

実践研究

生命科学部・薬学部「プロジェクト発信型英語プログラム」 における独自のプレイスメント評価モデル “English Test in Academic Context (e-TAC)” の実施について — 成果と課題 —

近 藤 雪 絵・山 中 司

要 旨

立命館大学生命科学部・薬学部の英語授業「プロジェクト発信型英語プログラム」ではプレイスメント・テストの一部として English Test in Academic Context (e-TAC) を実施している。同学部では、開講初年である 2008 年は TOEIC Bridge® を用いて習熟度別クラスを編成してきたが、上位層・下位層の編成をより適切に行うため、2009 年より独自テストである e-TAC を導入した。以降、テストの実施方法やクラス編成への反映方法に改善を加えながら、2011 年より e-TAC の本格的な導入を開始した。本稿では e-TAC の実施状況を報告するとともに、2011 年度～2013 年度の結果について古典的項目分析、並びに TOEIC Bridge® との相関分析から、今後の更なる改善に向けた課題を考察する。

キーワード

プレイスメント・テスト、プログラム評価、プロジェクト発信型英語プログラム、e-TAC

1 はじめに

本論文は、鈴木（2003, 2010, 2012）が開発した立命館大学生命科学部・薬学部独自の英語教育である「プロジェクト発信型英語プログラム」の評価実践について、その一部を報告し、議論するものである。以降では、2009 年度より鈴木が開発、導入し、改善を繰り返してきた“English Test in Academic Context (e-TAC)”について取り上げ、成果と今後の課題について検討する。

英語教育を論じる際、避けては通れない点に「評価をどうするのか」という議論がある。確かに昨今の日本の大学英語教育には、コミュニケーション重視の授業を目指し改革に挑む姿を数多く見ることができる。そしてこうした動きは大学以外にも、高等学校や中学校、さらには小学校の英語教育にまで広がり、英語でコミュニケーションすることを授業内で積極的に取り入れ、それを評価していこうとする新たな政策的潮流が生まれていると考えることができる。しかし、学習者が創り出す「コミュニケーション」をどう評価するのかについて、適切なモデルが提示でき

なければ、方法論に対する説得力は著しく低くなる。

国全体としての大学英語教育が停滞する中、コミュニケーションを重視し、学習者の確実な英語力の向上を成し遂げてきた「プロジェクト発信型英語プログラム」は、20年以上の進化と発展の歴史（鈴木, 2003, 2012）を経て、「うまくいく」英語教育として今後様々な教育機関で参考とされ、取り入れられていく可能性が高いといえよう。本論文で議論する e-TAC は、「プロジェクト発信型英語プログラム」の遂行過程で生み出された評価システムの一つであり、既存の評価システムとは一線を画すものである。本論文では e-TAC の概要についての説明からはじめ、これまでの成果と分析、更なる改善点について検討を行う。新たな潮流に向かって待ったなしの変化に曝されている大学英語教育にとって、本報告は一つの方向性を示し得るものと考えたい。

2 e-TAC とは何か

e-TAC については、既に鈴木（2012）に次のとおり言及があり、以降本論文はこの記述を前提として論を進める。

・・・プロトタイプが、筆者（鈴木）が前職で開発して幾つかの授業で実験した English Test in Academic Context (e-TAC) である。(中略) e-TAC は、プロジェクト発信型英語プログラムのリサーチ活動の流れを取り入れて、1. 情報を読んで理解し (reading comprehension/grammar/vocabulary/expressions)、2. 情報に関するレクチャーを聞いて理解し (listening comprehension)、3. 自分の考えについてエッセイを書き (writing)、4. プレゼンテーションをして質問に答える (presentation/discussion/debate) という4つのプロセスから成る。米国の大学や大学院の授業では、1. 授業前に reading assignments を読んでおく、授業では、まず、クイズで内容理解のチェックを受け、そして2. レクチャーを受けて、3. エッセイやペーパーを書く。presentation/discussion/debate はレクチャーに付随する。e-TAC は、米国の大学や大学院のこうした授業活動にどの程度対応できるかを測定するテストとしても用いることができる・・・(pp.218-219)

e-TAC の詳細については後章に譲るとして、ここで強調されるべき論点は、テストそのもののコンセプトが実用的であると共に、受験者にとって、テストの受験が「役に立つ」ものであることが認識され易いことを挙げておきたい。すなわち、e-TAC は英語の授業で必要となる能力を測定する試験であるから、このテストに取り組むこと自体が学習者にとって有益で、良い点数をおさめることは授業での成果に直結する。e-TAC を受験することは、英語で行うプロジェクト活動の実践ともいえるのであり、テストを通して学習者自身の能力を伸張させることができるのである。また e-TAC は学習者の受信能力のみでなく発信能力も評価する。さらにはリサーチ、プレゼンテーション、ディスカッション、ディベート等の、超言語運用スキル (professional skills または supra-skills) (鈴木, 2012, p. 14) をも評価の一部に組み込んでいることから¹⁾、学習者にとって、自身の様々な英語能力が多角的に評価されていることが否が応にも分かる。これは学習者にとって表面的妥当性 (face validity) を高め、これが結果的にテストそのものに対する信頼性を高

めることにも繋がる。

現在 e-TAC は、立命館大学生命科学部・薬学部「プロジェクト発信型英語プログラム」における 1 回生入学時のプレイズメント・テストの一部に用いられ、導入前に比較してプログラムにおける運営の効果を大いに高めた経緯がある。本章では、e-TAC 導入の背景から紐解き、その有効性と意義について検討を行う。

2.1 習熟度別クラス編成が抱える問題点

「英語が使える日本人」や、「実践的コミュニケーション能力」といった言葉に代表されるように、国の施策として、今後コミュニケーションを重視した英語教育に一層の力点が置かれることは明らかなように思われる。大学をはじめとした各教育機関は、旧来の訳読型や教員による一方的な授業形式からの転換を進めており、コミュニケーションを重視した英語教育についても、現場の教員や大学生からも大卒でコンセンサスが得られていると考えてよい。

大学英語教育にとって、一定の成果が求められる時代となりつつある今日、大学生に適切な英語教育を行うことによって、彼らの能力を伸ばし、社会に必要とされる人材に育て上げることに異論を唱える者はいないであろう。英語教育の観点からこれを実現するのに効率が良いと思われる方法の一つに習熟度別クラス編成がある。「プロジェクト」と「スキル・ワークショップ」の 2 つのモジュールで構成される「プロジェクト発信型英語プログラム」においても、後者で習熟度別クラス編成が採用されている。「プロジェクト」に必要な英語運用スキルと英語能力を学習する「スキル・ワークショップ」(鈴木, 2012, p. 40) では、英語のブラッシュアップを集中して行うため、学習者の英語能力のばらつきは少ないことが望ましい。

様々な利点を持つと思われる習熟度別クラス編成であるが、無論問題点も抱えている。特にここでは授業運営の観点に絞って言及するが、1 回生入学直後の第 1 回英語授業から習熟度別クラス編成で行う場合、何よりもテストそのものの採点に時間がかけられない。すなわち、TOEFL®, TOEIC® に代表されるような大規模テストを実施し、結果を待っている余裕がないのである。やむを得ず立命館大学びわこ・くさつ・キャンパス (BKC) における多くの学部は TOEIC Bridge® を用いることで採点にかかる時間を大幅に短縮させ、極めて限られた時間内にクラス編成を実施し、初回授業に間に合わせている。クラス編成を行うためには、クラス分けテストを実施しさえすればよいというわけではない。テストの結果をもとに、クラス人数、クラスレベル、能力のばらつきや男女比、留学生の有無等に配慮しながら、事務局と情報を共有し、相談しながらクラスを確定し、この後に、事務局は掲示用の一覧の作成、履修システムの学籍情報の登録、名簿の作成と印刷等を行わなければならない。また、履修する 1 回生に編成したクラスを周知させる期間も必要である。入学式を早めるわけにはいかず、初回授業を遅らすわけにもいかないため、毎年ぎりぎりのスケジュールでクラス編成を行っているのが実情なのである。理想的な状況下でクラス編成を行うことは極めて難しいことを改めて強調したい。

制約の多い状況の中でクラス編成がなされるわけであるが、ここで看過できない問題が発生する。それが、TOEIC Bridge®-IP で発生する、得点の天井効果と床面効果である。これらの発生は TOEIC Bridge®-IP に限ったことではなく、項目応答理論等を用いない「正当数に基づく得点」によって点数化するいかなるテストにおいても理論上必至の現象である。そしてこれは、TOEIC

Bridge® に付随する問題というよりは、そもそも難易度の設定が適切ではない TOEIC Bridge® を大学 1 回生のプレースメントで用いることによって生じる問題である。すなわち、一部の学習者にとっては問題が簡単過ぎるのであり、また一部の学習者にとっては問題が難し過ぎるが、マークシート形式であるため点数がある程度取れてしまうということである。ただし、TOEIC Bridge® を大学英語教育のプレースメントで用いることの不適切さは重々承知しつつも、現実的な採点時間の制約から利用せざるを得ない。また、たとえ天井効果と床面効果を認めたととしても、それ以外の箇所、すなわち中程度のスコア帯に属する学生の英語能力は適切に評価できている可能性が高く、これらを有効に活用する意義は十分にあるとあってよいだろう。

2.2 立命館大学 生命科学部・薬学部「プロジェクト発信型英語プログラム」における e-TAC の位置づけと活用

2.1 で議論したとおり、習熟度別クラスで英語プログラムを行うためには、非常に制約の多い中でクラス編成を実施しなければならず、TOEIC Bridge® を使用する限り天井効果と床面効果の問題が避けられない。こうした状況を発展的に解決するために導入されたのが e-TAC であった。2 で述べたとおり、e-TAC のアイデアは鈴木（2012）の中にあり、既に慶應義塾大学湘南藤沢キャンパス、千葉商科大学等で実践事例があった。生命科学部・薬学部の「プロジェクト発信型英語プログラム」における e-TAC は、当初の目的として、TOEIC Bridge® の天井効果と床面効果を解消することを主眼に据えた。TOEIC Bridge® の補完として活用することで、上位クラスと下位クラスの編成をより適切に行うことを目指したのである。当初は e-TAC を部分的に活用することで、TOEIC Bridge® によるクラス編成を補助する位置づけとしていたが、年を経る毎に e-TAC による評価の有効性を積極的に取り入れる方向に議論が進み、e-TAC による評価割合を拡張した。最終的に、TOEIC Bridge® だけでクラス編成を行っていた頃と比べ、多くの評価データを総合して実施することが可能となり、より妥当性の高い評価とプレースメントが実現できている。以降でその詳細について記載する。

3 e-TAC の問題構成

「スキル・ワークショップ」のプレースメント・テストの一部として実施された e-TAC は Part 1 リーディング、Part 2 リスニング、Part 3 ライティングで構成される全 14 問約 45 分 100 点満点のテストである。表 1 にパート毎の問題の種類、問題形式、問題数、配点、解答時間を示した。

表 1 e-TAC の問題構成

パート	問題の種類	問題形式	問題数	配点	解答時間
Part 1	リーディング	4 肢択一 (語句補充、内容把握)	8 問	40 点 (5 点× 8 問)	20 分
Part 2	リスニング	4 肢択一 (内容把握)	5 問	25 点 (5 点× 5 問)	約 10 分
Part 3	ライティング	自由記述 (エッセイ)	1 問	35 点	15 分

Part 1 のリーディングでは約 400 語のサイエンスに関連したショートストーリーを読み、語句補充問題 5 問と、内容把握問題 3 問に答える。Part 2 のリスニングでは、Part 1 で読んだショートストーリーに関するレクチャー動画を見た後、内容把握問題 5 問に答える。レクチャー動画には音声に加え、レクチャーの要点を箇条書きにしたスライドが表示される (図 1)。問題は音声で 2 回読み上げられ、問題用紙には記載されていない。Part 1、Part 2 は全て 4 肢択一形式である。Part 3 のライティングは自由記述形式のエッセイ問題で、これまでに読んだ小説や漫画、鑑賞した映画、テレビ番組等から自分が実現を望むアイデアを選び、200 語程度で作品の紹介をした後、独自のアイデアを記述する。なお、一旦各パートが終了すると、前のパートに戻って解答したり本文を読んだりすることは出来ない。Part 1 から Part 3 までは 1 つのトピックに基づいて作られており、2 で述べた、情報を読んで理解し、情報に関するレクチャーを聞いて理解し、自分の考えについてエッセイを書くという「プロジェクト発信型英語プログラム」におけるリサーチ活動の流れが短時間の中に取り入れられている。このようにして、e-TAC は限られた時間の中で受験者の受信能力のみでなく、発信能力も評価するテストとして構成されている。

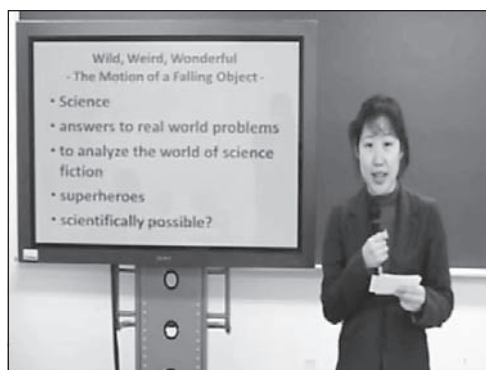


図 1 e-TAC リスニング問題よりレクチャー動画の例

4 e-TAC の実施要領

Part 1 のリーディング、Part 2 のリスニングはマークシート形式、Part 3 のライティングは解答用紙への書き込みによる自由記述形式を採用した。全パートを通じて、テストのインストラクションは音声と文字で動画に収録されている (図 2)。すなわち、e-TAC はテストの説明、問題の解答方法、各パートの説明、テストの開始と終了の合図が 1 つの動画にまとめられているため、試験監督は最初に再生ボタンを押すのみで試験を実施できる。こうすることにより、新任教員が試験監督を担当した場合でも、音声の再生や時間管理等の複雑な手順に伴う混乱を回避しやすくなる。e-TAC と同日に実施される TOEIC Bridge® に際しては、主監督の教員に加え、タイムキーパー等を担当する監督補助者が外部機関より派遣され、各教室に 3 名程度配属される。e-TAC は生命科学部・薬学部担当の英語教員 (主監督) のみで実施する必要があるため、インストラクションを動画に収録したのは運営面で大変効果的であった。

- このTestは約45分で3つのPartから構成されてい
ます。
- Part I は20分で読解問題です。
- Part II は約10分でPart I で読んだ英文についての聞き取
り問題です。
- Part III は15分で自由英作文です。テーマはPart I、Part
IIの内容に関連します。
- Part I、Part IIはマークシートに解答して下さい。
- Part IIIは自由英作文用解答欄に解答して下さい。
- 指示に従って解答して下さい。

図2 e-TAC インストラクション動画の例

5 e-TAC の採点

採点は試験監督を担当した英語教員が行った。2013年度は7名で作業を分担し、テスト実施日に Part 1、Part 2 のマークシートの読み込みとデータ入力、並びに Part 3 の自由記述問題の採点とデータ入力を行った。自由記述問題は学生の英文をまず7段階に評価した後、35点満点に調整した。ライティングの評価は英文の語数、文法、語彙、表現を対象とした。全てのパートの得点を合計して100点満点にしたものを、翌日結果が返却された TOEIC Bridge® の得点と合算した。

6 e-TAC によるプレイスメント（クラス分けへの反映方法）

2011年度よりプレイスメントはe-TACの得点と同日に実施される TOEIC Bridge® の得点との合計点に基づいて行われている。表2にプレイスメント・テストの全体の構成をまとめた。解答時間は実際に各パートの問題に解答する時間のみを表し、アナウンスの開始から問題用紙の回収までを考慮すると、テスト実施に TOEIC Bridge® は約90分、e-TAC は約60分を要した。

表2 2011年度～2013年度 英語プレイスメント・テストの構成

テスト	パート	問題の種類	問題数	配点		解答時間		テスト時間
TOEIC Bridge®	Part 1	リスニング	15問	90点	180点	25分	60分	90分
	Part 2		20問					
	Part 3		15問					
	Part 4	リーディング	30問	90点		35分		
	Part 5		20問					
e-TAC	Part 1	リーディング	8問	40点	100点	20分	45分	60分
	Part 2	リスニング	5問	25点		10分		
	Part 3	ライティング	1問	35点		15分		
合計			114問	280点		105分		150分

次に、2.2でも述べた現在のプレイズメント方法を採用した経緯を補足する。まず、「プロジェクト発信型英語プログラム」開講初年度に当たる2008年度は、TOEIC Bridge®の得点のみに基づいてプレイズメントを行った。ところが、前期 Semester 開始後からスキル・ワークショップを担当する外部教育機関との会議を重ねる中で、TOEIC Bridge®のみを使用したプレイズメントではスキル・ワークショップの授業に応じた英語力が十分に測れていないことが確認された。特に上位クラスではライティング力が低い学生がいること、下位クラスではリスニング力の差が大きく、講師の英語による説明が理解できない学生がいることが判明した。これを踏まえ、2009年度よりTOEIC Bridge®を補完するテストとしてe-TACの実施を開始した。2009年度、2010年度のプレイズメントでは全学生をまずTOEIC Bridge®の得点に基づいて26クラスに分けた後、上位4クラスと下位4クラス内のみe-TACの得点を加味して再編成した。2年間の実施を経てスキル・ワークショップ運営におけるe-TACの効果を確認した後、2011年度からはe-TACの全面実施を開始した。このような過程を辿り、2011年度からはTOEIC Bridge®とe-TACの合計点に基づいて受験者を26分割することでプレイズメントを行っている。

7 e-TAC の得点分布と基礎統計

2011年度から2013年度のe-TACの得点分布を図3に示した。何れの年も正規分布に近い単峰形を取り、やや負に裾を引いたなだらかな分布となっている。分布に関しては、3年を通じて重大な傾向の違いは見られない。

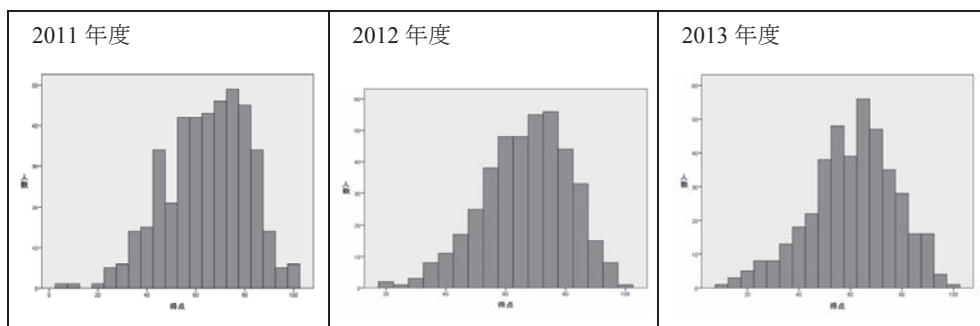


図3 2011年度～2013年度 e-TAC の得点分布

次に、e-TAC 基礎統計量を表3に示した。平均値、並びに中央値から、e-TAC は生命科学部・薬学部の新生者が6割強を取得するレベルの試験であるといえる。また、何れの年も満点取得者がおり、その数は2011年度が6名、2012年度が1名、2013年度が1名であった。平均値は2012年度が最も高く、次に2011年度が続き、2013年度は最も低い結果となった。平均値を比較するため分散分析を行ったところ、2013年度の平均点は0.05%の水準で、他の年に対し有意差が認められた。なお、平均値の差異は同日開催のTOEIC Bridge®の結果にも見られた(表4)。2012年度が最も高く、2011年度、2013年度が続く順番もe-TACと同様であり、分散分析の結果2011年度と2012年度、また2012年度と2013年度の得点に0.05%の水準で有意差が認められた。

e-TAC、TOEIC Bridge® に共通して、2013 年度の入学時点の英語学力は 3 年のうちで最も低いといえる。e-TAC と TOEIC Bridge® の相関については後の章で述べる。

表 3 2011 年度～2013 年度 e-TAC の基礎統計量

年度	受験者数	平均値	最大値	最小値	中央値	標準偏差
2011	424	64.52	100	5	65.00	16.83
2012	413	66.74	100	20	70.00	14.57
2013	406	60.87	100	10	62.50	16.82
合計	1243	64.07	100	5	65.00	16.28

表 4 2011 年度～2013 年度 TOEIC Bridge® の基礎統計量

年度	受験者数	平均値	最大値	最小値	中央値	標準偏差
2011	425	142.12	178	92	144.00	14.98
2012	413	144.72	178	86	146.00	15.14
2013	406	141.00	174	72	144.00	16.70
合計	1244	142.65	178	72	144.00	15.65

8 e-TAC の信頼性

e-TAC 全 14 問中、4 肢択一形式のリーディング 8 問とリスニング 5 問について信頼性を測るため、折半法による Spearman Brown の係数、及び Cronbach の α 係数を求めた。なお、ライティング問題は採点方法と配点が異なるため、この分析対象からは除外した。結果は表 5 に示されるように、Spearman Brown の係数が 0.56 ～ 0.68、Cronbach の α 係数が 0.57 ～ 0.65 であった。これらの係数は、いかにテストの各項目に整合性があるかを示すもので、一般的に信頼性が認められるテストと見なすには 0.7 以上が必要とされている (e.g. Morgan, Leech, Gloeckner, & Barrett, 2013)。今回の結果がやや低い数値となったのは、13 問という少ない項目で「プロジェクト発信型英語プログラム」で必要となる能力を多角的に測ろうとしたことが影響したのであろう。池田 (1992) が 4 肢択一形式で項目数が 10 のテストの信頼性係数は 0.37 ～ 0.61 になり、項目数が 20 に増えると 0.54 ～ 0.76 になることが予想されると述べていることから判断すると、表 5 に見られる結果は予想の範囲内であるといえる。信頼性係数はテストの総得点と相関の低い項目を見直

表 5 2011 年度～2013 年度 e-TAC の信頼性統計量

年度	項目数	Spearman Brown の係数 (折半法)	Cronbach の α 係数
2011	13	0.61	0.61
2012	13	0.56	0.57
2013	13	0.68	0.65

す他に、項目数を増やすことで改善できるが（Hatch and Lazaraton,1991; Henning 1987）、プレイスメント・テストとしての e-TAC が単独ではなく、100 問の TOEIC Bridge® との併用を前提としたテストであることから、項目数を極端に増やすことは現実的ではない。

9 e-TAC の項目分析

4 肢択一形式問題の項目分析を行うため、項目難易度、項目弁別力指数、点双列相関係数を求め、表 6 に示した。なお、この章でのデータ分析には TDAP Ver. 2（大友、中村、秋山）²⁾ を主に使用した。

表 6 2011 年度～2013 年度 e-TAC の項目難易度、項目弁別力指数、点双列相関係数

	年度	リーディング								リスニング				
		Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7	Q. 8	Q. 9	Q. 10	Q. 11	Q. 12	Q. 13
項目難易度	2011	0.63	0.77	0.56	0.58	0.86	0.86	0.65	0.67	0.93	0.74	0.75	0.72	0.22
	2012	0.66	0.79	0.60	0.62	0.89	0.89	0.57	0.64	0.91	0.76	0.75	0.75	0.26
	2013	0.69	0.76	0.59	0.56	0.87	0.84	0.57	0.65	0.88	0.75	0.70	0.63	0.15
	全体	0.66	0.77	0.58	0.59	0.87	0.86	0.60	0.65	0.91	0.75	0.73	0.70	0.21
項目弁別力指数	2011	0.61	0.44	0.61	0.52	0.32	0.36	0.39	0.48	0.24	0.50	0.47	0.39	0.42
	2012	0.55	0.36	0.63	0.62	0.26	0.29	0.51	0.21	0.24	0.38	0.46	0.41	0.63
	2013	0.47	0.36	0.65	0.66	0.36	0.41	0.50	0.44	0.36	0.52	0.56	0.45	0.34
	全体	0.54	0.39	0.63	0.60	0.31	0.35	0.47	0.38	0.28	0.47	0.50	0.42	0.46
点双列相関係数	2011	0.47	0.46	0.50	0.43	0.42	0.47	0.35	0.41	0.43	0.44	0.44	0.35	0.39
	2012	0.47	0.37	0.50	0.46	0.33	0.39	0.42	0.24	0.46	0.37	0.47	0.38	0.51
	2013	0.42	0.34	0.49	0.48	0.44	0.51	0.42	0.39	0.55	0.51	0.47	0.41	0.37
	全体	0.45	0.39	0.50	0.46	0.40	0.46	0.40	0.35	0.48	0.44	0.46	0.38	0.42

9.1 項目難易度

項目難易度は、言い換えれば正答率のことで、0～1 の値を取る。集団の中での個人の位置を相対的に測る集団規標準準拠テストでは 0.3～0.7 が許容範囲とされる（Brown, 2005）。この基準と表 6 の結果を照らし合わせると、e-TAC はリーディング・リスニング共にやや易しめの問題が多いと言える。基準を大きく逸脱して易しすぎる項目はリーディング Q.5、Q.6、リスニング Q.9 の 3 問で、項目難易度は 0.84～0.93 であった。反対に、難しすぎる項目はリスニング Q.13 で、項目難易度は 0.15～0.26 であった。

9.2 項目弁別力指数

項目弁別力指数は、ある項目がテスト全体の測定する上位群と下位群を弁別する度合いを示すものであり、-1～1 の値を取る。表 6 に示した結果より、一般的に許容の下限とされる 0.2 を下回る項目はなかったものの、3 年を通じて何れの年も良い弁別力を持つ項目とされる 0.4 以上（Brown, 2005）を満たさなかった項目はリーディング Q.5 とリスニング Q.9 の 2 問であった。

この2問は項目難易度においても易しすぎると判断された項目である。難しすぎると判断された Q. 13 に関しては、平均点が低い 2013 年の項目弁別力指数は 0.34 に留まったものの、平均点が高い 2012 年は 0.63 というやや高めめの弁別力指数を示している。

9.3 点双列相関係数

点双列相関係数は、ある項目の正誤とテストの総得点との相関から上位群と下位群を弁別する度合いを示すものであり、 $-1 \sim 1$ の値を取る。3 年間全体の数値は、全項目が弁別力の高い項目とされる 0.25 以上 (Henning, 1987) の規準に達しており、全体的に見て点双列相関係数から見た項目弁別に重大な問題は見られなかった。年度別の数値で唯一規準に達していなかったのは 2012 年のリーディング Q. 8 で、0.24 という結果であった。Q. 8 は項目弁別指数、点双列相関係数共に平均点が高かった 2012 年のみ低い弁別力を示している。

9.4 錯乱肢の有効性

項目に対して錯乱肢がどの程度有効に機能しているかを分析するために、実質選択肢数、実質選択肢適切度、標準実質選択肢適切度を求め、表 7 に示した。実質選択肢数は選択肢が実質的には何肢分の働きをしたかを検討するもので (中村, 2002)、4 肢択一問題の場合 1～4 の値を取る。実質選択肢適切度は実質選択肢数を 0～1 の値を取るように、その適切さを求めたもので、それをさらに標準化したものが標準実質選択肢適切度 (中村, 2002) である。9.1、9.2 に述べたとおり、難易度、弁別度共に低い数値を取ったリーディング Q. 5、リスニング Q. 9 は実質選択肢数も全体で前者は 1.63、後者は 1.47 と低い数値になった。この 2 問は、錯乱肢を差し替えることで、項目の難易度、弁別力は改善されると思われる。難易度が高かったリスニング Q.13 は、全体の実質選択肢数は 3.40 と高い数字をとっているものの、実質選択肢適切度は 0.60、標準実質選択肢適切度は 0.30 と低い数値になった。

表 7 2011 年度～2013 年度 e-TAC の実質選択肢数、実質選択肢適切度、標準実質選択肢適切度

	年度	リーディング								リスニング				
		Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7	Q. 8	Q. 9	Q. 10	Q. 11	Q. 12	Q. 13
実質選択肢数	2011	2.60	2.05	3.12	2.54	1.66	1.68	2.72	2.70	1.37	2.06	2.19	2.39	3.33
	2012	2.56	1.98	3.03	2.40	1.56	1.52	3.06	2.77	1.45	1.92	2.18	2.27	3.36
	2013	2.43	2.15	3.03	2.57	1.66	1.73	3.12	2.77	1.60	2.06	2.41	2.83	3.43
	全体	2.55	2.06	3.07	2.50	1.63	1.65	2.98	2.76	1.47	2.02	2.26	2.49	3.40
実質選択肢適切度	2011	0.71	0.79	0.79	0.60	0.80	0.84	0.82	0.86	0.81	0.68	0.78	0.83	0.58
	2012	0.88	0.89	0.99	0.71	0.91	0.87	0.95	0.95	0.95	0.73	0.89	0.97	0.79
	2013	0.77	0.81	0.82	0.59	0.85	0.78	0.82	0.84	0.83	0.71	0.78	0.83	0.61
	全体	0.75	0.80	0.82	0.60	0.83	0.81	0.82	0.84	0.84	0.68	0.79	0.84	0.60
標準実質選択肢適切度	2011	0.44	0.53	0.54	0.32	0.54	0.59	0.56	0.62	0.56	0.40	0.53	0.58	0.30
	2012	0.49	0.51	0.63	0.30	0.53	0.48	0.58	0.58	0.58	0.32	0.51	0.60	0.39
	2013	0.49	0.55	0.56	0.27	0.59	0.51	0.56	0.59	0.57	0.42	0.51	0.57	0.31
	全体	0.47	0.54	0.56	0.30	0.57	0.55	0.55	0.59	0.59	0.40	0.52	0.58	0.30

表8はQ.13において受験者を上位群、中位群、下位群に分け、各群がどの選択肢を選択したか、その割合を示したものである。上位群は過半数を超える受験者が正答を選択しているが、中位群から正答率は格段に低くなる。また、全体の50.7%、上位群の33.3%が錯乱肢Cを選択していることから、Cは受験者を引き付ける力を持った選択肢であるといえる。逆の見方をすれば、正答Aがより適切であるものの、錯乱肢Cにも正答と同じ方向性を示す表現が用いられていたため、不正解者の過半数がCを選択した可能性もある。錯乱肢C、正答Aに使われている表現や、選択肢と問題文の整合性を再度見直し修正することで、Q.13の弁別力も改善されるだろう。

表8 2011年度～2013年度 e-TAC Q.13 各選択肢の選択割合

受験者の水準	選択肢				
	A (正答)	B	C	D	無回答
上位	51.8%	6.8%	33.3%	8.0%	0.0%
中位	11.9%	17.3%	54.9%	15.9%	0.0%
下位	5.1%	17.9%	61.0%	13.7%	0.3%
全体	20.8%	14.6%	50.7%	13.2%	0.1%

10 e-TAC と TOEIC Bridge® の関係

10.1 相関係数

2011年から2013年の受験者のデータを合わせてe-TACとTOEIC Bridge®の相関を検証した結果、 $r=.65$ のやや強い相関が見られた。結果を散布図で表し、図4に示した。

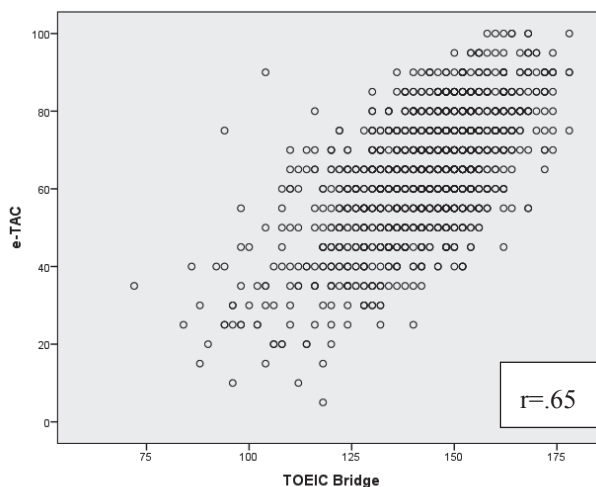


図4 2011年度～2013年度 e-TAC と TOEIC Bridge® の相関 (散布図)

e-TAC、TOEIC Bridge® の各パートの相関係数は表9に示されたとおり、ライティングを除い

た全ての項目に $r=.40$ を超える中程度の相関が見られた。特に e-TAC リーディングと TOEIC Bridge® リーディング、e-TAC リスニングと TOEIC Bridge® リスニングのように同じスキルを測るパートの相関係数は $r=.50$ を超え、相関はより強くなる傾向があった。以上のことから、e-TAC と TOEIC Bridge® は総括的には同じ性質の力を測るテストであると言える。e-TAC ライティングについては、TOEIC Bridge® リーディングとは $r=.30$ 、リスニングとは $r=.34$ の弱い相関しか見られなかった。e-TAC ライティングの問題は決まった日本語を英語に翻訳するのではなく、自分のアイデアを英語で説明することが求められる。これは唯一、受験者の発信能力が問われる問題であり、リーディング、及びリスニングパートとの方向性の違いは明らかである。得点の傾向に違いが見られるのは当然のことと言えよう。

表 9 2011 年度～2013 年度 e-TAC と TOEIC Bridge® の総得点、及び各パートの相関係数

	TOEIC Bridge® リーディング	TOEIC Bridge® リスニング	TOEIC Bridge® 合計
e-TAC リーディング	0.58	0.40	0.54
e-TAC リスニング	0.49	0.50	0.53
e-TAC ライティング	0.30	0.34	0.35
e-TAC 合計	0.61	0.57	0.65

11 TOEIC Bridge® のレベル別に見た e-TAC の得点分散

e-TAC 実施前の 2008 年度は TOEIC Bridge® の総得点に基づいてプレースメントを行っていたことを踏まえ、2011 年度から 2013 年度のプレースメント・テスト受験者を TOEIC Bridge® の総得点に基づいて上位群、中位群、下位群にレベル分けし、各群における e-TAC の得点分布を確認した。その結果は図 5 に示したとおりである。まず、e-TAC 総得点の分布は各レベルとも散らばりが大きく、2011 年度、2012 年度は上位群に下方向の外れ値が見られる。これは、TOEIC Bridge® では上位群に属する学生が、「プロジェクト発信型英語プログラム」の一環として開発された e-TAC では、属する水準から外れた得点を取得していることを示している。下方向の外れ値はパート別に見たリーディングとリスニングの結果にも散見される。e-TAC を用いることで、日常的で比較的平易な内容から出題している TOEIC Bridge® では高得点を取得したものの、サイエンスや大学での授業を想定した英語となると自分が属する集団の水準に達しない学生を、より適切なクラスに配置できる可能性は高まる。ただし、リーディングとリスニングでは下位層ほど散らばりが大きくなっていることには注意すべきであろう。8 で述べたように、各パートの項目数が少ないため、当て推量で高得点を取得した学生が生じ、その結果として下位群の分布が上向きにも広がっていることが示唆される。最後に、ライティングには、10 でも述べた TOEIC Bridge® との傾向の違いがボックスプロットからも確認できる。最も極端な傾向は 2011 年の結果に見られる。中位群と上位群のボックスプロットは全く同じ形をしており、最大値、最小値、中央値の違いが認められない。下位層はボックスが下に伸びるものの、最大値は同じ値（満点）である。

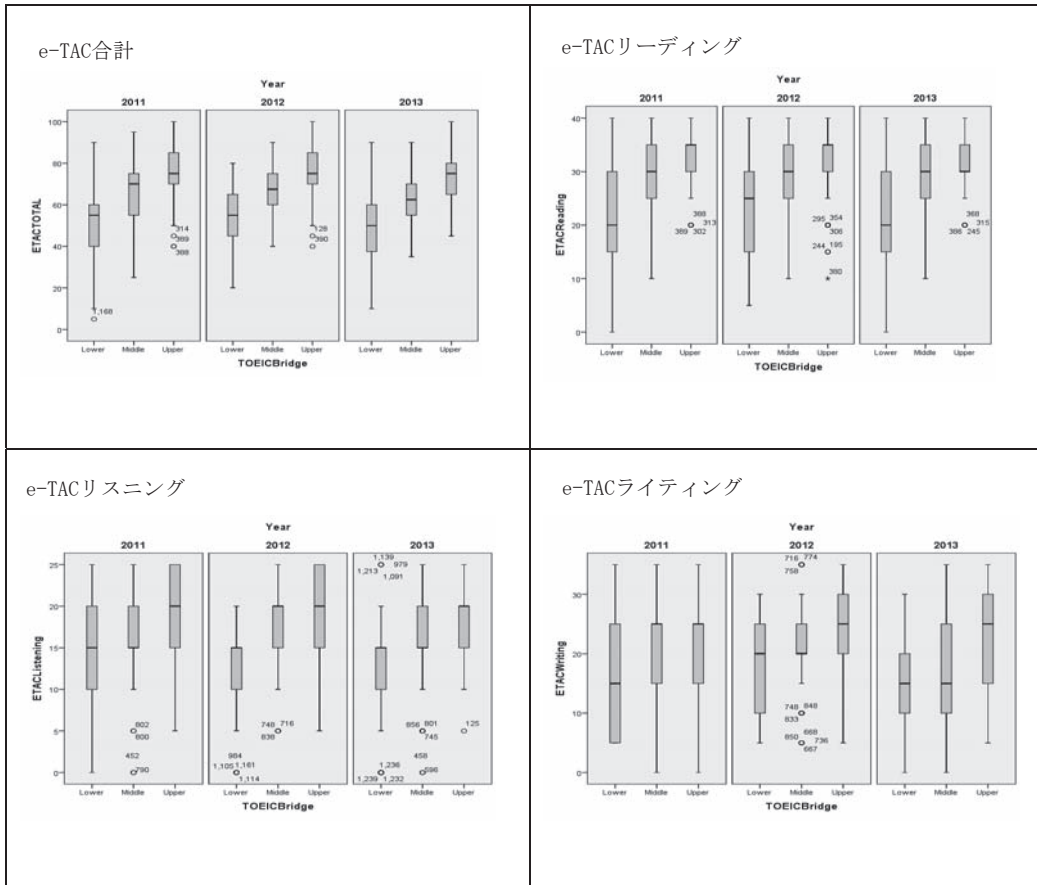


図5 TOEIC Bridge® のレベル別に見た e-TAC の合計、及び各パートのボックスプロット

12 e-TAC の改善へ向けた考察

まず、信頼性については項目数を増やすことが肝要であるが、英語のプレイズメント・テストの実施時間が全体で150分を要している現状から判断すると、項目増加のために時間を延長するのは避けるべきであろう。しかしながら、4肢択一問題1問に対するe-TACの解答時間はTOEIC Bridge®の3倍以上である。アカデミック・コンテストの問題に取り組むe-TACでは解答により時間が掛かるのは当然のことだが、9で述べたようにe-TACの項目の大半は易しめであると判断されるため、受験生が同問題をより短い時間で解答できる可能性は十分にある。実施時間は現状を維持し、項目数を増やした新たなバージョンへと改定することが、信頼性を向上させる上での1つの解決策である。次に、項目分析を振り返ると、e-TACがTOEIC Bridge®の天井効果と床面効果を考慮して開発された経緯を踏まえるならば、項目難易度が許容範囲に留まらない項目があるのは必然であると言えよう。しかし、項目難易度と項目弁別力の両者が許容範囲を逸脱していた項目、並びに受験者の水準により弁別力が大きく影響を受けた項目は改善すべきで

ある。具体的には 9.4 で述べたように錯乱肢を見直し、問題文と選択肢の整合性を高めることにより難易度、弁別力共に向上するだろう。最後に、e-TAC は TOEIC Bridge® とやや強めの相関を見せながらも、ライティングに限っては TOEIC Bridge® と関係が浅いことがわかった。つまり、ライティング問題は e-TAC を TOEIC Bridge® とは一線を画すテストにする主要因であると言える。TOEIC Bridge® 下位群の中にも上位群に負けず劣らず評価の高いエッセイを書いた学生がおり、またその逆も同様であった。e-TAC ライティングに関しては、学生の「スキル・ワークショップ」でのテスト結果、及び 2013 年度 9 月に実施される TOEIC SW® の結果と併せて分析することで、改善に向けた新たな発見があるだろう。

13 おわりに

英語能力、そして英語コミュニケーション能力をいかにして評価するのか、これは現代の英語教育が直面し、そしてこれからも直面していかなければならない難題である。しかも、いわゆる英語 4 技能の中の「聞く」、「読む」という受信型スキルのテストに比して、「話す」、「書く」という発信型スキルの評価モデルについては、未だ開発・研究の余地が大きいことは、大方異論がないものと思われる。こうした中、e-TAC は「プロジェクト発信型英語プログラム」におけるプレイズメントモデルとして機能し、評価モデルとしての妥当性を備え始めた。今後の課題として、プレイズメント時における e-TAC の信頼性や妥当性を高めることもさることながら、「プロジェクト発信型英語プログラム」における学習者の総合的な能力評価が可能なモデルとして、新たなバージョンを開発していくことが必要となろう。すなわち、入口（＝プレイズメント）だけでなく、出口（＝プログラム受講後の評価）における評価システムとして機能することが目指されるべきである。

大胆に変化する社会の変化を柔軟に取り入れ、現場で役に立ち、学習者を成長させることができるテスト開発として、今後も e-TAC モデルを継続的に深化、発展させていきたい。

注

- 1) 鈴木は 2000 年度～2005 年度にアドバイジングをしていた千葉商科大学制作情報学部用の入学試験（英語）の Part 2、Part 3 で 2 に述べたモデルの問題を作成したが、生命科学部・薬学部のプレイズメント・テストでは時間の制約上 presentation/discussion/debate 能力のテストは実行していない。
- 2) TDAP (Test Data Analysis Program) は『テストで言語能力は測れるか言語テストデータ分析入門』（中村, 2002）に添付されている。

参考文献

- 池田央『テストの科学—試験にかかわるすべての人に』日本文化科学社、1992 年。
- 鈴木佑治『英語教育のグランド・デザイン：慶應義塾大学 SFC の実践と展望』慶應義塾大学出版会、2003 年。
- 鈴木佑治「立命館大学生命科学部・薬学部『プロジェクト発信型英語プログラム—Project-based English Program』の理論的基盤と実践」『立命館高等教育研究』第 10 号、2010 年、43-61 頁。
- 鈴木佑治『グローバル社会を生きるための英語授業：立命館大学 生命科学部・薬学部・生命科学研究科

プロジェクト発信型英語プログラム』創英社／三省堂書店、2012年。

中村洋一『テストで言語能力は測れるか—言語テストデータ入門』桐原書店、2002年。

Brown, J. D., *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*, New York: McGraw-Hill, 2005.

Hatch, E., & Lazaraton, A., *The Research Manual: Design and Statistics for Applied Linguistics*, New York: Newbury House, 1991.

Henning, G., *A Guide to Language Testing: Development, Evaluation, Research*, Cambridge, MA: Newbury House, 1987.

Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C., *IBM SPSS for Introductory Statistics: Use and Interpretation (5th ed.)*, New York: Routledge/Taylor and Francis, 2013.

Development and Implementation of “English Test in Academic Context” (e-TAC) as a Placement Test for “Project-based English Program” in the College of Life Sciences and Pharmaceutical Sciences:

Achievements and Challenges

KONDO Yukie (Foreign Language Lecturer, Language Education Center, Ritsumeikan University)

YAMANAKA Tsukasa (Associate Professor, College of Life Sciences, Ritsumeikan University)

English Test in Academic Context (e-TAC) is used as a part of the placement test for English classes in the “Project-based English Program” in the College of Life Sciences and Pharmaceutical Sciences. During the first year of the program, in 2008, TOEIC Bridge® was the only placement test; however, after pursuing better placements for lower and upper level students, e-TAC was developed and started to be used in 2009. Through revisions on the test and its implementation method, e-TAC has been fully established for all the students in the program since 2011 and is now used in conjunction with TOEIC Bridge®. This paper reports on the development and implementation of e-TAC and discusses the improvements from the classical test analysis and a correlation analysis with TOEIC Bridge®.

Keywords

Placement test, Program assessment, Project-based English Program, e-TAC