す。

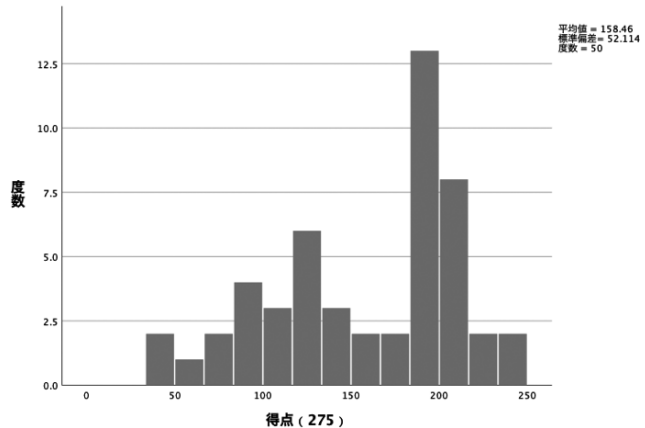
診断テスト β 版は、TOEIC L&R Tests及びTOEIC Bridge L&R Testsの問題形式に準拠したもので、被験者の語彙・表現力、文法・語法力、リスニング力と、読解力の4つの領域に関わる英語運用能力の測定を意図した80の質問項目からなる診断テストとして開発した。

β 版は、総合英語のクラス編成と入学生のリスニングとリーディングに関わる英語運用能力の実態を把握し、授業の最適化を図るために必要なデータを得ることを目的に、独自に開発した診断テストで、スマートフォンを用いたWeb受験ができるリスニング問題、語彙・表現問題、文法・構文問題、読解問題により構成されている。

診断テスト β 版の開発に際しては、リスニング問題25項目、語彙・表現問題20項目、文法・構文問題18項目、読解問題17項目、計80項目によりテストを構成・配列した。開発した診断

(表2：OIDAI 英語力診断テストβ版の実施概要)

英語力診断テストβ版	
総問題数	80
得点	275
受験者数	50
最高点	246
最低点	48
平均値	158.46
分散	2715.85
標準偏差	52.11
歪度	-0.49
尖度	-0.84



(図1：英語力診断テストβ版の得点分布)

テストβ版は、国際教養学科1、2年生50名を対象に3月上旬にWeb上で実施した。平均回答時間は110分であった。

なお、国際教養学科の学生をサンプリングした理由は、診断テストの妥当性を検定するための外部基準として、TOEIC L&Rのスコアを有していることであった。

本研究では、古典的テスト理論の観点から、診断テストで測定する4つの能力カテゴリ（語彙・表現力、文法・語法力、リスニング力、読解力）毎の信頼性の検定、ならびに、統計学的にみた項目の良否についての分析である項目分析を行った。

5.1 診断テストβ版のカテゴリ別信頼性の検定

表3に、各カテゴリごとに求めたテストの信頼性係数の推定値であるクロンバックの α 係数を示す。 α 係数は、「内部一貫性」(internal consistency)の推定を目的とするもので、当該項目間の等質性や一貫性がどの程度あるかを、数量的に示すことができる指標である。

いずれの項目も、項目数としては17から25、被験者数は50名であり、 α 係数を算出するための十分なサンプリング数ではないものの、その値は0.72から0.86を示しており、いずれのカ

(表3：診断テストβ版の項目分析概要)

β版の項目分析概要	リスニング	語彙・表現	文法・語法	読解
項目(問題)数	25	20	18	17
通過率(正答率)※項目困難度	55.29	63.50	51.10	60.20
上位下位項目弁別力指数	0.34	0.32	0.47	0.65
成績上位(H)群(27%)	0.70	0.80	0.68	0.86
成績下位(L)群(27%)	0.36	0.49	0.20	0.20
点双列相関係数による項目弁別力指数	0.29	0.31	0.41	0.52
信頼性係数(α)	0.72	0.80	0.79	0.86

テゴリにおいても十分な精度の測定結果が得られていると判断できる。

5.2 項目困難度の検証

テストの信頼性の検定によって、各カテゴリごとにテスト全体としては、望ましい精度の測定ができていたことが確認された。次に、診断テストβ版を構成している個々の項目が、それぞれ期待通りの機能を果たしているかどうかを検証するために、古典的テスト理論に基づく項目分析を行った。

結果の概要は、表3に示すとおりである。

項目困難度として、各質問項目の通過率を算出し、各カテゴリの平均値を求めた。その結果、各カテゴリの通過率は51.1%から63.5%を示し、診断テストやブレースメントテストに求められる項目困難度については、妥当であることが判明した。

5.3 弁別力の検証

項目弁別力の評価を行うに当たっては、H-L分析と点双列相関係数による検定を行った。

A. H-L分析による分析結果

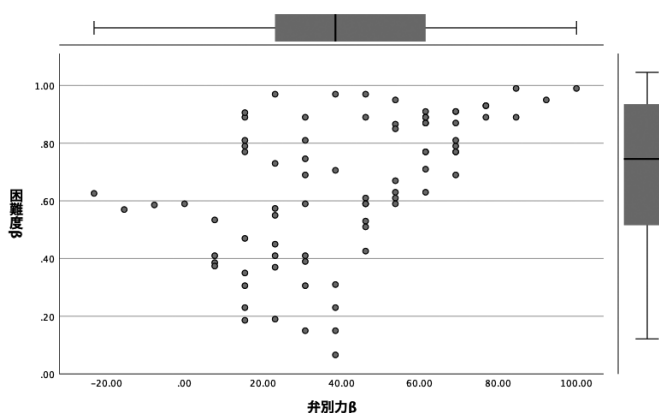
項目弁別力の算出に際しては、診断テストβ版の総得点で被験者総数の上位27%を高能力群（H群）、下位27%を低能力群（L群）とし、両群の各項目の通過率の差を算出した。項目弁別力の評価においては、.4を分割点として、両群の差が大きければ大きいほど項目弁別力が高いと判断される。

その結果、上位下位項目弁別力指数においては、.4を下回る項目群（リスニングと語彙表現問題）があり、難易度の調整が必要であることが判明した。一方、文法・語法問題と読解問題の指数はそれぞれ.47と.65であり、十分な弁別力を有しているものと判断することができた。

B. 点双列相関係数による分析結果

点双列相関係数は、その項目とその項目が含まれているテストの得点との相関を計算したものである。その意味で、項目－テスト得点（Item-Test）間相関、あるいはIT相関と呼ばれることもある。この指標は、いわば、その項目がそのテストで測定している特性を、どの程度とらえているかを見るものであり、識別力とよばれる統計量のひとつである。

点双列相関係数の評価においては、一般的に点双列相関係数の値は.2以上が望ましいとされることが多いが、本研究では、項目弁別力指数として、.3以上を基準とした。その結果、各項目群共に、基準に達していることが判明した。すなわち、全体的にみて、良い問題が作成できたと評価することができる。参考までに、図2にβ版の項目困難度と弁別力の散布図を示す。

(図2: β 版の困難度と弁別力の散布図)

5.4 妥当性の検証

本研究では、測定しようとしている被験者の英語運用能力を、開発した英語力診断テストが本当に測定し得るものであるかということを、基準関連妥当性に基づき検証する。すなわち、開発した診断テストの妥当性の度合いを、当該テスト以外の「外部基準」(external criterion)と照合して検証するもので、用いる外部基準は測定しようとしている項目内容に最も近いものが前提となる。

そのために、被験者としてサンプリングした国際教養学科の学生が受験してきたTOEIC L&Rのテストスコアを採用し、本研究で開発し実施した診断テストスコアとの相関係数を算出し、併存的妥当性 (concurrent validity) を検証することとした。

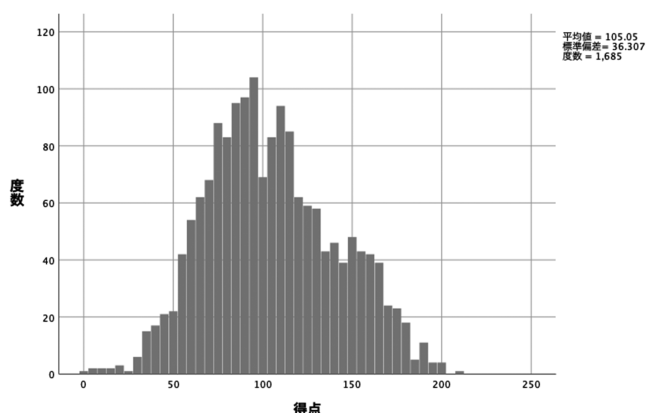
その結果、開発した診断テスト β 版のテストスコアと、外部基準として採用したTOEIC L&Rの相関係数は、 $r = .714$ ($p < .000$) で、プラスのかなり強い相関があることが判明した。また、算出された相関係数 (妥当性係数) から決定係数を導き出すと、.51 となり、開発した診断テスト β 版の妥当性は、一定レベルを有していることが判明した。

6. 「OIDAI 英語力診断テスト」の開発と分析結果

β 版で得られたデータに基づき、項目困難度、項目弁別力や信頼性をはじめとする項目分析を行い、 β 版の80項目から難易度、弁別力及び信頼性を考慮し、選択した67項目からなる新入生を対象とした「OIDAI 英語力診断テスト」を編纂した。これにより、Web 上での平均回答時間をより短縮することが期待された。

(表4：OIDAI英語力診断テスト実施概要)

OIDAI英語力診断テスト2021	
総問題数	67
得点	230
受験者数	1685
最高点	208
最低点	0
平均値	105.05
分散	1318.22
標準偏差	36.31
歪度	0.25
尖度	-0.42



(図3：OIDAI英語力診断テスト得点分布)

(表5：OIDAI英語力診断テストの項目分析概要)

概要	リスニング	語彙・表現	文法・語法	読解
項目(問題)数	19	19	13	16
通過率(正答率)	51.82	59.21	38.35	41.50
上位下位項目弁別力指数	0.34	0.34	0.32	0.48
成績上位(H)群(27%)	0.64	0.79	0.53	0.67
成績下位(L)群(27%)	0.30	0.44	0.20	0.19
点双列相関係数による項目弁別力指数	0.28	0.32	0.31	0.43
信頼性係数(α)	0.65	0.76	0.62	0.79

6.1 診断テストの実施概要と項目分析結果

2021年4月1日(木)から4月5日(月)までの5日間、新入生を対象として、スマートフォンからの受験も可能とし、Web上で診断テストを実施した。

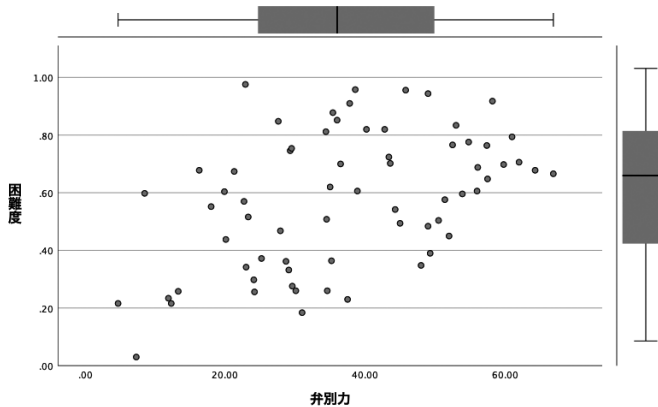
総受験者数は1,685名で、4月1日時点の入学予定者名簿に登録されているクラス編成対象者1,957名の86.1%であった(表4)。

また、実施した診断テストの基礎統計量概要、及び、項目分析結果を表5に、テスト得点分布のヒストグラムを図3に示す。

6.2 信頼性の検証

信頼性の一つの指標となる内部一貫性を検証するために、クロンバックの α 係数を算出した。その結果、各テスト項目カテゴリの信頼係数は、.62から.79と、 β 版よりもやや低めの値を示しているが、編纂した診断テストの内部一貫性の程度は満足できるレベルであることが確認された。

中でも、リスニング項目群と文法・語法項目群が、.8を満たすことができていないことが明



(図4：困難度と弁別力の散布図)

らかとなった。質問項目の内容や形式を一部改変することが求められている。しかし、その他の項目群は、一定レベルの内部一貫性を保持していることが確認された。

6.3 項目困難度の検証

診断テストやプレースメントテストに求められる項目困難度については、ほぼ妥当であること(3～.6)が判明した。被験者が新入生であることから、 β 版の項目分析結果とは異なり、文法・語法問題と読解問題の項目困難度が高くなっている(通過率が低くなっている)ことが判明した。また、上位下位項目弁別力指数においては、読解問題群を除いて、.4を下回っており、一部、難易度の調整が必要であることが示唆されている。

一方、読解問題の指数は.48であり、十分な弁別力を有しているものと判断できる。

6.4 弁別力の検証

一般的に、.3以上が求められる点双列相関係数による項目弁別力指数においては、23項目が基準を下回っており、今後の検討が望まれるところである。

なお、項目弁別力の算出には、上位下位項目弁別力指数(成績上位群27%、下位群27%による項目困難度差)と点双列相関係数による項目弁別力指数を掲載している。図4に、項目困難度と弁別力の散布図を示す。一定レベルの弁別力が示された β 版の散布図(図3)とは異なり、分散が大きいことが窺える。

6.5 妥当性の検証

国際教養学部1年生(国際教養学科・国際日本学科)の学生は、全員がTOEIC Bridgeを新入生ガイダンスの日に受験するため、妥当性を検証するために、TOEIC Bridgeのテストスコアを、

基準関連妥当性を算出するための外的基準として採用した。

その結果、開発した診断テストのテストスコアと、外的基準としてのTOEIC Bridgeの相関係数は、 $r = .669$ ($p < .000$) で、プラスの中程度の相関があることが判明した。また、これによる決定係数は、.45であり、国際教養学科の新2年生と新3年生50名を対象とした β 版の妥当性と比較すると、やや低くなっているものの、十分な妥当性が示されていることが確認できた。

7. 考察と課題

7.1 考察

本研究では、言語テストの一つである英語力診断テストを開発・実施する目的の一つとして、学生の英語運用能力の向上を図ることをその中心テーマとして取り組んできた。また、テスト開発においては、言語運用能力測定に用いる診断テストは、本当にテストの備えるべき条件を満たし、精度の高い測定を行うことができるツールとなっているのかということを検証してきた。

その結果、古典的テスト理論で求められる信頼性、難易度、弁別力、妥当性等の各指標については、一定レベルの基準を満たす診断テストを開発し、学生の英語運用能力を、設定した4つの能力カテゴリ（語彙・表現力、文法・語法力、リスニング力、読解力）に基づき、満足できる精度で測定することができたものと評価することができる。

しかし、本研究で用いたテスト理論と項目分析の枠組みは「古典的テスト理論」に依拠したものであるため、標本依存やテスト依存をはじめとする分析の限界と困難さを併存したものとなっている。

7.2 課題

今後は、これら古典的テスト理論の限界を克服した項目応答理論（Item Response Theory）に基づく分析やテスト開発を行うことが期待される。すなわち、すべてのテスト項目が、被験者の同一の潜在特性や能力を測定することができるもので、同一のテスト項目を異なる被験者集団で実施したとしても、その項目の持っている特性の値に変わりはないという結果を得ることができるテスト開発が望まれる。

また、本研究の項目困難度や項目弁別力で算出した値は、標本依存やテスト依存を強く反映したものであったが、その項目が、項目応答理論のモデルに一致していれば、被験者集団に関わらず、ひとつの不変的な値として示すことができるということになる。

さらに、測定の精度という観点で古典的テスト理論を評価すると、本研究の分析においても、測定の標準誤差や信頼性係数として、被験者集団全体に関する値を得ているのみで、本来の目

的としていた、学生一人ひとりの英語運用能力の向上に資する個別のデータを得ることはできていない。すなわち、測定の対象となる被験者集団が決まれば、その集団に対して1つの測定の精度が決定されるという標本依存のものであったため、学生一人ひとりに対するテストの精度は明らかとはならないという限界が生じていたのである。

これについては、項目応答理論では、項目情報関数 (Item information function) という、それぞれの能力水準に対応する測定精度の指数が用いられ、分割点付近における情報関数の値が最大となるように、被験者の能力に応じたテスト項目をコンピュータが提示し、被験者の能力にあった項目で能力を効率的に推定することができる適応型テスト (adaptive testing) の開発が可能となる。

このように、項目応答理論を用いることにより、項目情報関数により導き出される指数を、被験者集団全体という平均的なものではなく、被験者一人ひとりについて検討し活用することで、教育の個別最適化を図ることが期待できる。

参考文献

- 阿部真理子、ブライアン・ウィスナー、酒井英樹 (2008). 古典的項目分析とラッシュモデリングを用いた英語熟達度テストの分析. 『高崎経済大学論集』, 50, 3-4, 125-134.
- Council of Europe (2001). Common European Framework of Reference for Languages (CEFR) (Global scale - Table 1 (CEFR 3.3): Common Reference levels.
<https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale> (参照2021年7月)
- J.D. ブラウン. (1999). 『言語テストの基礎知識：正しい問題作成・評価のために』 東京：大修館書店.
- 水野康一 (2012). 英語多肢選択テストの項目分析. 『香川大学教育研究』. (香川大学教育開発センター). 1, 19-30.
- 大友賢二 (1996). 『項目応答理論入門：言語テスト・データの新しい分析法』 東京：大修館書店.
- Y. Wei (2017). Monitoring Score Change Patterns to Support TOEIC® Listening and Reading Test Quality.
<https://doi.org/10.1002/ets2.12186> (参照2021年7月)

資料 1-1：英語運用能力診断テストβ版項目分析結果-1

項目No.	項目特性	正解数	回答者数	正解率(通過率)%	最適項目困難度	項目困難度適切度	H正答者数(N=13)	L正答者数(N=13)	項目弁別力	点双列相関係数	p値
β項目1	L1	26	50	52.0	66.7	0.706	9	4	38.5	0.327	*
β項目2	L2	18	50	36.0	66.7	0.386	5	4	7.7	0.14	
β項目3	L3	49	50	98.0	66.7	0.374	13	12	7.7	0.25	
β項目4	L4	46	50	92.0	62.5	0.410	13	10	23.1	0.444	**
β項目5	L5	43	50	86.0	62.5	0.530	13	7	46.2	0.494	**
β項目6	L6	45	50	90.0	66.7	0.534	12	11	7.7	0.14	
β項目7	L7	16	50	32.0	66.7	0.306	7	3	30.8	0.24	
β項目8	L8	44	50	88.0	66.7	0.574	13	10	23.1	0.26	
β項目9	L9	23	50	46.0	66.7	0.586	6	7	-7.7	0.04	
β項目10	L10	27	50	54.0	66.7	0.746	9	5	30.8	0.301	*
β項目11	L11	24	50	48.0	66.7	0.626	5	8	-23.1	0	
β項目12	L12	30	50	60.0	66.7	0.866	11	4	53.8	0.465	**
β項目13	L13	16	50	32.0	66.7	0.306	5	3	15.4	0.22	
β項目14	L14	37	50	74.0	62.5	0.770	10	8	15.4	0.22	
β項目15	V1	41	50	82.0	62.5	0.610	13	6	53.8	0.619	**
β項目16	V2	41	50	82.0	62.5	0.610	13	7	46.2	0.49	**
β項目17	V3	34	50	68.0	62.5	0.890	10	6	30.8	0.379	**
β項目18	V4	32	50	64.0	62.5	0.970	9	6	23.1	0.24	
β項目19	V5	36	50	72.0	62.5	0.810	10	8	15.4	0.22	
β項目20	V6	39	50	78.0	62.5	0.690	11	7	30.8	0.342	*
β項目21	V7	34	50	68.0	62.5	0.890	11	3	61.5	0.464	**
β項目22	V8	46	50	92.0	62.5	0.410	12	11	7.7	0.17	
β項目23	V9	46	50	92.0	62.5	0.410	13	9	30.8	0.477	**
β項目24	V10	21	50	42.0	62.5	0.590	7	7	0.0	-0.1	
β項目25	V11	38	50	76.0	62.5	0.730	11	8	23.1	0.22	
β項目26	V12	42	50	84.0	62.5	0.570	9	11	-15.4	0.03	
β項目27	V13	36	50	72.0	62.5	0.810	11	7	30.8	0.299	*
β項目28	V14	45	50	90.0	62.5	0.450	13	10	23.1	0.291	*
β項目29	V15	47	50	94.0	62.5	0.370	13	10	23.1	0.487	**
β項目30	LRC1	34	50	68.0	62.5	0.890	9	7	15.4	0.24	
β項目31	LRC2	28	50	56.0	62.5	0.870	11	3	61.5	0.421	**
β項目32	LRC3	13	50	26.0	66.7	0.186	4	2	15.4	0.2	
β項目33	V16	10	50	20.0	66.7	0.066	6	1	38.5	0.335	*
β項目34	LRC4	19	50	38.0	66.7	0.426	9	3	46.2	0.353	*
β項目35	LRC5	31	50	62.0	66.7	0.906	8	6	15.4	0.07	
β項目36	LC1	19	50	38.0	62.5	0.510	9	3	46.2	0.33	*
β項目37	LC2	22	50	44.0	62.5	0.630	11	3	61.5	0.415	**
β項目38	LC3	12	50	24.0	62.5	0.230	5	3	15.4	0.14	
β項目39	G1	29	50	58.0	62.5	0.910	11	3	61.5	0.482	**
β項目40	G2	35	50	70.0	62.5	0.850	12	5	53.8	0.518	**

※最も適切な困難度（理想の困難度）は、.5であるが、被験者の当て推量によってそれぞれの選択肢が選ばれる確率を加味したものが、最適困難度である。

※選択肢が4択の場合の最適困難度は.625で、3択の場合には.667となる。

※項目困難度適切度は、項目困難度と最適困難度を用いて算出した水準で、項目困難度が.0の場合-.250に、1.0の場合.25となり、困難度を容易に評価することができる。

資料1-2：英語運用能力診断テストβ版項目分析結果-2

項目No.	項目特性	正解数	回答者数	正解率(通過率)%	最適項目困難度	項目困難度適切度	H正答者数(N=13)	L正答者数(N=13)	項目弁別力	点双列相関係数	p値
β項目41	G3	33	50	66.0	62.5	0.930	12	2	76.9	0.585	**
β項目42	G4	21	50	42.0	62.5	0.590	9	3	46.2	0.323	*
β項目43	G5	37	50	74.0	62.5	0.770	13	4	69.2	0.73	**
β項目44	G6	26	50	52.0	62.5	0.790	10	1	69.2	0.563	**
β項目45	LC4	37	50	74.0	62.5	0.770	13	5	61.5	0.574	**
β項目46	LC5	37	50	74.0	62.5	0.770	12	4	61.5	0.579	**
β項目47	LC6	29	50	58.0	62.5	0.910	12	3	69.2	0.591	**
β項目48	G7	20	50	40.0	62.5	0.550	8	5	23.1	0.17	
β項目49	G8	12	50	24.0	62.5	0.230	5	0	38.5	0.323	*
β項目50	G9	21	50	42.0	62.5	0.590	7	1	46.2	0.453	**
β項目51	G10	18	50	36.0	62.5	0.470	7	5	15.4	0.05	
β項目52	G11	34	50	68.0	62.5	0.890	12	4	61.5	0.483	**
β項目53	V17	10	50	20.0	62.5	0.150	4	0	30.8	0.326	*
β項目54	RC1	32	50	64.0	62.5	0.970	11	5	46.2	0.482	**
β項目55	G12	28	50	56.0	62.5	0.870	12	3	69.2	0.513	**
β項目56	G13	11	50	22.0	62.5	0.190	5	2	23.1	0.22	
β項目57	G14	10	50	20.0	62.5	0.150	5	0	38.5	0.396	**
β項目58	G15	24	50	48.0	62.5	0.710	10	2	61.5	0.455	**
β項目59	G16	23	50	46.0	62.5	0.670	9	2	53.8	0.484	**
β項目60	G17	21	50	42.0	62.5	0.590	7	3	30.8	0.27	
β項目61	G18	15	50	30.0	62.5	0.350	4	2	15.4	0.07	
β項目62	RC2	16	50	32.0	62.5	0.390	8	4	30.8	0.24	
β項目63	RC3	28	50	56.0	62.5	0.870	11	3	61.5	0.495	**
β項目64	RC4	29	50	58.0	62.5	0.910	12	3	69.2	0.532	**
β項目65	RC5	26	50	52.0	62.5	0.790	8	6	15.4	0.18	
β項目66	RC6	34	50	68.0	62.5	0.890	13	3	76.9	0.634	**
β項目67	RC7	37	50	74.0	62.5	0.770	13	4	69.2	0.599	**
β項目68	RC8	31	50	62.0	62.5	0.990	13	0	100.0	0.767	**
β項目69	RC9	14	50	28.0	62.5	0.310	5	0	38.5	0.282	*
β項目70	RC10	33	50	66.0	62.5	0.930	13	3	76.9	0.651	**
β項目71	V18	31	50	62.0	62.5	0.990	13	2	84.6	0.683	**
β項目72	RC11	39	50	78.0	62.5	0.690	13	4	69.2	0.629	**
β項目73	RC12	34	50	68.0	62.5	0.890	12	6	46.2	0.377	**
β項目74	RC13	30	50	60.0	62.5	0.950	10	3	53.8	0.488	**
β項目75	V19	22	50	44.0	62.5	0.630	9	2	53.8	0.566	**
β項目76	V20	32	50	64.0	62.5	0.970	11	6	38.5	0.35	*
β項目77	RC14	30	50	60.0	62.5	0.950	13	1	92.3	0.685	**
β項目78	RC15	34	50	68.0	62.5	0.890	13	2	84.6	0.737	**
β項目79	RC16	21	50	42.0	62.5	0.590	8	1	53.8	0.416	**
β項目80	RC17	36	50	72.0	62.5	0.810	13	4	69.2	0.648	**

項目分析と古典的テスト理論に基づく「OIDAI英語力診断テスト」の研究開発

資料2-1：OIDAI英語運用能力診断テスト項目分析結果-1

項目No.	β 版	項目特性	正解数	回答者数	正解率(通過率)%	最適項目困難度	項目困難度適切度	H正答者数(N=464)	L正答者数(N=453)	項目弁別力	点双列相関係数	p値
	β 項目1*											
	β 項目2*											
項目1	β 項目3	LP1	1338	1685	79.4	66.7	0.746	430	287	29.3	0.204	**
項目2	β 項目4	LP2	1416	1685	84.0	62.5	0.570	432	319	22.7	0.170	**
項目3	β 項目5	LP3	1091	1685	64.7	62.5	0.956	398	181	45.8	0.348	**
項目4	β 項目6	LR1	1226	1685	72.8	66.7	0.878	408	238	35.4	0.279	**
	β 項目7*											
項目5	β 項目8	LR2	926	1685	55.0	66.7	0.766	385	138	52.5	0.424	**
	β 項目9*											
	β 項目10*											
	β 項目11*											
項目6	β 項目12	LR3	710	1685	42.1	66.7	0.508	278	115	34.5	0.293	**
項目7	β 項目13	LR4	499	1685	29.6	66.7	0.258	164	100	13.3	0.130	**
項目8	β 項目14	LR5	1103	1685	65.5	66.7	0.976	348	236	22.9	0.139	**
項目9	β 項目15	V1	1305	1685	77.4	62.5	0.702	441	233	43.6	0.362	**
項目10	β 項目16	V2	805	1685	47.8	62.5	0.706	379	89	62.0	0.493	**
項目11	β 項目17	V3	1129	1685	67.0	62.5	0.910	378	198	37.8	0.250	**
項目12	β 項目18	V4	1204	1685	71.5	62.5	0.820	417	225	40.2	0.308	**
項目13	β 項目19	V5	1325	1685	78.6	62.5	0.678	392	309	16.3	0.090	**
項目14	β 項目20	V6	1088	1685	64.6	62.5	0.958	387	203	38.6	0.297	**
項目15	β 項目21	V7	1006	1685	59.7	62.5	0.944	379	148	49.0	0.342	**
項目16	β 項目22	V8	1328	1685	78.8	62.5	0.674	407	301	21.3	0.170	**
項目17	β 項目23	V9	1501	1685	89.1	62.5	0.468	454	317	27.9	0.262	**
項目18	β 項目24	V10	676	1685	40.1	62.5	0.552	228	141	18.0	0.149	**
項目19	β 項目25	V11	1181	1685	70.1	62.5	0.848	381	247	27.6	0.194	**
項目20	β 項目26	V12	1386	1685	82.3	62.5	0.604	415	315	19.9	0.129	**
項目21	β 項目27	V13	1211	1685	71.9	62.5	0.812	384	219	34.4	0.223	**
項目22	β 項目28	V14	1374	1685	81.5	62.5	0.620	443	274	35.0	0.313	**
項目23	β 項目29	V15	1286	1685	76.3	62.5	0.724	439	232	43.4	0.367	**
項目24	β 項目30	LRC1	916	1685	54.4	66.7	0.754	316	175	29.5	0.214	**
項目25	β 項目31	LRC2	660	1685	39.2	66.7	0.450	311	68	52.0	0.402	**
項目26	β 項目32	LRC3	500	1685	29.7	66.7	0.260	224	62	34.6	0.318	**
項目27	β 項目33	V16	532	1685	31.6	66.7	0.298	206	92	24.1	0.212	**
項目28	β 項目34	LRC4	714	1685	42.4	62.5	0.598	198	155	8.5	0.030	**
項目29	β 項目35	LRC5	929	1685	55.1	62.5	0.852	329	158	36.0	0.314	**
項目30	β 項目36	LC1	491	1685	29.1	62.5	0.332	211	74	29.1	0.285	**
項目31	β 項目37	LC2	404	1685	24.0	62.5	0.230	214	39	37.5	0.354	**
項目32	β 項目38	LC3	392	1685	23.3	62.5	0.216	114	90	4.7	0.055	**
項目33	β 項目39	G1	801	1685	47.5	62.5	0.700	316	143	36.5	0.293	**
項目34	β 項目40	G2	914	1685	54.2	62.5	0.834	379	130	53.0	0.425	**
項目35	β 項目41	G3	782	1685	46.4	62.5	0.678	371	71	64.3	0.476	**
項目36	β 項目42	G4	645	1685	38.3	62.5	0.516	238	127	23.3	0.226	**
項目37	β 項目43	G5	855	1685	50.7	62.5	0.764	373	104	57.4	0.431	**
項目38	β 項目44#	G6	366	1685	21.7	62.5	0.184	197	52	31.0	0.339	**
項目39	β 項目45	LC4	791	1685	46.9	62.5	0.688	366	103	56.1	0.472	**
項目40	β 項目46	LC5	864	1685	51.3	62.5	0.776	375	118	54.8	0.421	**

※項目番号に付記した*は、削除された問題を表す。

※項目番号に付記した#は、問題・選択肢が一部変更されたものを表す。

資料2-2：OIDAI英語運用能力診断テスト項目分析結果-2

項目No.	β 版	項目特性	正解数	回答者数	正解率(通 過率)%	最適項目困 難度	項目困難度 適切度	H正答者数 (N=464)	L正答者数 (N=453)	項目弁別力	点双列相関 係数	p値
項目41	β 項目47	LC6	503	1685	29.9	62.5	0.348	281	57	48.0	0.423	**
項目42	β 項目48	G7	498	1685	29.6	62.5	0.342	195	86	23.0	0.183	**
項目43	β 項目49	G8	236	1685	14.0	62.5	0.030	83	48	7.3	0.111	**
β 項目50*												
β 項目51*												
項目44	β 項目52	G9	902	1685	53.5	62.5	0.820	349	147	42.8	0.308	**
β 項目53*												
項目45	β 項目54	RC1	515	1685	30.6	62.5	0.362	213	78	28.7	0.239	**
項目46	β 項目55	G10	429	1685	25.5	62.5	0.260	205	64	30.1	0.269	**
β 項目56*												
β 項目57*												
項目47	β 項目58	G11	443	1685	26.3	62.5	0.276	205	66	29.6	0.284	**
項目48	β 項目59	G12	393	1685	23.3	62.5	0.216	140	81	12.3	0.117	**
項目49	β 項目60	G13	407	1685	24.2	62.5	0.234	132	75	11.9	0.111	**
β 項目61*												
項目50	β 項目62	RC2	426	1685	25.3	62.5	0.256	190	76	24.2	0.282	**
項目51	β 項目63	RC3	618	1685	36.7	62.5	0.484	298	69	49.0	0.389	**
項目52	β 項目64	RC4	696	1685	41.3	62.5	0.576	338	97	51.4	0.423	**
項目53	β 項目65	RC5	518	1685	30.7	62.5	0.364	238	73	35.2	0.310	**
項目54	β 項目66	RC6	722	1685	42.8	62.5	0.606	344	82	56.0	0.450	**
項目55	β 項目67	RC7	635	1685	37.7	62.5	0.504	307	71	50.5	0.394	**
項目56	β 項目68	RC8	626	1685	37.2	62.5	0.494	298	87	45.0	0.399	**
項目57	β 項目69	V17	580	1685	34.4	62.5	0.438	201	105	20.1	0.146	**
項目58	β 項目70	RC9	757	1685	44.9	62.5	0.648	352	83	57.5	0.480	**
項目59	β 項目71	V18	772	1685	45.8	62.5	0.666	379	67	66.9	0.506	**
項目60	β 項目72	RC10	984	1685	58.4	62.5	0.918	395	122	58.2	0.444	**
項目61	β 項目73	RC11	712	1685	42.3	62.5	0.596	334	82	53.9	0.438	**
項目62	β 項目74	RC12	668	1685	39.6	62.5	0.542	300	92	44.3	0.408	**
項目63	β 項目75	V19	524	1685	31.1	62.5	0.372	213	94	25.2	0.225	**
β 項目76*												
項目64	β 項目77	RC13	540	1685	32.0	62.5	0.390	287	57	49.3	0.433	**
項目65	β 項目78	RC14	799	1685	47.4	62.5	0.698	382	102	59.8	0.488	**
項目66	β 項目79	RC15	721	1685	42.8	62.5	0.606	288	105	38.9	0.273	**
項目67	β 項目80	RC16	880	1685	52.2	62.5	0.794	399	113	61.0	0.468	**

※項目番号に付記した*は、削除された問題を表す。

※項目番号に付記した#は、問題・選択肢が一部変更されたものを表す。