

数学プレースメントテスト成績と期末試験成績の比較

廣瀬 英雄*

(平成27年9月24日受付)

Comparison of Testing Evaluations between the Placement Test and the Final Examination

Hideo HIROSE

(Received Sep. 24, 2015)

Abstract

It is considered to be difficult to design mathematics lectures for undergraduate classes if a variety of students exist in the same class. To grasp the structure of the classes as early as possible, placement tests are often performed. Using the testing results, students are classified into different level classes. However, the students in the lowest classes may lose pride and cooperation attitude. Thus, our university decided not to provide such classes. In such a situation, much more efforts are required for teachers to enhance students' skills. To check if the progress of mathematics in undergraduate classes, we have compared the final test results with the placement test results. The item response theory (IRT) is used to evaluate the abilities of students rather than the classical test theory.

We have found that 1) students made progress by well-arranged lecture design, 2) there seems to be a positive correlation between the placement test results and the final test results in analysis classes, 3) much more progresses were seen for students without higher level high school education than those with that, 4) in linear algebra classes, no clear relationships were observed between the placement test results and the final test results. Above all, students made progress even though they were not educated well in high schools.

Key Words: placement test, final test, item response theory, progress, analysis, linear algebra

1 はじめに

今では多くの大学が多様な学生を受け入れている。理数系大学を目指して高校数学をきちんと学んできた学生もいれば、高校で数I、数Aは教わってはいても、数IIや数B、あるいは数IIIや数Cについては十分に教わってこなかった学生もいる。後者にはAO入試や推薦入試によって数学の筆記試験を免れて入学してくる学生もいる。こういった状況をいち早く把握して入学してくる学生にあわせた授業の進め方を設計しなくてはならない。そのために、入学直後にプレースメントテストを実施することによって高校で学

んだ基本的な知識を確認している。

入学直後の学生の知識に合わせた内容（これには高校の復習も含められる）から、大学生として身に付けるべき素養（学士力）まで高める授業設計を行うことが求められるが、問題となるのは入学時の学生の知識レベルの大きなばらつきである。プレースメントの結果によって、ある程度同じようにまとまった知識レベルの学生のクラスをいくつか設けて授業設計を行うことがかまわず考えられるが、こうすることでうまくいくとは限らないこともあると聞く。つまり、もっとも基本的なところから始めるクラスの学生には、誇りが失われる、となりどおしの助け合いが功を奏し

* 広島工業大学環境学部環境デザイン学科

ないといった問題点が指摘されているのである。このため、広島工大では上位クラスは設けるが、中位と下位クラスは合同としている。

今回、この合同クラスについて、プレースメント成績と期末試験の成績とを比較してみた。このように知識レベルが大きくなればつきがある場合でも授業設計がうまくいっているかどうかを見るためである。今回はすべてのクラスについて調べたものではなく、一部のクラスについて調べたサンプリング結果の報告である。30人クラスと60人クラスについての1年生の解析学のコース（必修）、60人クラスについての1年生の線形代数学のコース（選択）、および2年生の解析学のコース（15人、20人程度の選択科目）を対象とした。また、プレースメントテストについては、数Ⅰ、数Ⅱ、数Ⅲ、数Ⅳ、数Ⅴ、数Ⅵ、数Ⅶ、数Ⅷ、数Ⅸ、数Ⅹ、数Ⅺ、数Ⅻすべてを調べることとした。2014年度以前の古い学習指導要領に従う学生もいることを考慮して、新しい指導要領内容には従わず以前のものに従うテスト内容とした。そのため、行列の問題も1部含まれている。まとめると、

- 1) 数ⅠA、ⅡB、ⅢCすべてを含むプレースメントテストの成績と期末試験の成績とを比較する。
- 2) 成績評価に普遍性を持たせるため、素点ではなく現代テスト理論で用いられる項目反応理論（IRT）から得られる評価結果を用いる。
- 3) ED学科1年生の約半数、ICT系の1年生の約1/4、GES学科2年生を調査対象とした。
- 4) 調査した数学科目は解析学および線形代数である。

2 成績評価の方法

これまでプレースメントテストの成績評価については、あらかじめ決められた素点について正解した点数を合計した総合点で行っていることが多い。いわゆる古典的テスト評価法である。この方法では問題の難易度については考慮されていない。これに対して学生の習熟度評価に加えて問題の難易度も考慮された評価法が現代テスト理論である。項目反応理論（Item Response Theory, IRT）¹⁻³⁾ とよばれる。ここではIRTを用いた評価法を用いる。これによって、問題毎の評価が可能になり、それを使うことによってアダプティブなテスト⁴⁾が可能になる。IRTの理論についてはその概略を付録に示す。

問題数は、数Ⅰが7問、数Ⅱが3問、数Ⅲが6問や数Ⅳが2問、および数Ⅴが9問や数Ⅵが7問の合計34問である。回答法はマークシート方式で、与えられた解答例から正解を選択する方法（9択）であるので、デタラメに正解を選択する可能性は低い。

3 プレースメントテストと期末試験の成績の比較

学生を、1) 一般入試受験者および一般入試非受験者（推薦、AO、その他）の両方の場合、2) 一般入試非受験者（推薦、AO、その他）のみの場合、3) 一般入試受験者のみの場合について区別して分析してみた。

3.1 ED学科2015年度1年生（解析）の結果との比較

図1に、プレースメントテストの結果とED学科2015年度1年生（解析）の結果との比較を示す。横軸がプレースメントテストの結果、縦軸が期末試験結果である。図から次のようなことが分かる。

- 1) プレースメントの成績と期末の成績とは弱い正の相関が見られる。相関係数=0.43
- 2) 一般入試受験者（平均=(-0.25, 0.29)）と一般入試非受験者（平均=(-1.12, -0.44)）の間には、プレースメント、期末試験の両方に差が見られ、一般入試受験者の方が良い成績を示す（平均の差=(0.87, 0.73)）。
- 3) 一般入試受験者と一般入試非受験者には、入学後両方ともに努力の後が見られる（平均(-0.25=>0.29), 平均(-1.12, =>-0.44)）。
- 4) 一般入試受験者よりも一般入試非受験者の方により成績向上（大きい上がり具合）が見える（プレースメントの差 $0.87 >$ 期末の差 0.73 ）。

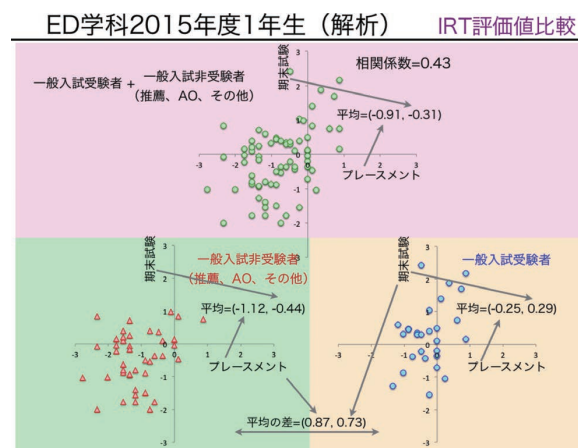


図1 ED学科2015年度1年生（解析）との比較1

3.2 数Ⅱ、Bと数Ⅲ、Cの未履修効果について

図2に、プレースメントテストの結果とED学科2015年度1年生（解析）の結果との比較を、1) プレースメント数Ⅰ、Aのみと仮定、2) プレースメント数Ⅱ、Bのみと仮定、3) プレースメント数Ⅲ、Cのみと仮定した場合について示す。

- 1) プレースメントテスト結果には数ⅠAだけ履修している不利益が見られるが、期末試験の結果にはこの不利

ED学科2015年度1年生（解析） IRT評価値比較

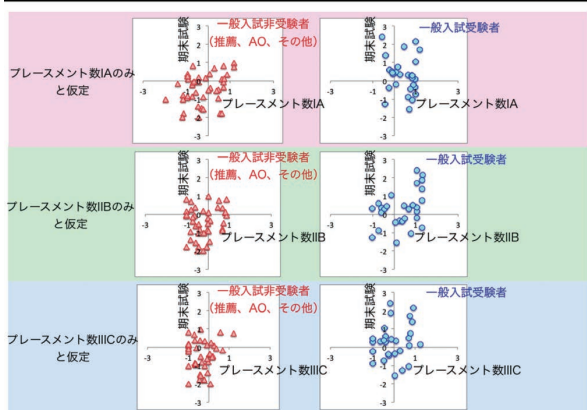


図2 ED学科2015年度1年生（解析）との比較2

益はあまり反映されているようには見えない。つまり、入学後の数II Bと数III Cの復習の効果がみられる。

- 2) プレースメントテストを数I A, 数II B, 数III Cと細分化して期末試験結果と比較したとき、いずれの間にも目立った相関は見られない。

3.3 ICT系2015年度1年生（線形代数）の結果との比較

図3に、プレースメントテストの結果とICT系2015年度1年生（線形代数）結果との比較を示す。図から次のようなことが分かる。

- 1) プレースメントの成績と期末の成績との間に正の相関があるようには見えない（相関係数=0.27）。
- 2) 一般入試受験者と一般入試非受験者には、入学後両方ともに努力の後が見られる。なぜなら、平均（ $-0.69 = > 0.42$ ）、および平均（ $-0.62 = > 0.19$ ）の変化があるからである。
- 3) 入学前の評価値と入学後の評価値に大きな変化がないのは、線形代数の内容が解析学に比べて高校数学と大きく異なるからであると考えられる。

ICT系2015年度1年生（線形代数） IRT評価値比較

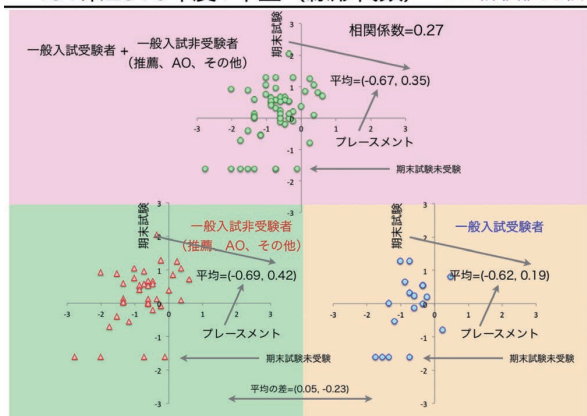


図3 ICT系2015年度1年生（線形代数）との比較

3.4 GES学科2015年度2年生（解析）の結果との比較

図4に、プレースメントテストの結果とGES学科2015年度2年生（解析）の結果との比較を示す。図から次のようなことが分かる。

- 1) プレースメントの成績は1年生よりも上がっているように見える。
- 2) プレースメントの成績と期末の成績との間には関係があるようには見えない。
- 3) 期末試験の成績評価が悪いように見えるが、これは試験問題のレベルをかなり高くしたためである。成績評価の多くの割合を反転授業での発表評価に充てているため、総合成績は総じて高い。

GES学科2015年度2年生（解析） IRT評価値比較

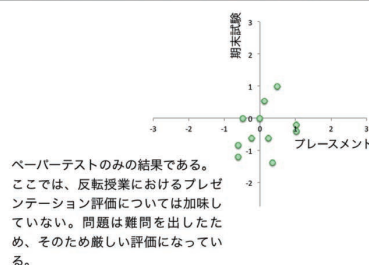


図4 GES学科2015年度2年生（解析）との比較

4 まとめ

多様な学生を受け入れる大学で基礎数学の教育をうまく機能させることは難しい。通常プレースメントテストによってクラスを分けることが行われているが、広島工大では、学生の意気を失わせないように、また学生同士の助け合いの精神を保てるように、低いレベルのクラスは設けていない。このため、教育効果がどのようになっているかを調べることにした。調査方法は全数調査ではなく、一部のクラスのサンプリングによって行った。クラスは解析学と線形代数の1年生と2年生である。

プレースメントテストの結果と期末試験の結果の比較を行った結果、1) 学生は伸びている、2) 解析学のクラスではプレースメントテストの結果と期末試験の結果の間に正の相関が見られる、3) 高校で十分な数学教育を受けてこなかった学生は、そうでない学生に比べてより努力をしている、4) 線形代数のクラスではプレースメントテストの結果と期末試験の結果の間に明確な相関は観察されなかった。

文 献

- 1) R. K. Hambleton and H. Swaminathan, Item Response Theory: Principles and Applications. Springer, 1984.

- 2) R. Hambleton and H. Swaminathan, and H. J. Rogers, Fundamentals of Item Response Theory. Sage Publications, 1991.
- 3) W. J. D. Linden and R. K. Hambleton, Handbook of Modern Item Response Theory. Springer, 1996.
- 4) H. Hirose, T. Sakumura, Item Response Prediction for Incomplete Response Matrix Using the EM-type Item Response Theory with Application to Adaptive Online Ability Evaluation System, IEEE International Conference on Teaching, Assessment, and Learning for Engineering 2012, pp. 8-12, 2012.

付録 項目反応理論 (Item Response Theory, IRT)

これまでの評価法では、各問題にはあらかじめ配点が与えられ、それぞれの問題の得点を合計した総得点が評価値であった。同じ試験を多くの人に課せば全員の総得点が得

られる。そこから平均や標準偏差を算出すれば、自分の相対的な評価値を偏差値という形で求めることができる。しかし、問題の配点を変えれば総得点が違ってくる場合がある。配点によって評価値が変わるのは公正な評価法とはいえないかもしれない。そこで、問題の難易度と各受験者の学習習熟度とを同時に求めながら、公正で公平な評価法が提案された。これがIRTによる評価法である。この理論は、これまでにTOEFLや情報処理検定など多くの公的な場面で適用されている。ここではこの評価法を用いている。

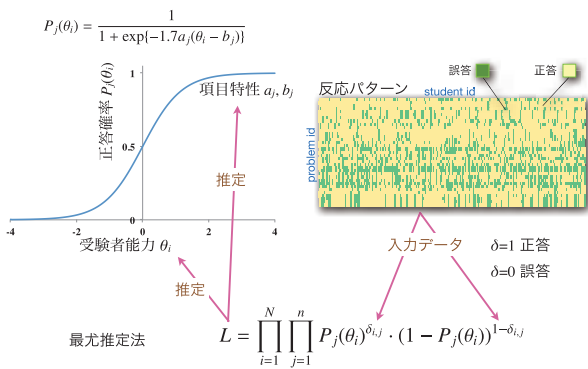
IRTでは、各問題*j*に対する受験者*i*の評価確率 $P_j(\theta_i)$ が2パラメータロジスティック分布、

$$P_j(\theta_i) = \frac{1}{1 + \exp(-1.7a_j(\theta_i - b_j))}$$

に従っていると仮定する。 a_j, b_j は問題*j*の識別力（簡単にいうと、問題の良し悪しを表す）と困難度（文字どおり、問題の難易度を表す）を、 θ_i は受験者*i*の習熟度を表している。受験者 $i=1, \dots, N$ が項目 $j=1, \dots, n$ に対して取り組んだ結果、その解答が正答なら、 $\delta_{i,j}=1$ 、誤答なら $\delta_{i,j}=0$ と書き表すと、すべての受験者がすべての問題に挑戦した結果（これを反応パターンという）の確率は、独立事象を仮定すれば、

$$L = \prod_{i=1}^N \prod_{j=1}^n P_j(\theta_i)^{\delta_{i,j}} (1 - P_j(\theta_i))^{1-\delta_{i,j}}$$

と表される。これを尤度関数という。付録図に、IRTによる評価の過程のイメージを示す。誤答0と正答1からなる $\delta_{i,j}$ を上の尤度関数 L に代入し、それを最大にするような a_j, b_j, θ_i を同時に求めるのがIRTによる評価法である。



付録図 項目反応理論の概念図