

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—

竹安大・三重野佳子・東真千子・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、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を報告するものである。

教養英語の必修の授業では、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。

こうした取り組みの概要と、中間結果の一部は竹安ほか(2014)にて報告したところであるが、本稿では、その後の経過報告とともに、2012年度及び2013年度に別府大学教養英語において実施した学内統一テスト(入学時のプレイスメントテスト、前期末および後期末の成果測定テスト)の結果を分析して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利用した教育効果の測定

別府大学の日本語及び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並行したテストを一定の期間をおいて実施し、そ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。教養英語では、2012年度は4月(プレイスメントテスト)7月(前期末成果テスト)に、また、2013年度は4月(プレイスメントテスト)と7

月(前期末成果測定テスト)、12月(後期末成果測定テスト)に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他、2013年度には後期末の授業で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付録参照)。

2.1. テストの構成

竹安ほか(2014)にて報告した通り、教養英語のプレイスメントテストおよび成果測定テストはいずれも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2割程度はテスト間比較のための共通項目となっている。また、所要時間は30分程度で、受験者数は500名弱である¹⁾。

2.2. 分析方法

プレイスメントテストと成果測定テストは問題形式や問題数は同じであるが、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)により等化を行ない、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/成果測定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により行った。

2.3. 結果

2.3.1. 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(2014)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。

2.3.1.1. 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なった。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

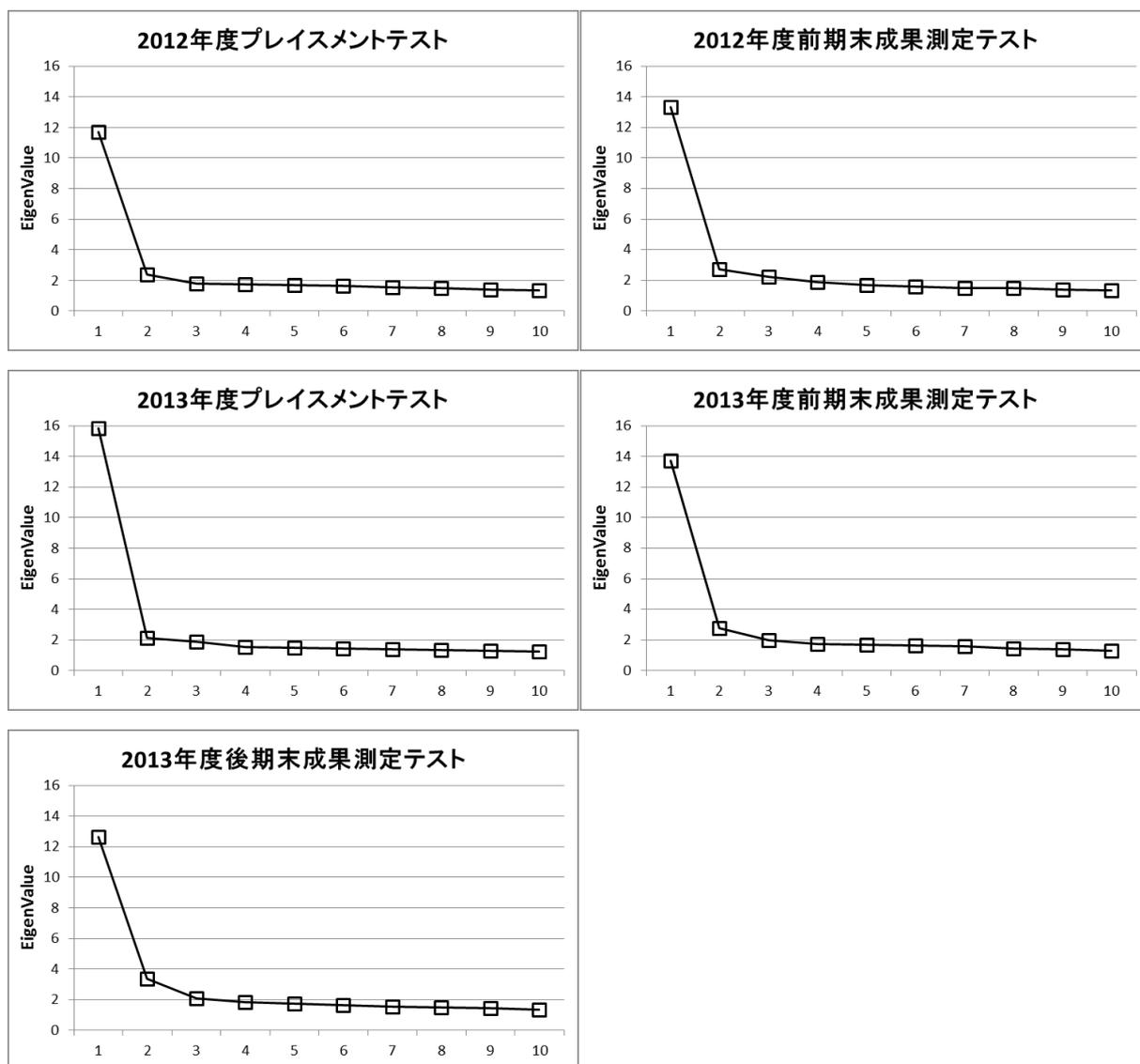


図 1 過去のプレイスメントテスト・成果測定テストのスクリープロット

表1 2013年度後期末成果測定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	0.787	0.697	31.06	-0.108
2012年7月	421	50	0.702	0.607	33.13	0.001
2013年4月	457	50	0.694	0.573	37.55	-0.074
2013年7月	427	50	0.754	0.655	31.59	0.132
2013年12月	437	50	0.709	0.632	34.44	0.090

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無ければ、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2013年度12月までの5回のテストのスクリープロットを図1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無いと判断できる。

2.3.1.2. 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5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。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、モデルの適合度に問題はないと判断できる²⁾。また、テスト総情報量については、2012年4月に実施したテストから見た場合、回数を重ねるにつれて徐々に値が大きくなりつつあり、テスト全体の有効性が改善されつつある傾向が示されている一方で、個別にみるとばらつきも見られる。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く、こうしたばらつきが生じてしまっているものと思われるが、今後はこうしたばらつきを減らし、より情報量の高いテストを構築していくことが望ましい。

2.3.1.3. 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メントテスト・成果測定テストの妥当性の検証の一環として、

これらのテスト結果と外部テスト（TOEIC）との相関を調べた。

別府大学では、2012年度より年に2回（7月頃と12月頃）の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成果測定テストが実施される時期と近い（以下の表2参照）。従って、これらのTOEIC IPテストを受け、かつ直近の成果測定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）と、成果測定テストの θ の値についてピアソンの積率相関係数を求めたところ、 $r=0.628$ ($p<0.001$)であった（散布図を図2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、別府大学教養英語成果測定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であ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。

2.3.2. 潜在能力値（ θ ）

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成果測定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。なお、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。

2.3.2.1. 2012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 -0.108 であったのに対し、前期末時点では 0.001 であった。テストを2回とも受けた学生の θ の値について対応のあるt検定で分析したところ、 θ の平均値の間に有意な差が観察された（ $t=-3.923$, $df=352$, $p<0.001$ ）。以上のことから、プレイスメントテストから前期末テストの間に、学生の能力が伸びたと判断できる。

2.3.2.2. 2013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 -0.074 であったのに対し、前期末時点では 0.132 、後期末時点では 0.090 であった。3回のテストをすべて受けた学生の θ の値について反復測定ANOVA（SPSS混合分析（線形）モデルによる）で分析したところ、3つの θ の平均値の間に有意な差が観察された（ $F_{(2, 768)}=22.470$, $p<0.001$ ）。多重比較（Bonferroniの修正による）を実施したところ、プレイスメント時点と前期末時点、プレイスメント時点と後期末時点の θ の平均値の間に有意な差が観察された。一方、前期末時点と後期末時点の θ の平均値の間には有意な差は観察されなかった。以上のことを合わせると、プレイスメントテストから前期末テストの間に学生の能力が伸

表2 相関分析における別府大学TOEIC IPテストと成果測定テストの対応関係

TOEIC IPテスト		直近の成果測定テスト	TOEIC IPテストと成果測定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年度前期末成果測定テスト	2
2012年12月2日	69	なし（実施せず）	—
2013年7月13日	126	2013年度前期末成果測定テスト	92
2013年12月21日	87	2013年度後期末成果測定テスト	62

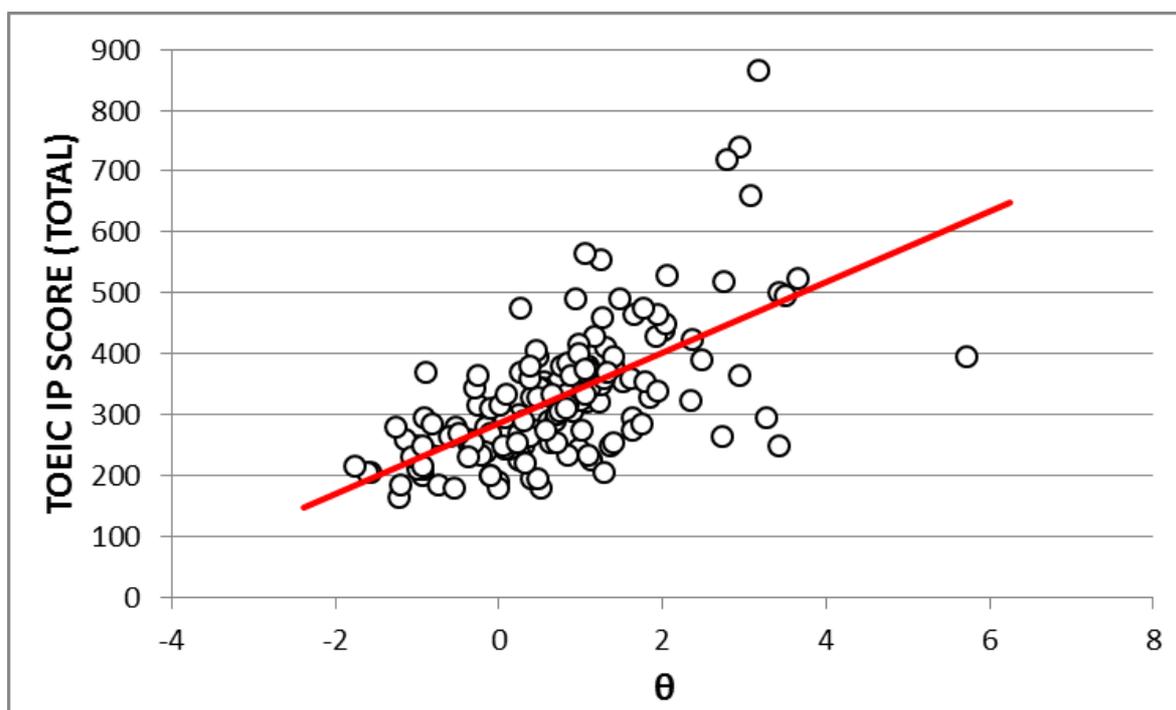


図 2 TOEIC IP テストの得点と成果測定テスト結果 (θ) の散布図

びたが、前期末から後期末の間には学生の能力には大きな変化が見られなかったと判断できる。

2.4. 考察と今後の課題

いずれの年度も、入学時から前期末にかけては平均潜在能力値が上昇していた。日本で普通に生活している大学生が授業外で英語に触れる機会は限られているため、この上昇は主に英語の授業を履修した結果によるものであると判断すると、前期の英語授業は学生の英語能力向上について一定の効果を発揮したと言える。一方、後期末にも成果測定テストを実施した2013年度の結果を見ると、前期末から後期末の間に平均潜在能力値の有意な変化が見られず、後期末の授業があまり有効に機能していないことがわかる。別府大学教養英語授業では、個々のクラスは同じ教員が前期・後期を続けて担当する体制になっているため、前期と後期の教育効果の違いを生じさせた原因が教員側にあるとは考えにくい。この点について、テスト間比較結果を教養英語教員に対してフィードバックした際、ある教員から「前期に比べて後期は学生の中だるみが目立ち、教えていて成果を感じにくかった。テスト間比較結果は授業における教員の体感とも一致している。」とのコメントが得ら

れており、こうした中だるみが後期授業の教育効果の弱化につながった可能性がある。これが事実であれば、今後は学生の学習意欲を継続させるような仕掛けを教養英語グループ全体で考えていく必要があるだろう。

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ー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、英語のプレースメントテストおよび成果測定テストは、今後改善の余地はあるものの、有効に機能していると判断できることがわかった。今後は、項目分析を行って、個々の問題の質を高めながら、テスト全体としての有効性をさらに高めていく必要がある。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することができた。現状では、後期の授業の教育効果が弱く、その原因の解明および教育の質の改善が必要であることが明らかとなった。

付記

本研究は別府大学GP（研究支援；平成24年度・25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) 英語の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、成果測定テストについては、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。本稿は1年生の必修英語授業の教育効果の測定が目的であるため、2年生以上の受験生については分析の対象としない（TOEICとの相関を議論するセクションを除く）。
- 2) 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。

参考文献

- 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（2014）「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.
- 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.
- (2014年3月10日受付)

付録

教養英語プレイスメントテスト・成果測定テスト結果の学生へのフィードバックの例

2013年度

一年間の英語学習成果

学科	史学・文化財
学籍番号	A13
氏名	

受験者コード 13

テストコード ##

入学時 ##

前期末 ##

後期末 ##

テスト結果

入学時	547	▼
547 点		1000
前期末	601	▼
601 点		1000
後期末	595	▼
595 点		1000

出題文法項目	正答数／出題数		
	入学時	前期末	後期末
1 動詞と文型 (be動詞・一般動詞現在形、5文型、自動詞と他動詞)	4 / 5	2 / 5	4 / 5
2 時制 (過去形・未来形、時制の一致)	2 / 5	3 / 5	4 / 5
3 完了形・進行形・態 (現在完了形、現在進行形、過去進行形、受動態)	3 / 4	4 / 4	3 / 4
4 準動詞 (不定詞・動名詞・分詞)	3 / 3	1 / 3	3 / 3
5 助動詞 (助動詞の意味・用法)	2 / 2	2 / 2	1 / 2
6 形容詞・副詞 (形容詞・副詞の意味・用法)	2 / 2	0 / 2	1 / 2
7 比較 (原級、比較級、最上級、比較表現)	3 / 3	3 / 3	3 / 3
8 名詞・代名詞 (単数と複数、冠詞、人称/指示/不定代名詞、itの特別用法)	3 / 4	3 / 4	3 / 4
9 前置詞・接続詞 (前置詞、接続詞の意味・用法)	2 / 3	2 / 3	2 / 3
10 特殊な文の形 (命令文、感嘆文、There is(are)の文)	1 / 3	3 / 3	3 / 3
11 疑問文 (疑問詞疑問文・付加疑問文・否定疑問文・間接疑問文)	3 / 3	3 / 3	3 / 3
12 関係詞 (関係代名詞、関係副詞の用法)	2 / 2	2 / 2	1 / 2
13 その他構文・語彙 (仮定法、語彙、構文)	9 / 11	7 / 11	5 / 11

コメント

入学時点から48点の伸びがありました。1年間の学習を通して、ある程度理解が深まっているようです。少しずつ結果が出てきていますので、この調子で頑張ってください。基本の文法はある程度理解できていますが、複雑な文の仕組みの理解がまだ不足しています。点が取れていない項目を勉強し直しましょう。現時点であなたがTOEICを受験した場合、330点前後の点を取得できると予想されます。努力次第ではもっと高い点を取ることも期待できますので、さらに上を目指して今後も勉強を続けて下さい。