

別府大学語学教育におけるプレイスメントテストの活用 —2014年度分析結果から見える今後の課題—

竹安大・三重野佳子・船田佐央子・内山和也・松田美香

1. はじめに

別府大学の日本語および教養英語の授業では、それぞれプレイスメントテストによる習熟度別クラス編成授業を行っており、平成24年度から日本語及び英語教員グループの有志でプレイスメントテストを効果的に活用するための共同研究を実施している。本稿は、こうしたプロジェクトのうち、教養英語グループによる取り組みとして報告した竹安ほか(2014a, 2014b)のその後の経過を報告するものである。

別府大学教養英語の必修の授業¹⁾では、2009年度より入学時にプレイスメントテストを実施して、習熟度別クラス編成による授業を行っている。2012年度からは、このプレイスメントテストを教育効果の測定や学生・教員へのフィードバックにも利用するプロジェクトが始動し、教養英語授業におけるPDCAサイクルの確立や問題点の可視化を目指して取り組んでいるところである。2014年度には、竹安ほか(2014a, 2014b)で報告済みのプレイスメントテストを活用した教育効果測定の取り組みを継続して実施することに加えて、2013年度より導入したe-learning教材(ALC Net Academy 2)の学生の学習時間と、学生の英語力の伸びの相関や、TOEIC IPテストを複数回受験した学生のスコアの伸びを調べるなど、教養英語授業の教育効果を測定して授業改善に役立てるための取り組みを新たに行った。本稿では、こうした一連の取り組みの結果と、そこから明らかとなった課題について報告する。

2. プレイスメントテストを活用した

教育効果の測定

別府大学の教養英語の授業では、必修の授業の習熟度別クラス編成のために入学時に実施したプレイスメントテストとその後に実施する同形式

のテストの結果を比較することによって教育効果の測定を実施している。このような形での教育効果の測定を始めたのは2012年度からであるが、2012年度は年に2回(4月のプレイスメントテストと7月の前期末共通テスト²⁾)、2013年度と2014年度はそれぞれ年に3回(4月のプレイスメントテスト、7月の前期末共通テスト、12月の後期末共通テスト)のペースで学内統一テストを実施した。テストの結果は、英語教育に関する自己点検の一環として、教員に対してフィードバックしたほか、2013年度以降からは後期末の授業で1年間の学習成果を振り返ってもらう材料として、学生にも個人の3回分のテストの成績を返却した(フィードバック内容については竹安ほか(2014b)参照)。

以下では、教養英語における自己点検の一環として、2014年度に実施した3回のテストの結果を報告し、それ以前の結果との比較を行いながら今後改善していくべき点を議論する。なお、過去の結果との比較がしやすいように、2014年度の3回のテスト作成や分析は竹安ほか(2014a, 2014b)で報告されたものと同様の手法を用いている。先行研究を検証しつつ新たな視点を加えていくことにする。

2.1. テストの構成

別府大学教養英語のプレイスメントテストおよび共通テストはいずれも50問の空所補充問題から構成されており、高校までの語彙・文法を問う問題である。50問中、語彙に関する問題が10問程度、文法事項に関する問題が40問程度となっており、テスト全体としては高校までの文法を理解できているかどうかを主に把握するための構成となっている。

項目応答理論による等化を実施するため、前

年度までと同様、50問中2割程度はテスト間比較のための共通項目とし、残りの40問は新規に作成された問題とした。各テスト実施の所要時間は30分程度、受験者数は500名前後である³⁾。

2.2. 分析方法

プレイスメントテストと共通テストは問題形式や問題数は同じであるが、個々の設問内容が異なるため、難易度が等しいという保証はない。そこで、項目応答理論(2母数モデル)により等化を行い、テスト間の結果を比較した。なお、教養英語のプレイスメントテスト/共通テストに関する分析で主に考慮しているのは、1次元性、項目パラメータ(項目難易度、傾き)、受験者の潜在能力値(θ)である(個々の事項の詳細については、Embretson & Reise (2000)、de Ayala (2009)など項目応答理論に関連する文献を参照のこと)。分析にはEasyEstimation(熊谷2009)を用い、推定はconcurrent calibration法により行った。

2.3. 分析対象

プレイスメントテストの受験生は基本的に1年生だけであるが、7月と12月の共通テストについては、TOEICの得点との相関を調べるためにTOEIC関連の選択授業でも実施されているため、2年生以上も受験する場合がある。本稿では、項目パラメータや受験者の潜在能力値はすべての受験者の解答に基づいて推定し、必要に応じてそれらの推定値を用いた統計分析を実施した⁴⁾。

2.4. 結果

2.4.1. テストの妥当性

どのような教育方法を用いるにしても、その教育効果の測定のためには妥当性の高いテストを用いて評価することが不可欠である。竹安ほか(2014a, 2014b)にて報告したとおり、別府大学教養英語ではTOEIC等の外部試験ではなく、自作の学内統一テストを作成して習熟度別クラス編成や教育効果の測定を行っているため、用いたテストの妥当性をしっかりと検証し、問題点を改善していく仕組みを整えることが望ましい。このような観点から、受験者の潜在能力値の変化と教

育効果の測定の議論の前に、テストそのものの妥当性に関する分析結果を提示する。

2.4.1.1. 一次元性

テストの一次元性の確認は、テスト作成/実施者が測定しようとしている能力をそのテストが本当に測定できているかを評価するうえで重要である。一次元性が確認できれば、そのテストは全体としてある一つの能力(例:英語の能力)を測定していると判断できるのに対し、一次元性が確認できない場合には、そのテストが全体として複数の能力(例:英語の能力と情報処理能力)を測定していることを示し、テスト結果に調べたいこととは無関係の要素が紛れ込んでいることが示唆される。

本稿では、スクリープロットをもとにして一次元性の判断を行った。スクリープロットとは、因子分析など、得られたデータから潜在的な共通因子を探るための分析において、因子数決定の判断のために用いられる手法の一つであり、固有値(eigenvalue)の大きい成分から順にならべて図示することにより、第一成分およびそれ以降の各成分がデータのばらつきを説明できる度合いをわかりやすく示したものである。実施したテストの一次元性に問題がなければ、第一成分の固有値のみがそれ以外の成分の固有値よりも著しく高くなるため、スクリープロットはL字型に近い形状を取ることが予想される。

2012年度4月から2014年度12月までの8回のテストのスクリープロットを図1(次頁)に示す。いずれについても、かなりL字型に近い形を描いており、テストの一次元性には問題がないと判断できる。

2.4.1.2. モデルの適合度とテスト総情報量(Total I(θ))

項目応答理論による分析においては、モデル(ここでは2母数モデル)がデータによく当てはまるとは限らない。モデルの適合の度合いが悪ければ、テストの結果から推定される潜在能力値も不安定なものとなり、これらの値を用いて行う教育効果の測定の妥当性にも影響を及ぼしてしまう。従って、テストの妥当性の確認の一環として、モデル

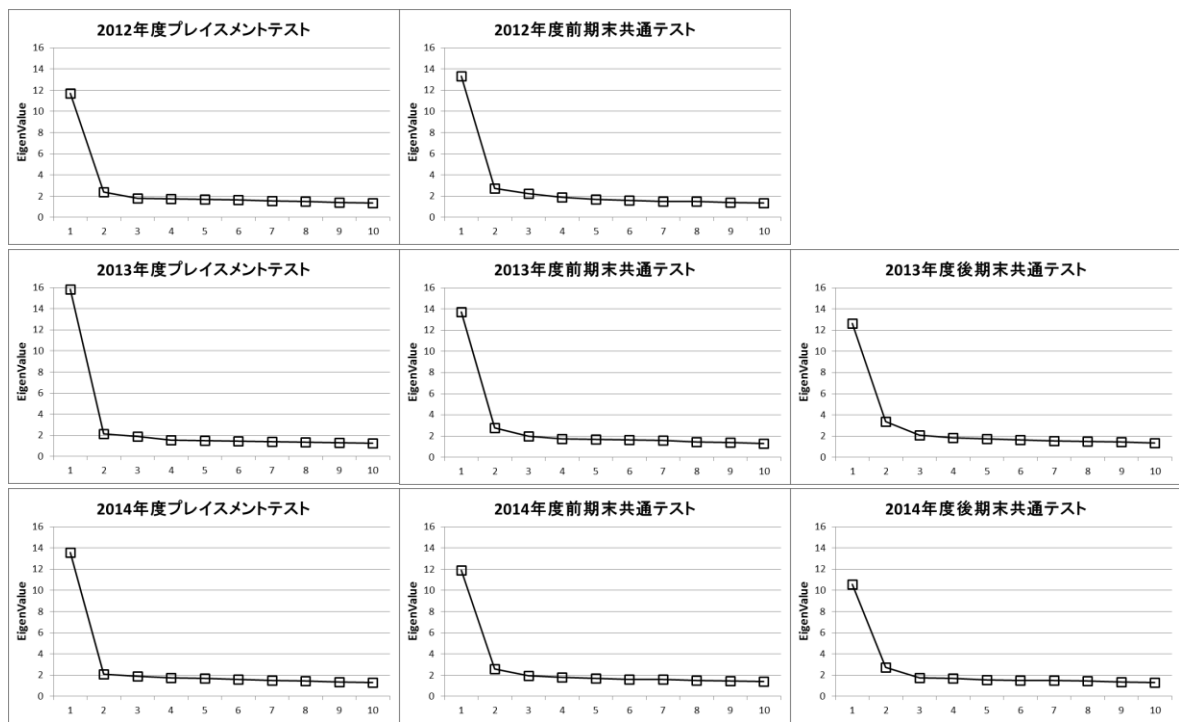


図1 過去のプレイズメントテスト・共通テストのスクリープロット

のデータに対する適合度を確認する必要がある。本稿では、de Ayala (2009: 51-53) で紹介されている OutFit (MnSQ) と InFit (MnSQ) の値（項目に関する平均値）をもとに、これまでに実施した8回のテストの適合の度合いを判断することとした。また、現在別府大学教養英語ではテスト実施のための項目バンクの構築と、テストの質の向上に取り組んでいるところであるため、テスト全体の有効性を把握するために、各テストのテスト総情報量も求めた。

各テストの OutFit (MnSQ)、 InFit (MnSQ) およびテスト総情報量の値は表1のとおりである⁵⁾。

これらの値はともに期待値が1で、1よりも大きい場合にはデータがモデルの予測とうまく適合していない (misfit) ことを示すのに対し、1よりも小さい場合には項目間依存などによりデータがモデルに過剰に適合している (overfit) ことを示す。de Ayala (2009: 53) によれば、これらの値が0.5～1.5の範囲内であれば問題ないと見なすことができるため、 OutFit (MnSQ)、 InFit (MnSQ) のどちらを用いた場合でも、2013年4月のInFit値が0.5をわずかに下回った点を除けば、モデルの適合度におおむね問題はないと判断できる⁶⁾。また、テスト総情報量については、2012年～2013年に実

表1 2014年度後期末共通テスト終了時点でのモデル適合度、テスト総情報量および平均潜在能力値

テスト実施日	受験者数	テストの項目数	OutFit (MnSQ)	InFit (MnSQ)	TOTAL I(θ)	潜在能力値(θ)の平均
2012年4月	446	50	0.805	0.649	50.94	-0.089
2012年7月	421	50	0.720	0.558	54.56	0.003
2013年4月	457	50	0.686	0.492	60.80	0.028
2013年7月	427	50	0.789	0.625	50.79	0.160
2013年12月	437	50	0.750	0.616	55.71	0.192
2014年4月	593	50	0.846	0.671	50.22	-0.076
2014年7月	425	50	0.898	0.730	41.93	0.062
2014年12月	575	50	0.886	0.744	41.56	-0.007

施したテストに比べて、2014年に実施した3回のテストの総情報量は相対的に低いものであった。現状では問題のうち8割が新規に作成された問題であるため、事前の難易度の予測が難しいことに加えて、2014年に実施したテストの問題作成の際には、項目バンクの問題にバリエーションを増やすため、これまでの問題とは少し傾向の異なる問題となるように各教員が問題を作成した経緯があり、これがテスト総情報量の値に影響した可能性がある。しかしながら、項目バンクに登録された問題数が増えてきたこともあり、今後は項目の識別力や難易度が判明している問題の出題割合を高めていくことで、テスト間のばらつきを少なくし、より情報量の高いテストを構築しやすくなっていくものと思われる。

2.4.1.3. 個々の項目の難易度、傾き（識別力）

別府大学教養英語では、プレイスメントテストや共通テスト実施の際の共通項目候補を増やすため、テスト実施のたびに40問の新問題を作成し、

項目バンクを構築してきた。結果として、2014年12月の後期末共通テストの分析終了時点で、不備のあった問題を除く308問の項目が実施済み項目として登録されることとなった。

教養英語における統一テスト実施の目的は、習熟度別クラス編成と、その後の教育効果の測定であるため、難易度の低い問題から高い問題まで、様々な難易度の問題を共通項目として用いることができることが望ましい（問題作成の際には、難易度が-2～+2の範囲に収まるように意図した作問を依頼している）。また、個々の学生の能力についてより正確な推定が可能となるよう、傾きの値がある程度高い問題が揃っている必要がある。こうした当初の目標・期待に対して、現状がどのような状態であるかを以下で簡単に報告し、今後改善すべき点を議論する。

各項目の難易度と傾きに関する代表値と散布図を表2と図2にそれぞれ示す。散布図からも明らかのように、難易度については比較的広い範囲に値が分布しており、7～8割程度の項目が-

表2 過去問（308項目）の項目難易度、傾き（識別力）に関する代表値

	平均値	SD	中央値	最大値	最小値
項目難易度	0.362	1.708	-0.015	5.971	-5.094
傾き（識別力）	0.949	0.526	0.885	3.842	0.103

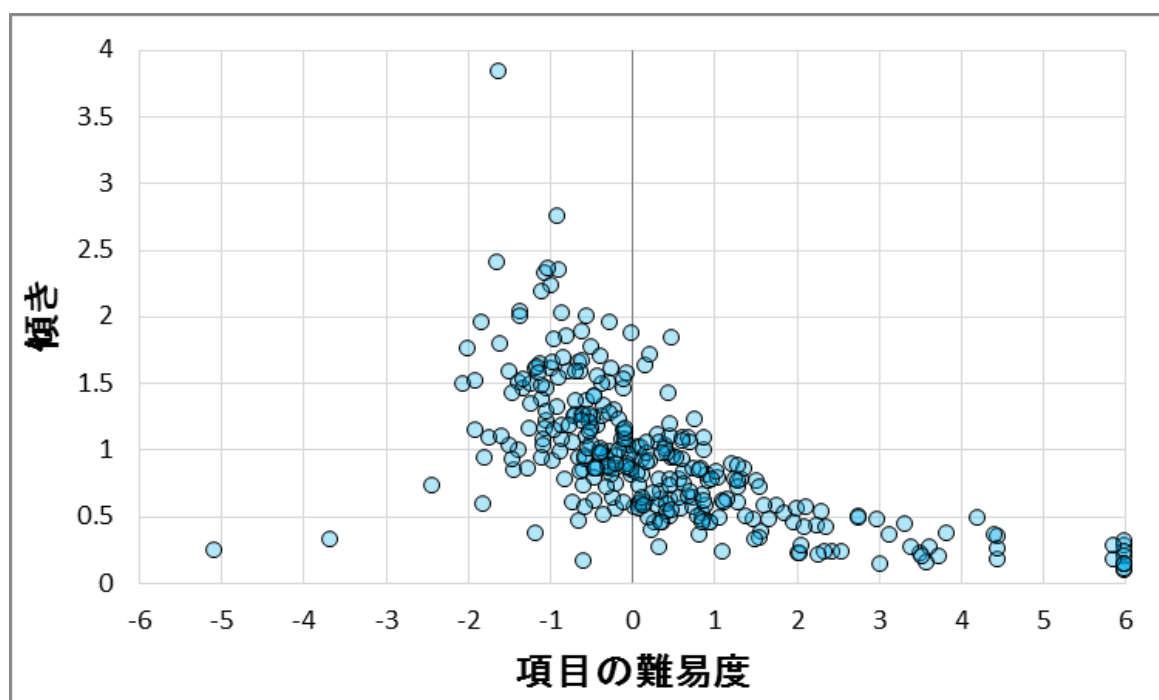


図2 過去問（308項目）の項目難易度と傾きの散布図

2～+2の範囲に収まっていることから、使用目的におおむね合致した項目のリストを作成できていると判断する。また、難易度の値が低すぎるまたは高すぎる項目については、共通項目としての利用は望めないが、一方でこれらの項目は学生たちがすでに完全に理解しているまたは全く理解できていない内容であることを示しているため、これらの設問で使用されている語彙や文法事項を詳細に検討していくことで、学生に対するより効果的な指導につなげることができる可能性がある。

傾きについては、難易度が低すぎるまたは高すぎる項目において低くなる傾向が予測され、散布図からもその傾向が読み取れるが、その一方で難易度が中程度の範囲においても傾きの値がそれほど高くない項目も見られる。こうした項目については、設問内容そのものが良くないという可能性と、扱っている語彙・文法事項が長期間かけて習得される性質のものであるという可能性が考えられるが、個々の項目に関する詳細な検討は現時点では出来ていない状態であり、今後の検討課題である。

2.4.1.4. 外部テストとの相関分析

別府大学教養英語におけるプレイスメントテスト・共通テスト（以下、まとめて学内テストとする）の妥当性の検証の一環として、これらのテストと外部テスト（TOEIC IPテスト）の結果の相関を調べた。

別府大学では、2012年度より年に2回（7月頃と12月頃）TOEIC IPテストを実施しており、TOEIC IPテストが実施される時期がちょうど前

期末もしくは後期末共通テストが実施される時期と近い（表3参照）。従って、これらのTOEIC IPテストを受け、かつ直近の学内テストを受けた学生の成績を対象として相関分析を実施することで、TOEICとの相関を調べることができる。

TOEIC IPテストの得点 (Total Score) と、直近の学内テストの θ の値 (307ペア) についてピアソンの積率相関係数を求めたところ、 $r=0.665$ ($t=15.562$, $df=305$, $p<0.001$) であった⁷⁾ (散布図を次頁図3に示す)。これは必ずしも強い相関であるとは言えないが、TOEIC IPテストがリスニング、リーディングから構成されており、幅広い内容の問題が出題されるのに対し、教養英語の学内テストはリスニングの出題が無く、高校までの語彙・文法に関する短文の空所補充形式のみで構成されていることを考慮すれば、相関の度合いとしては想定の範囲内であると思われる。

2.4.2. 教育効果の測定

2.4.2.1. 学内テストでの潜在能力値 (θ) の伸び

竹安ほか (2014b) では、必修の教養英語授業の教育効果を測定するために、プレイスメントテストと共通テストをいずれも受験した学生について、各テストにおける θ の値を比較したところ、2012年度、2013年度ともに入学時から前期末の間に θ の値に有意な伸びが見られたことを報告した。

以下では、2014年度にプレイスメントテスト・前期末共通テスト・後期末共通テストをすべて受験した学生の θ 値について、昨年と同様の分析を実施した結果を以下に報告する。なお、2014年

表3 相関分析における別府大学学内テストとTOEIC IPテストの対応関係

TOEIC IP テスト		直近の学内テスト	TOEIC IP テストと共通テストを共に受けた受験者の数
実施日	受験者数		
2012年6月30日	11	2012年度前期末共通テスト	2
2012年12月2日	69	なし (実施せず)	—
2013年7月13日	126	2013年度前期末共通テスト	92
2013年12月21日	87	2013年度後期末共通テスト	62
2014年7月12日	126	2014年度前期末共通テスト	89
2014年12月20日	78	2014年度後期末共通テスト	62

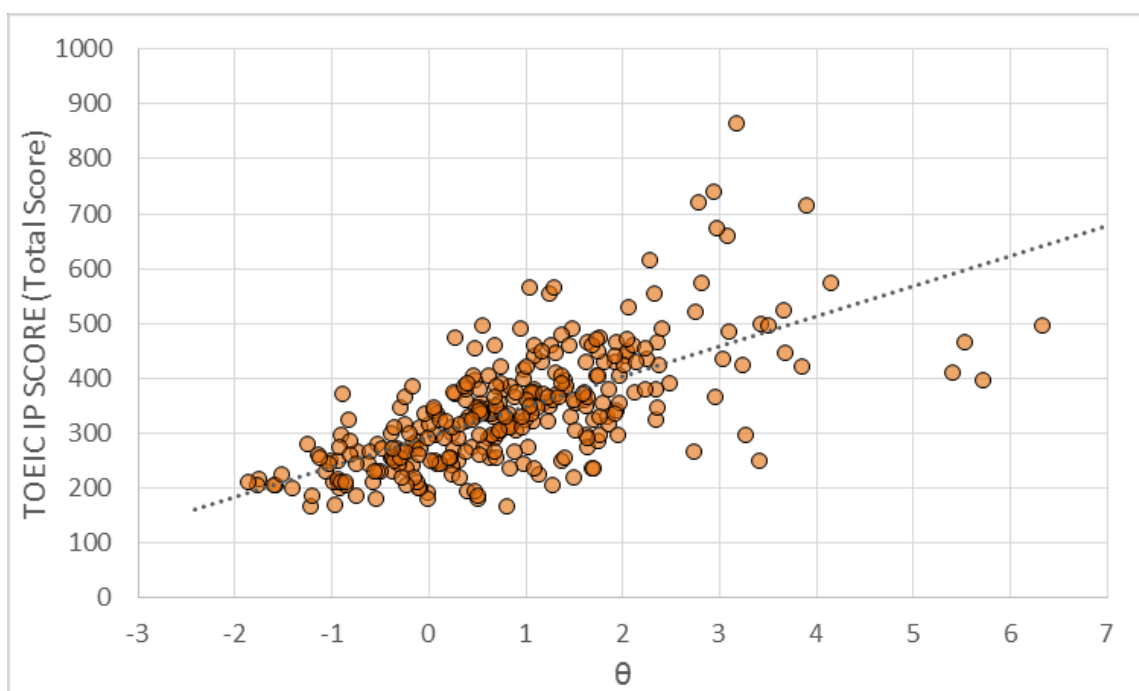


図3 TOEIC IPテストの得点と共通テスト結果（ θ ）の散布図

度は別府大学短期大学部の学生も初年次英語教育の教育効果測定プロジェクトに参加しているが、短期大学部の学生は前期末共通テストを受験していないため、この節での分析の対象からは外れている。同様に、外国人留学生についても、日本語の授業が語学の単位として認められる関係で必修英語授業は半期しか受けない場合が多く、ほとんどが分析の対象から外れることとなった。従って、この節での分析の主な対象となっているのは別府大学の2014年度入学の日本人学生ということになる。この節における θ の平均値などの値が表1のものとは異なるのは、こうした事情のためである。

2.4.2.2. 2014年度の結果

プレイスメントテスト時点での θ の平均値は0.138であったのに対し、前期末時点では0.084、後期末時点では0.152であった⁸⁾。反復測定ANOVA（SPSS混合分析（線形）モデルによる）で θ の値の変化を分析したところ、3つの θ の平均値の間には有意な差が観察されなかった（ $F(2, 612)=1.576, p=0.208$ (n.s.), 偏 $\eta^2=0.005$ ）。このことから、2014年度には1年間を通して学生の英語力に変化が確認できなかったということになる⁹⁾。

2.4.2.3. 過去の結果との比較

別府大学教養英語では、プレイスメントテスト（入学時）と2回の期末共通テスト結果を比較し、プレイスメントテストから前期末共通テストの間の伸びから前期授業の教育効果を、また、前期末共通テストから後期末共通テストの間の伸びから後期授業の教育効果を測定し、自己点検の際の目安の一つとしてきた。2012年度から2014年度までの教育効果測定結果をまとめると、表4のようになる。

表4 潜在能力値（ θ ）の値の変化

	入学時～前期末	前期末～後期末
2012年度	上昇	
2013年度	上昇	変化なし
2014年度	変化なし	変化なし

2013年度までとは異なり、2014年度に学生の潜在能力値に変化が観察されなくなったということは、2014年度の教養英語の必修授業による教育効果が弱くなっている可能性を示しており、授業に携わる教員としては残念な結果である。この点については、授業へのe-learning学習導入の効果や、外部テストに基づく学生の能力の伸びの分析結果と含めて、後ほど議論する。

2.4.3. 授業へのe-learning学習導入の効果

2013年度には、学生の授業外での学習促進のため、同年度に導入されたe-learning教材（ALC Net Academy 2）を後期から必修の教養英語授業に組み込み、学生に対して毎週指定されたユニットを学習することと、その内容に関する授業中の小テスト（語句の書き取り）を課したが、2014年度からはその取り組みを拡充し、1年間（前期・後期）を通して、必修の教養英語授業において同様の課題を課し、さらに半期の間に学習対象とするユニットと小テストの数を増やした（詳細は表5参照）。つまり、2014年度は前年度までと比べて学生が授業時間外学習をしている時間が圧倒的に多くなっていると考えられるため、1年間のθの値の伸び方もより大きくなることが期待されるところである。

こうした学習環境の整備の取り組みにもかかわらず、2014年度に学生の英語力に伸びが観察されなくなってしまうことは大きな問題である。表5は必修英語授業における2012年度からのe-learning教材利用状況をまとめたものであるが、この表と表4との対応関係を見ると、むしろe-learning教材を利用しない方が学生の能力が伸びているようにも解釈できる結果が得られたようにも見える。もちろん、これが直ちにe-learning教

材を用いた課題や教材そのものが無意味であるという議論につながるわけではないが、少なくとも、課題の内容が妥当であったのかを検証し、今後授業におけるe-learning教材の活用方法の見直すことも検討してみる必要があることが示されたと言える。この点について、e-learning教材に関する学生へのアンケート結果や、各学生のe-learning学習時間などから、今後解決すべき点を明らかにし、授業改善への一歩としたい。

2.4.3.1. e-learning学習に関するアンケート

2014年度後期末テスト実施時に、e-learningを用いた課題に関するアンケートを実施した。アンケートで学生に尋ねた項目、それぞれの項目に対する学生の回答は以下のとおりであった。

1. あなたは、この1年間、ネットアカデミーにきちんと取り組めましたか。

1. きちんと取り組めた	98
2. まあまあ取り組めた	135
3. あまり取り組めなかった	65
4. ほとんど取り組めなかった	18
2. あなたにとって、ネットアカデミーの「リーディング」の難易度はどうでしたか。

1. 難しかった	36
2. 少し難しかった	134
3. ちょうどよかった	145
4. 簡単だった	3

表5 e-learning教材を用いた授業時間外学習と、その内容に沿った授業中の小テストの実施状況

		前期	後期
2012年度	e-learning教材を用いた授業時間外学習	実施せず	実施せず
	時間外学習内容にリンクした授業中の小テスト	実施せず	実施せず
2013年度	e-learning教材を用いた授業時間外学習	実施せず	英語2にて実施 (計5ユニット)
	時間外学習内容にリンクした授業中の小テスト	実施せず	英語2にて実施 (5回)
2014年度	e-learning教材を用いた授業時間外学習	英語1・3にて実施 (計20ユニット)	英語2・4にて実施 (計20ユニット)
	時間外学習内容にリンクした授業中の小テスト	英語1にて実施 (10回)	英語2にて実施 (10回)

※授業時間外に学習するユニットについては、習熟度別クラスごとに個別に指定

3. あなたにとって、ネットアカデミーの「リスニング」の難易度はどうでしたか。
- | | |
|-------------|-----|
| 1. 難しかった | 45 |
| 2. 少し難しかった | 123 |
| 3. ちょうどよかった | 142 |
| 4. 簡単だった | 8 |
4. あなたにとって、ネットアカデミーの小テストは英語力の向上に役立ちましたか。
- | | |
|---------------|-----|
| 1. 非常に役立った | 37 |
| 2. 少し役立った | 225 |
| 3. あまり役立たなかった | 49 |
| 4. 全然役立たなかった | 8 |
5. あなたにとって、ネットアカデミーの小テストの難易度はどうでしたか。
- | | |
|-------------|-----|
| 1. 難しかった | 30 |
| 2. 少し難しかった | 159 |
| 3. ちょうどよかった | 122 |
| 4. 簡単だった | 7 |
6. 大学に入ってから1年間の英語の授業で、英語についての知識は増えたり広がったりしましたか。
- | | |
|---------------|-----|
| 1. 増えた・広がった | 37 |
| 2. 少し増えた・広がった | 187 |
| 3. 変わらない | 84 |
| 4. 減少した・狭くなった | 10 |

アンケート結果を見る限りでは、学生にとってe-learning教材を用いた学習は有益であり、その内容とリンクした小テストの難易度もある程度適切であったと思われる。また、半数以上の学生が、

1年間の教養英語授業の履修を通して英語に関する知識が「増えた」・「少し増えた」と回答しているが、こうした学生の実感、プレイメントテストを活用した教育効果の測定の結果には現れているとは言えない。

2.4.3.2. 授業時間外のe-learning学習時間

授業でのe-learning教材の使用方法を見直す上での手がかりを得るために、各学生が1年間にe-learning教材を使って授業時間外に学習した時間（ログインしていた時間）の総計と、その学生の入学時点から後期末時点までの θ 値の変化について相関分析を実施した。その結果、ピアソンの積率相関係数を用いた分析では両者の間には有意な相関が観察されなかった ($r=0.080$, $p=0.073$ (n.s.))。散布図（図4）より、e-learning教材の学習時間を全くしていない者も一定数存在しており、この変数は正規分布に従っていない可能性が高いため、Spearmanの順位相関係数 (r_s) と Kendallの順位相関係数 (τ_b) についても検討したが、いずれにも有意な相関が観察されなかった ($r_s=0.086$, $p=0.053$ (n.s.); $\tau_b=0.060$, $p=0.054$ (n.s.))。これらの相関係数はいずれも絶対値がきわめて小さく、仮に有意であったとし

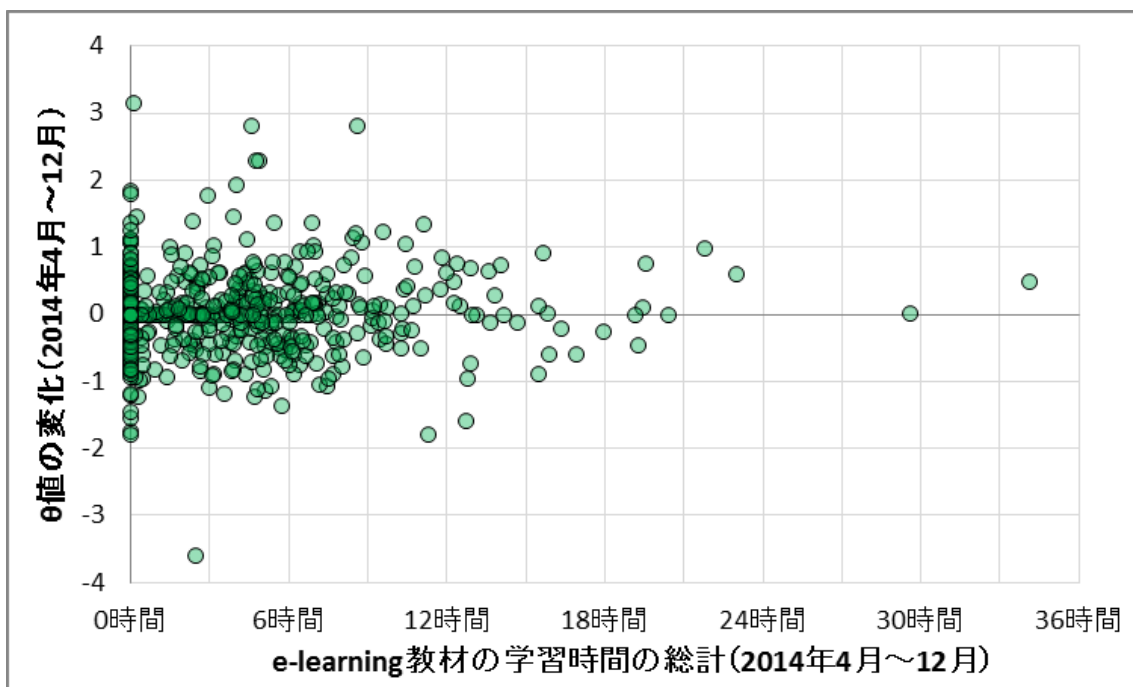


図4 2014年度（4月～12月）のe-learning教材学習時間の総計と、プレイメントテストから後期末共通テストの間の θ 値の伸びの散布図

ても、相関は極めて弱いと判断せざるを得ない。

さらに、個々の学生の学習状況を分析した結果、ログイン時間の大半をユニットの学習に充てていた学生と、ログインしていた時間は長いもののユニットの学習をほとんどしていない学生、ログイン自体していない学生が混在しており、学生の総学習時間に相当の誤差があることが判明した。この事実は、本来相関があったとしても、現状ではデータにおける誤差の影響で相関が観察されない事が容易に予測される状態であることを示しており、これがe-learning教材学習時間と θ 値の変化の間に相関が見られなかった理由の一つとなっている可能性がある。

2.4.4. その他：TOEIC IP スコアの伸び

別府大学教養英語では、選択科目としてTOEIC関連授業を開講しており、これらの授業を履修している学生には原則としてTOEIC IPテストを受験するよう促していることから、年に2回（6～7月・12月）のTOEIC IPテストをともに受験している学生が一定数存在する。これらの学生の中には、TOEIC関連の選択科目も必修英語もともに履修している1年生もいれば、選択科目のみを履修している2年生以上の学生もいるが、そうした違いはさておき、2012年度から2014年度まで、同一年度内に実施されたTOEIC IPテストを2回とも受験した学生を対象に、TOEIC IPスコアの伸びを測定したところ、2012年度には有意な伸びが観察されなかったが（ $t=0.418$, $df=8$,

$p=0.687$ (n.s.))、2013年度と2014年度にはそれぞれ平均で30点と17点の伸びが観察された（2013年度: $t=-3.601$, $df=54$, $p<0.001$; 2014年度: $t=-2.063$, $df=57$, $p<0.05$ ）。

この結果と、TOEIC IPテストと直近の学内テスト間の潜在能力値の伸びを比較してみると、2013年度、2014年度ともに学内テストでは伸びが観察されなかったが、TOEIC IPスコアについては伸びているようである。しかしながら、表6の結果には2年生以上の学生や外国人留学生のデータも含まれており、前節までの新入生を分析対象とした学内テスト結果との比較をするのにふさわしくない面がある。そこで、それぞれの年度について、同一年度内に実施されたTOEIC IPテストを共に受験し、かつその直近の学内テストをすべて受験した1年生（例：2014年7月・12月のTOEIC IPテストを受験し、2014年前期末共通テスト、2014年後期末共通テストを受験した1年生）を抽出し、TOEIC IPスコアと学内テストの平均の伸びをそれぞれ求めた（次頁表7）。その結果、TOEIC IPスコアの伸びについては2013年度のみ有意な伸びが観察された¹⁰⁾。また、学内テストにおける潜在能力値は、全体傾向と同様で、2013年度、2014年度ともに有意な伸びが観察されなかった。表4において、2014年度にはそれ以前と比べて教育効果が弱まってしまった可能性を指摘したが、残念なことにTOEIC IPスコアの面からも同様の可能性が見て取れる。

表6 同一年度内に実施されたTOEIC IPテストをともに受験した学生のテスト結果 (Total Score)

		平均値	SD	中央値	最大値	最小値
2012年度 (該当者数：9名)	6月	375	168	370	740	200
	12月	363	148	325	705	225
	伸び	-12	84	-35	110	-145
2013年度 (該当者数：55名)	7月	321	84	315	515	170
	12月	351	106	350	660	185
	伸び	30	63	30	230	-100
2014年度 (該当者数：58名)	7月	359	99	348	675	190
	12月	377	113	370	715	195
	伸び	17	64	20	175	-165

表7 表6のうち、1年生に関するTOEIC IPテストおよび学内テストの結果

			平均値	SD	中央値	最大値	最小値
TOEIC IP テスト	2012年度 (該当者数：0名)	6月	—	—	—	—	—
		12月	—	—	—	—	—
		伸び (t検定結果)	—	—	—	—	—
	2013年度 (該当者数：11名)	7月	339	62	350	430	195
		12月	403	65	380	495	275
		伸び (t検定結果)	64 ($p < .001$)	41	80	110	-5
	2014年度 (該当者数：19名)	7月	398	105	375	675	220
		12月	418	121	420	715	235
		伸び (t検定結果)	21 (n.s.)	68	40	125	-165
学内テスト	2012年度 (該当者数：0名)	前期末	—	—	—	—	—
		後期末	—	—	—	—	—
		伸び (t検定結果)	—	—	—	—	—
	2013年度 (該当者数：11名)	前期末	1.811	1.453	1.312	5.526	0.449
		後期末	1.445	0.802	1.335	3.474	0.492
		伸び (t検定結果)	-0.365 (n.s.)	1.076	0.023	0.977	-2.784
	2014年度 (該当者数：19名)	前期末	1.841	1.246	1.691	5.404	0.542
		後期末	1.810	1.383	1.592	6.322	0.168
		伸び (t検定結果)	-0.031 (n.s.)	0.933	0.067	1.776	-2.258

2.5. 総合的考察と今後の課題

必修英語授業の受講者は年度ごとに変化するが、教えている立場からすると、全体として見れば年度による学習意欲の違いはほとんど感じられない。また、2014年度からは教養英語必修授業においてe-learning学習を全面的に導入するなど、それ以前に比べて学習環境も整備された。しかしながら、以上の分析結果は2014年度にはそれ以前と比べて必修英語授業の教育効果が弱まってしまったことを示していた。直ちにその原因を明らかにすることは簡単ではないが、存在が明らかになった問題点を改善し、教育効果をより高めるための取り組みを続けていく必要がある。

一つの可能性として、現状では授業に導入したe-learning学習のシステムを十分に生かす授業モデルが確立できていないことが挙げられる。学生のe-learning学習状況調査から、e-learningの課題にまったく取り組んでいなかったり、ログインはするもののユニットの中身をしっかりと学習していない学生がいるなどの事実が明らかとなったことから、授業へのe-learning教材の組み込みにおいて、単にe-learning学習の課題を課すだけでなく、実際に学生がそれらに真剣に取り組むようにする仕掛けが必要である。また、潜在能力値に伸びが観察されなかった点については、学生が毎週課されるe-learningの課題をこなすのに追われ

てしまい（もしくは、それだけで英語の勉強をしっ
かりしたという気になってしまい）、それ以外の
部分（例えば、個々の授業で用いているテキスト
内容など）の勉強に割く時間が減ることにより、
全体としてはかえって英語を学習する時間が減っ
てしまったという可能性もありうる。もしそうだ
とすれば、e-learningの課題の量や頻度を見直す
などして、全体として学習量と学習の質が高まる
ように工夫すべきである。どちらにしても、e-
learning教材をより有効に活用することができる
よう、その運用方法についてさまざまな角度から
見直しを進めていくことが必要である。

3. まとめ

別府大学の日本語・英語教員グループによるプ
レイスメントテストを効果的に活用するための共
同研究の成果の一部として、教養英語の取り組み
を報告した。項目応答理論を用いた分析により、
英語のプレイスメントテストおよび共通テストは、
今後改善の余地はあるものの、テストとしての有
効性をかなりの程度持っていることがわかった。
また、潜在能力推定値を用いて学生の英語能力を
追跡調査した結果、教養英語授業の教育効果の
測定を実施し、現状の把握と今後解決すべき問
題点を明らかにすることができた。教育効果に関
する分析結果自体は必ずしも教養英語担当者
にとって歓迎されるものではなかったものの、
授業改善の第一歩は現状の問題点を客観的に可
視化することであるから、来年度以降の授業改
善のヒントが得られたものと解釈し、今後も多
面的な取り組みを続けていく必要がある。

付記

本研究は別府大学GP（研究支援；平成24年度・
25年度・平成26年度）「語学教育におけるプ
レイスメントテストの効果的な活用と教育効果の
測定に関する共同研究」の助成による研究成果の
一部である。

注

1) 別府大学教養英語の必修授業は、英語1（前
期、1単位）、英語2（後期、1単位）、

英語3（前期、1単位）、英語4（後期、
1単位）の4種類から成り、英語1・2は英作文
や文法、読解を中心とした授業、英語3・4は
聴解や会話などの運用力を高めることを目的
とした授業という位置づけとなっている。

- 2) これまで竹安ほか（2014a, 2014b）では
「成果測定テスト」の名称を用いていたが、
今後はテストを行う際の実際の名義である「共
通テスト」を用いる。
- 3) 2014年度については、プレイスメントテ
スト（4月）と後期末共通テスト（12月）は
短期大学部1年生（約160名）も受験してお
り、これらも分析の対象としている。
- 4) この点は竹安ほか（2014b）の手順と同様
であるが、竹安ほか（2014b）では項目パラ
メータや受験者の潜在能力値（ θ ）の推定等
を1年生のデータのみに基づいて実施したと
も解釈できる表現を用いて記述してしまった。
誤解を招く記述をしたことについて、この場
を借りて訂正してお詫びする。
- 5) 別府大学教養英語における項目パラメータ値
や受験者の潜在能力値（ θ ）の推定には
concurrent calibration法を用いているため、テ
スト実施に伴い個々の項目や受験者の推定
値が更新され、それに伴ってモデル適合度や
テスト総情報量も変化する。表1の2012年
～2013年の間のテストに関する表中の値が
竹安ほか（2014b）の表1で報告されている
値と若干異なっているのはこのためである。
- 6) de Ayala (2009)には、0.5～1.5の範囲内を基
準とする判定方法以外にも、OutFitについて
は $1 \pm 6/\sqrt{N}$ の範囲内、InFitについては $1 \pm$
 $2/\sqrt{N}$ の範囲内を基準とする方法も紹介さ
れている。こちらの判定方法によれば、本稿
のテスト結果については特にInFitの値がモデ
ルに過剰に適合していると判断される。理由
としては、英語プレイスメントテストおよび
成果測定テストが主に文法項目を中心とした
構成となっており、似たパターンの問題が多
いために項目間の依存性が高くなりやすい

こと、また、共通項目を選択する際にできるだけ識別力の高い項目を選択しているが、これらの共通項目に適合度が小さなものが多かったことなどが挙げられる。

- 7) 昨年に竹安ほか (2014b) で2013年度分のデータ (156ペア) に対して同様の分析を行った際には、 $r=0.628$ であった。今回の分析でも、ほぼ同じ程度の相関の強さが観察されたと言える。
- 8) 項目応答理論では受験者の能力 (プレイスメントテストなどで測定した英語力) は潜在能力値 (θ) で表される (θ は受験者の解答パターンを分析して求められる値)。 $\theta = 0$ が平均的な能力を示しており、正の値であれば平均よりも能力が高いことを、負の値であれば平均よりも能力が低いことを示す。また、絶対値の大きさは平均と比較したときの能力の高さ (低さ) の度合いを表している。
- 9) $\alpha=0.05$ における観測検定力の値が0.335と低いため、本来は平均に差があるのに検知できなかった可能性もあるが、少なくとも昨年までは同規模のデータに有意な平均の差が観察されていたことを考えると、今年度は教育効果が弱まってしまったと言えるだろう。
- 10) 2年生以上の分析結果や学年を分けずに分析した結果を総合すると、2014年度の1年

生のTOEIC IPテストにおける得点は全体として伸びてはいるが、各学生の伸び具合に相当な個人差があり、データ数もそれほど多くないために標準誤差が大きくなってしまい、有意水準に達しなかったものと考えられる。

参考文献

- de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York, NY: The Guilford Press.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- 熊谷龍一 (2009) 「初学者向けの項目反応理論分析プログラムEasyEstimationシリーズの開発」『日本テスト学会誌』5, pp.107-118.
- 竹安大・三重野佳子・東真千子・内山和也・松田美香 (2014a) 「学生のレベルに合わせたプレイスメントテスト構築とその運用」, 『第62回九州地区大学一般教育研究協議会議事録』, pp.65-70, 九州地区大学一般教育研究会.
- 竹安大・三重野佳子・東真千子・内山和也・松田美香 (2014b) 「別府大学語学教育におけるプレイスメントテストの活用：教養英語の事例報告」『別府大学日本語教育研究』4, pp.21-27, 別府大学日本語教育研究センター.

(2015年3月25日受付)