

大学数学科目の CBT と記述式による試験の評価結果の比較： 線形代数の場合

廣瀬 英雄*

(令和元年 8 月 6 日受付)

Comparison of Evaluation Results Between Multiple-Choice Testing and Description Type Testing: Linear Algebra Case

Hideo HIROSE

(Received Aug. 6, 2019)

Abstract

Using the LCT (learning check testing) results in the follow-up program system, we have shown that we can predict the students' future results of success/failure to the final examination with a probability to some extent. However, we also anticipate that we may not be able to enhance a certain level of prediction accuracy for evaluation because the description type questions and their evaluations to the final examinations depend on two or more teachers resulting unfair scores although the LCT provides the multiple-choice questions and the evaluations are performed automatically using computers. We had a chance to perform such a CBT using the multiple-choice questions and the automatic evaluations. Thus, in this paper, we investigate whether such a prediction is correct or not. Amazingly, there may be no differences of the results between the description type examination and multiple-choice testing in terms of the correlation coefficients. The subject is Linear Algebra for first semester undergraduate students.

Key Words: computer based testing, multiple-choice paper questions, description type questions, item response theory, Linear Algebra.

1 はじめに

2016年度から開始したカリキュラムでは、解析基礎、線形代数などの大学数学基礎科目について、毎授業でのオンラインテスト(LCT, learning check testing)を行い、習熟度が不十分とみなされる場合には補習授業を受講できるような、フォローアッププログラム (FP) 体制のもとで、よりきめ細かい教育を行ってきた。補習授業は FP クラス (FPC) と呼ばれる。

これまで、LCT の結果と期末試験結果を対比させると、ある時期から一定程度の確率で学生の期末試験の合否を予測できることが示されている。ただし、期末試験に失敗する学生に注目すると、失敗を予測した学生のおよそ半分は予測が的中しているが、的中率はなかなかそれ以上にはいれないと思われる。その理由として、LCT は全学生が同一条件で同一問題を解き、採点もコンピュータが自動的にを行うため、公平で公正な評価が行えているのに対し、期末試験では、同じクラスでも教育効果を考えて少人数に分割

* 広島工業大学データサイエンス研究センター & 環境学部建築デザイン学科

しているため、複数の教員が（記述式の）期末試験の採点を行うことになり、その結果、教員間での評価による不公平さが出てくるのではないかと考えていた。ここでは、その理由がそのように想定しているような原因に起因するものかどうかを検証してみたい。

そのため、これまでのような、LCTと記述式期末試験の評価結果を比較したものと、2018年度の期末試験ではじめて採用した多肢選択式の問題による評価結果をLCTの結果と比較したものを比べてみることにした。パイロット的な試みなので、2018年度は情報系学部の1年生後期の選択科目である線形代数Bについてのみ行っている。

評価はクラス分割して実施するような直接的な方法では行えないのでここでは間接的な方法に依った。

また、期末試験結果のLCTの結果との比較に加えてブレースメントテスト（PT）との比較も加えておきたい。

2 LCTと中間試験と期末試験の実施状況

2016年度からのカリキュラムでは、1年生前期では解析基礎Aと線形代数Aが必修科目であり、全学生が受講する。しかし、後期になると工学部では解析基礎Bと線形代数Bが必修科目になるが、それ以外の学部（情報学部、環境学部、生命学部）では選択に変わる。LCTは必修科目に課せられるため、前期は全学生がLCTを受験しているが、後期になると工学部しか受験していない。今回対象とするのは情報学部の選択科目線形代数Bである。

表1に、LCTと中間試験および期末試験の実施状況を示す。情報学部以外は中間試験と期末試験は記述式である。情報学部では、しかし、後期のLCTは実施されておらず、LCTとCBTとの比較は間接的に行うことになる。それには、1) 前期のLCT結果と後期のCBT結果を比較する、2) 工学部（あるいは全学部）での前後期のLCT結果間、

表1a LCTと中間試験および期末試験の実施状況（前期）

	前期LCT	前期中間	前期期末
工学部	CBT.A	記述式	記述式
情報学部	CBT.A	記述式	記述式
環境学部	CBT.A	記述式	記述式
生命学部	CBT.A	記述式	記述式

CBT.A: 線形代数Aに対応するCBT

表1b LCTと中間試験および期末試験の実施状況（後期）

	後期LCT	後期中間	後期期末
工学部	CBT.B	記述式	記述式
情報学部	-	CBT.B*	MSS
環境学部	-	記述式	記述式
生命学部	-	記述式	記述式

あるいは年度をまたがって比較する、という方法を用いることとした。

なお、一つのクラスを二つに分けて、一つはCBT、一つは記述式として対比させるのが好ましいが、大学正規の期末試験のため、公平さからまったく同じ条件下での試験環境とした。また、中間試験はLCTとまったく同様なオンラインテストに依ったが、期末試験はWiFiの遮断の危険性に配慮してマークシート方式（ここではMSSと呼ぶことにする）の多肢選択式とした。

表1で、CBT.Bは、線形代数Bに対応するCBT、CBT.B*は、CBT.Bと同等なCBT、MSSは、マークシート方式を表す。

3 2018年度情報学部での比較

3.1 選択科目によるバイアスについて

情報学部での線形代数Bは選択科目である。前期のLCT.Aは線形代数Aに対応しており、線形代数Aは必修科目であるため学部全員が受験している。そこで、比較のためLCT.Aを使うのは、後期のCBT.BあるいはMSSの結果は学生全員の結果を反映していないため、正しい比較ができないのではないかと疑問がある。つまり、科目選択によって結果にバイアスが発生する恐れはないかということである。そこで、線形代数B受講者のLCT.Aと非受講者のLCT.Aの正解率の比較を行ってみた。それを図1に示す。ヒストグラムからは両者の間に顕著な差はみられない。クラスの中からランダムサンプリングしたとみて差し支えないと思われる。

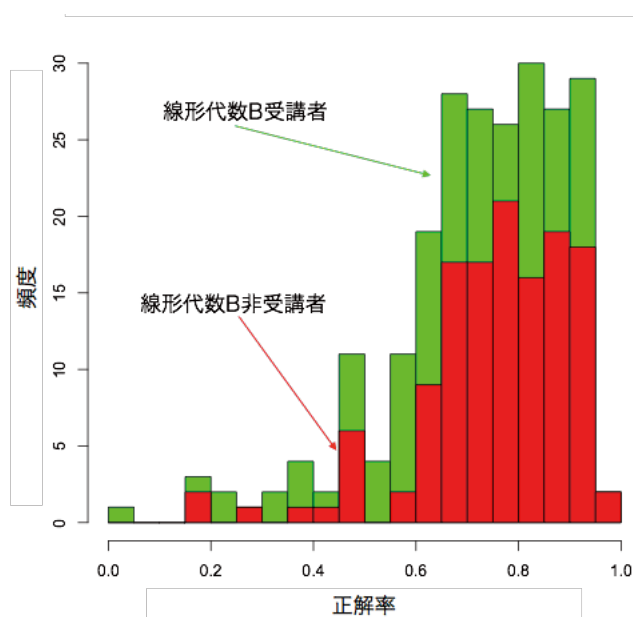


図1 線形代数B受講者のLCT.Aと非受講者のLCT.Aの正解率のヒストグラム（情報学部）

3.2 LCT.A と CBT.B* の比較

情報学部での LCT.A の正解率はかなり高い。問題がやさしいとも言える。それに対して、後期になると、問題が難しくなるため正解率は低くなると考えられる。実際に LCT.B は（選択科目であるため）実施されていないので直接的な比較はできないが、中間テストの CBT では LCT.B と同等なオンライン試験 LCT.B* を用いているため、間接的な比較は可能である。このヒストグラムを図 2 に示す。確かに、LCT.B* の正解率は LCT.A の正解率よりも低くなっている。

一方、同じ結果を両者について IRT（項目反応理論）を使って ability の計算し、比較したヒストグラムを図 3 に示す。正解率では見えていた両者の差が ability を使って比較すると差が見えなくなっていることがわかる。これは、IRT で ability を求める際に基準化が行われているためである。

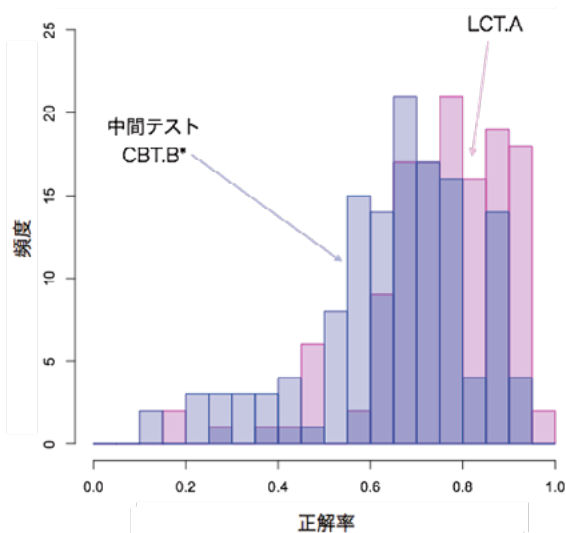


図 2 線形代数 B 受講者の LCT.A と CBT.B* の正解率のヒストグラム（情報学部）

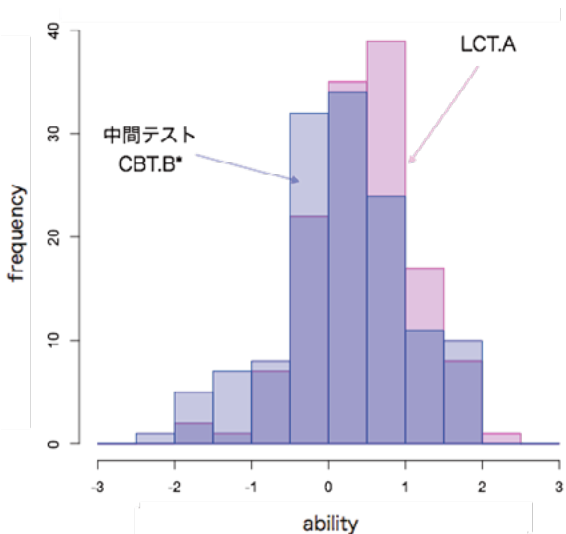


図 3 線形代数 B 受講者の LCT.A と CBT.B* の ability のヒストグラム（情報学部）

図 4、5 は、線形代数 B 受講者の LCT.A と非受講者の CBT.B* を、散布図を使って比較したものである。LCT.A と CBT.B* を比較すると、ある程度相関があるように見える。実際に相関係数を計算してみた値はそれぞれ 0.63、0.62 である。

同じような形式で同じような内容の問題による CBT の結果の相関は高いことが予想されるが、実際はこの程度の値であり、個人の習熟度が確率的にかなりゆらいていることが観測される。

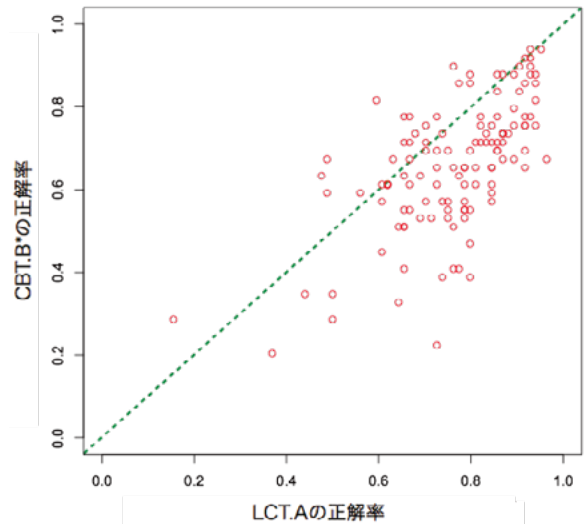


図 4 線形代数 B 受講者の LCT.A と CBT.B* の正解率の散布図（情報学部 2018 年入学生のみ）

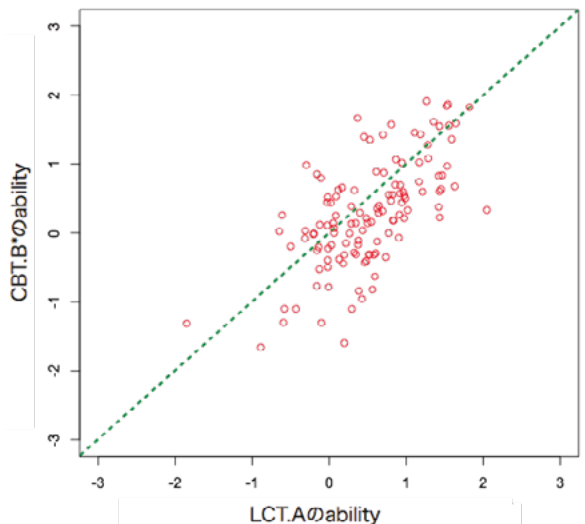


図 5 線形代数 B 受講者の LCT.A と CBT.B* の ability の散布図（情報学部 2018 年入学生のみ）

3.3 CBT.B* と MSS の比較

中間テストと期末試験の両方とも多肢選択式である。ここではそれらの間の関係を見てみよう。図 6、7 は、線形代数 B 受講者の CBT.B* と MSS の正解率のヒストグラム、および ability のヒストグラムの比較図である。正解

率はCBT.B*よりもMSSの方が低くなっていることがわかるが、abilityについては、LCT.AとCBT.B*のときと同様、基準化によるものである。

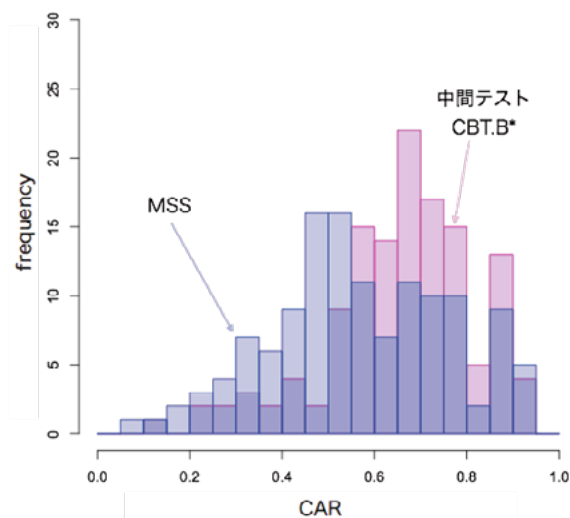


図6 線形代数B受講者のCBT.B*とLCT.Bの正解率のヒストグラム (情報学部)

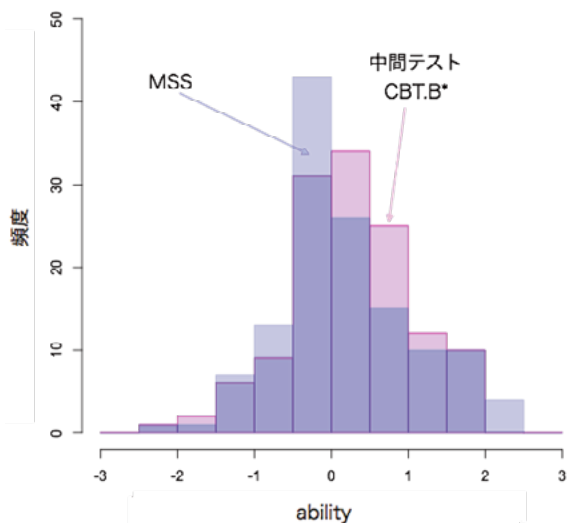


図7 線形代数B受講者のCBT.B*とLCT.Bのabilityのヒストグラム (情報学部)

図8、9は、線形代数B受講者のCBT.B*とMSSの関係を散布図に表したものである。それぞれ、正解率とabilityを評価の指標としている。相関係数を計算してみた値はそれぞれ0.62, 0.62である。先に示したLCT.AとCBT.B*の間の相関係数とほとんど変わらない。やはり、個人の習熟度が確率的にかなりゆらいで測定されていると考えられる。

4 2018年度工学部との比較

4.1 情報学部のLCT.Aと工学部のLCT.Aの比較

図10に、情報学部全学生と工学部全学生のLCT.Aの結果

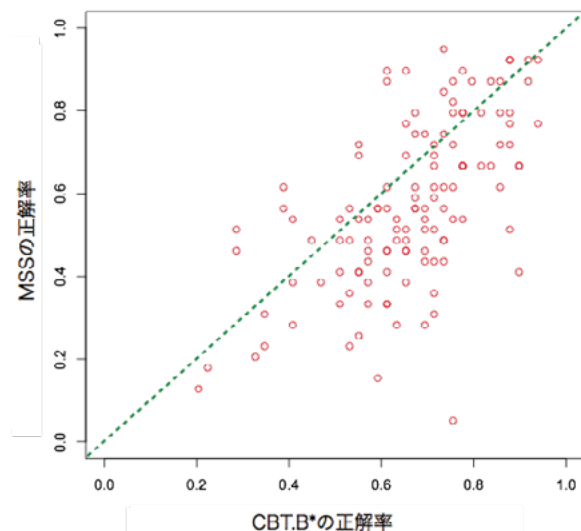


図8 線形代数B受講者のCBT.B*とMSSの正解率の散布図 (情報学部2018年入学生のみ)

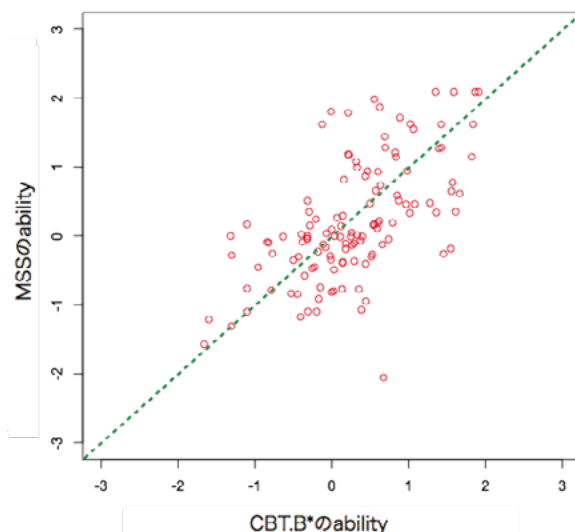


図9 線形代数B受講者のCBT.B*とMSSのabilityの散布図 (情報学部2018年入学生のみ)

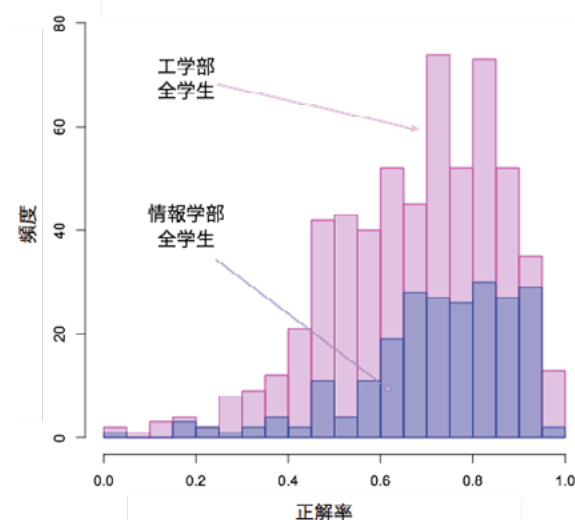


図10 情報学部全学生と工学部全学生のLCT.Aの正解率のヒストグラム

果の正解率のヒストグラムを示す。工学部より情報学部
の学生の方が正解率がやや高い傾向にあることがわかる。

4.2 工学部の LCT.A と LCT.B の比較

工学部では、中間試験はすべて記述式である。そこで、
情報学部での多肢選択式どうしでの比較に相当するもの
として、ここでは LCT.A と LCT.B の比較を行った。

図11に、工学部の LCT.A と LCT.B の正解率のヒストグ
ラムを示す。正解率は、LCT.A より LCT.B の方がやや低
くなっている傾向にあることがわかる。これは、線形代数
の内容が高度になっていき学生が難しく感じていることの
現れと思われる。工学部の LCT.B は、情報学部では CBT.
B* に相当し、ほぼ同じ比較を行っていると考えることが
できる。つまり、図9は図2に対応するものである。図12
に、工学部の LCT.A と LCT.B の ability のヒストグラム
を示す。

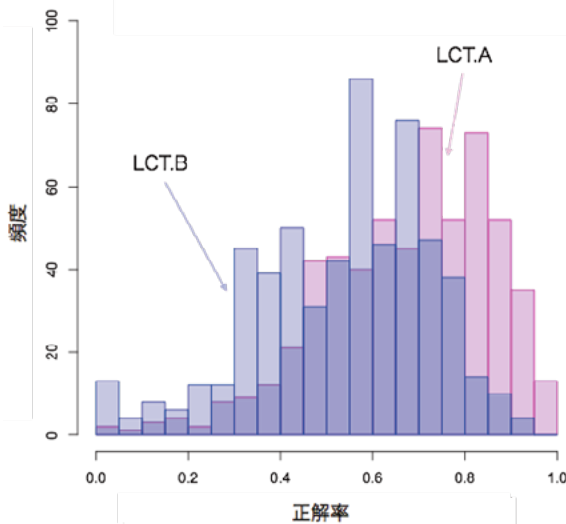


図11 工学部の LCT.A と LCT.B の正解率のヒストグラム

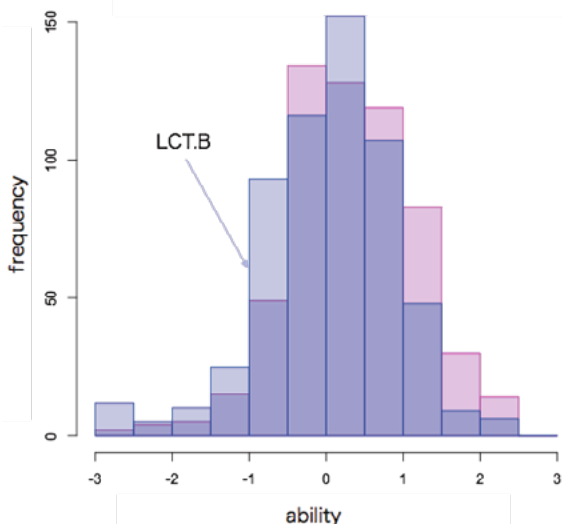


図12 工学部の LCT.A と LCT.B の ability のヒストグラム

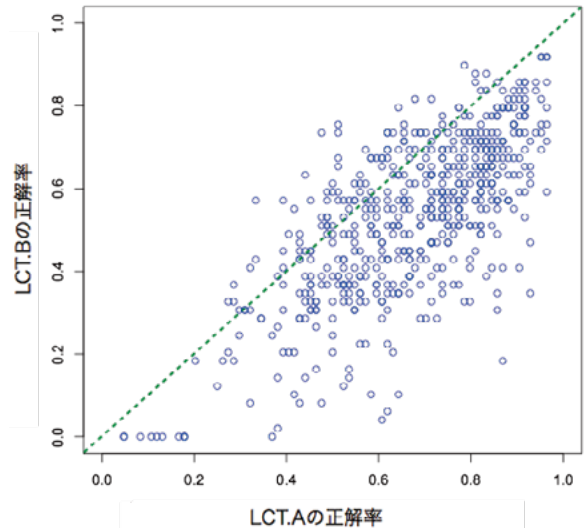


図13 工学部の LCT.A と LCT.B の正解率の散布図

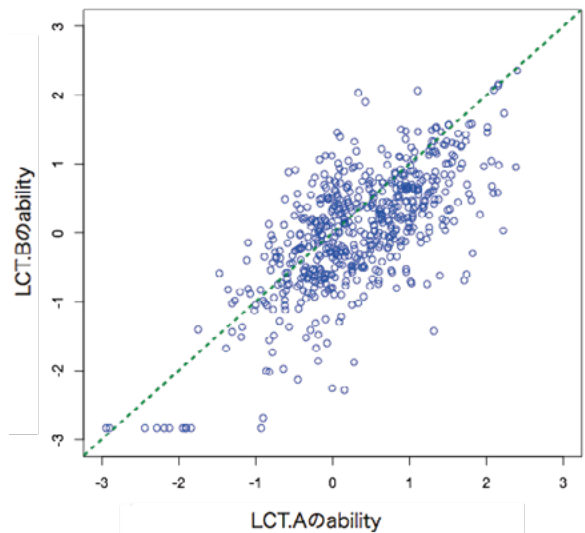


図14 工学部の LCT.A と LCT.B の ability の散布図

図13、14に、LCT.A と LCT.B の正解率および ability の
散布図を示す。両者の相関係数を計算してみると、図13の
正解率の場合は0.68、図14の ability の場合は0.64であっ
た。

5 2017年度前期での LCT.A と期末試験の比較

これまでほとんどの期末試験は記述式に依っていた。今
回、はじめて全問題を多肢選択式として中間試験および期
末試験で実施した。期末試験の評価が記述式の場合には教
員の評価基準が異なるために、両者の間には違いが出てく
るのではないかと想定される。しかし、直接的な比較はで
きないので、2017年度前期の LCT.A と期末試験の関係を
調べることによって、今回の試みの参考資料としてみたい。

図15、16、17は、2017年度の LCT.A の正解率と期末試
験評価結果 (0-59、60-69、70-79、80-89、90-100点の 5

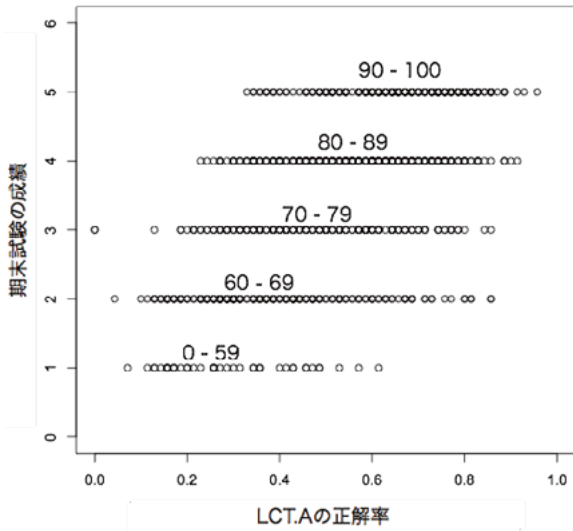


図15 全学部の LCT.A の正解率と期末試験の成績

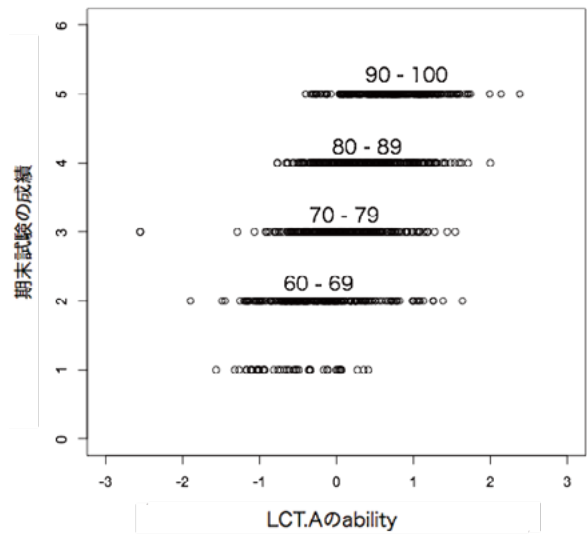


図18 全学部の LCT.A の ability と期末試験の成績

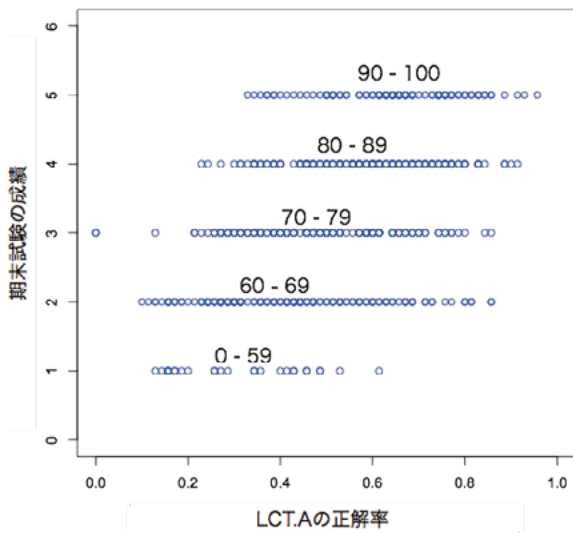


図16 工学部の LCT.A の正解率と期末試験の成績

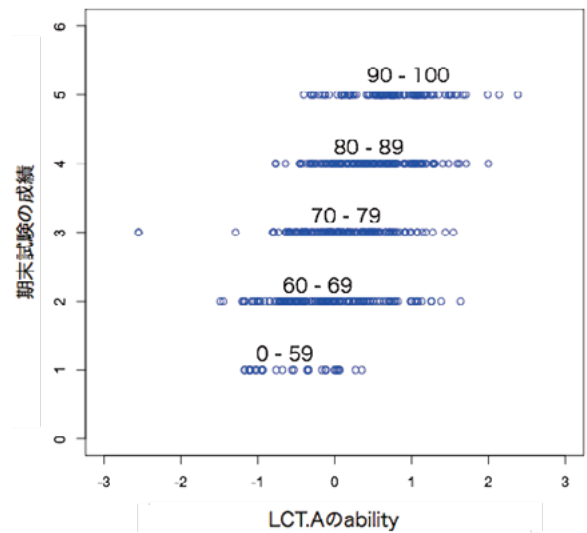


図19 工学部の ability と期末試験の成績

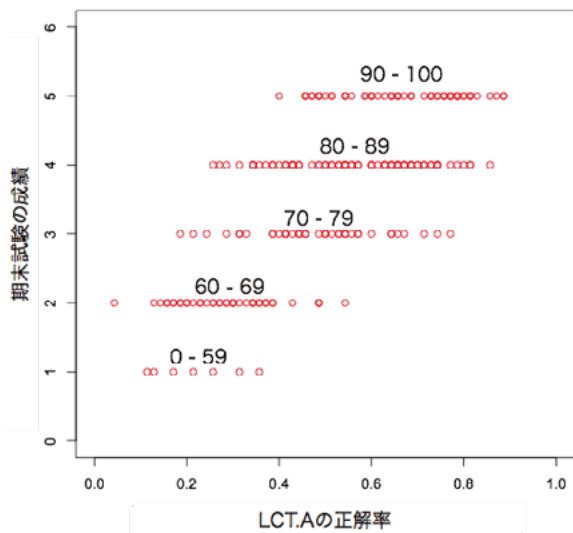


図17 情報学部の LCT.A の正解率と期末試験の成績

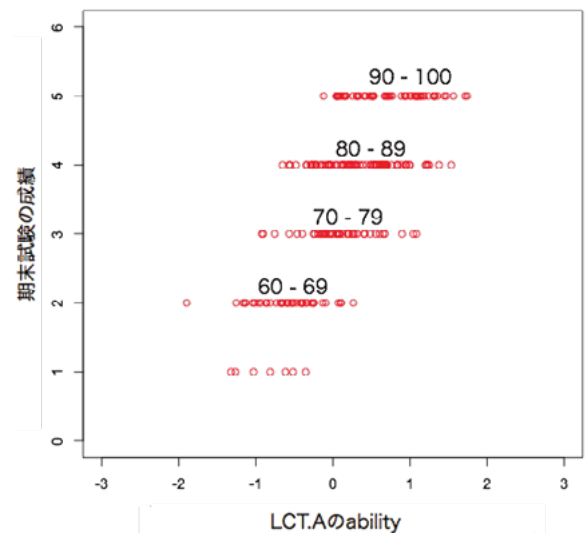


図20 情報学部の ability と期末試験の成績

段階評価) を、全学部、工学部、情報学部についての散布図を表したものである。

図15、16、17における相関係数は、それぞれ、0.58、0.51、0.70であった。

図18、19、20は、2017年度の LCT.A の ability と期末試験評価結果 (0-59、60-69、70-79、80-89、90-100点の5段階評価) を、全学部、工学部、情報学部についての散布図を表したものである。

図18、19、20における相関係数は、それぞれ、0.58、0.50、0.69であった。

6 プレースメントテスト結果との比較

LCTの結果から期末試験結果を予測することに主眼を置いてきたが、ここで、参考のため、プレースメントテストの結果と期末試験結果の相関を観てみたい。入学直後のプレースメントテストによって習熟度に応じてクラス分けを行い、習熟度別のクラスでの教育が普通に行われているが、プレースメントテストでのしきい値で分断されたため、たまたま異なった教育を受けさせられることへの判断の誤りはないのだろうか、という観点からである。

図21に、情報学部でのプレースメントテストのスコア(数IAまでの比較的基本的な問題22個(PTA)と数IIBまでの主に知識を問う55個の問題(PTB)の両方の成績(PTA+PTB)から求めた)と期末試験の成績との関係を示す。図で点線で囲んだ部分は、仮にプレースメントスコアが満点の半分程度しかなかった場合の期末試験の成績(IRTによるabilityの結果)を観てみたものである。極めて成績の悪い学生と極めて良い学生が同じ領域に存在していることがわかる。このときの両者の成績の間の相関係数は0.5-0.6程度であり、高くはない。

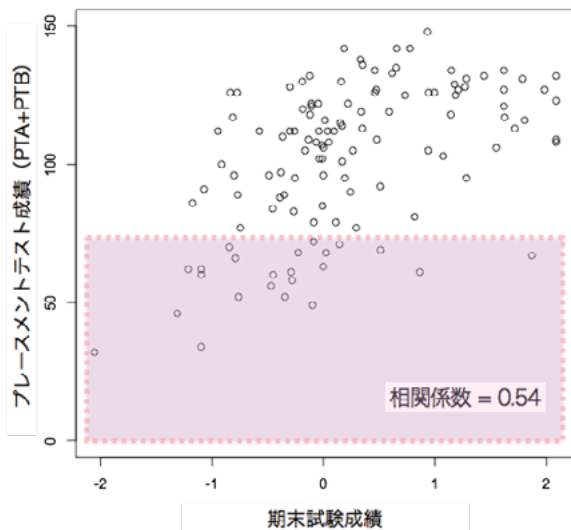


図21 プレースメントテストのスコア(総合スコア)と期末試験の成績の関係(情報学部)

プレースメントテストの総合スコアと期末試験の成績との相関だけでなく、もっと詳しく、PTA、PTB、PTA+PTB、期末試験の問題の正解数、IRTを使ったability値の関係をすべて比較してみたものが図22である。期末試験の成績とプレースメントテストのスコアとは単独でも総合でも軒並み相関係数が低い。

仮に、両者の成績が2変量正規分布に従っており、相関係数が0.6とした場合に、正規分布の平均から標準偏差だ

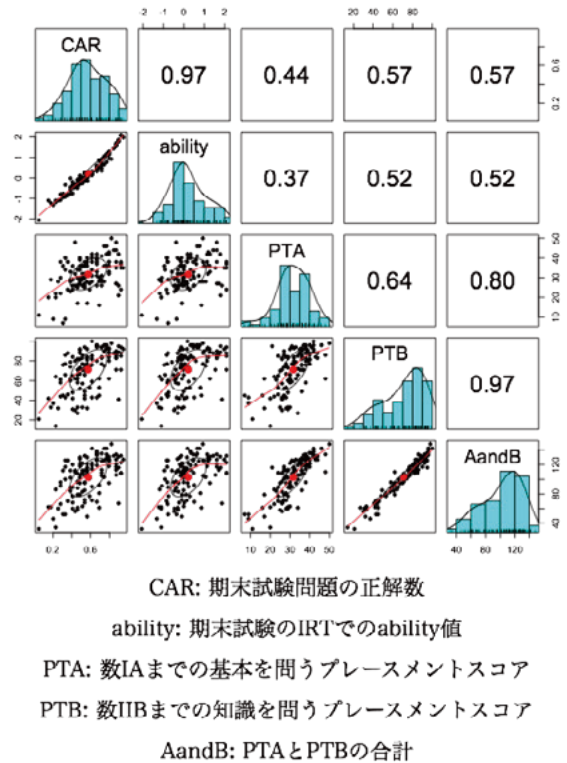


図22 プレースメントテストのスコアと期末試験の成績の関係(情報学部)

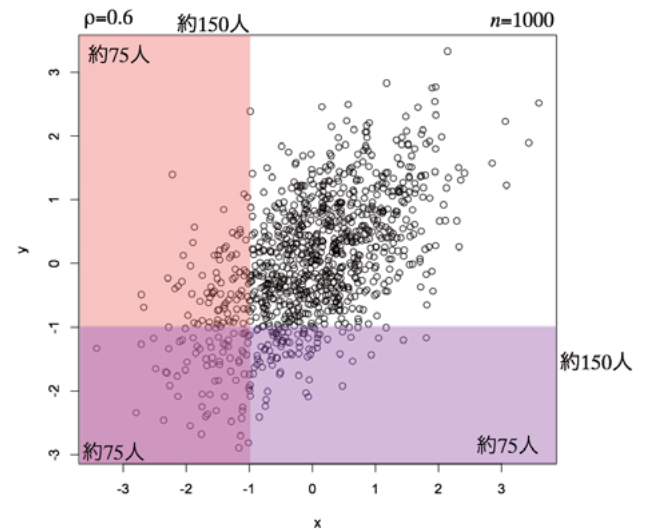


図23 2変量正規分布を仮定したときのドロップアウトリスクを持つ学生の把握

け下位にしきい値を置いた場合、1000人中150人程度がその中に入り、どちらの変数でもしきい値以下の学生数となると、さらにその半数で75人程度がその対象となることが図22から示される。この数は、1年生がドロップアウトする人数に対応しているとも考えられる（実際には50程度）。つまり、プレースメントテストでリスクのある学生であると予測された学生の半数はそうのように言い当てられているようにも見えるが、残りの半数はそれに該当しない、ということがわかる。

7 考察

これまでさまざまな評価結果について相関係数を計算してきた。見やすいようにここで一覧表（表2）にして再掲する。

表からは、すべての場合にわたって、相関係数の値は0.50-0.70の範囲に入っており、特に情報学部内だけでの比較では0.62-0.70と更に狭い範囲に納まっていることがわかる。このことは、試験の形態による成績の変動よりも、個人が持っている習熟度の確率的なゆらぎの影響が大きく、また、そのゆらぎは一定の値（0.6程度）からそれほど外れないことがわかった。

表2 相関係数の一覧表

学部	テスト	学部	テスト	評価数	相関
情報	LCT.A	情報	CBT.B*	正解率	0.63
情報	LCT.A	情報	CBT.B*	ability	0.62
情報	CBT.B*	情報	MSS	正解率	0.62
情報	CBT.B*	情報	MSS	ability	0.62
工学	LCT.A	工学部	LCT.A	正解率	0.68
工学	LCT.A	工学部	LCT.A	ability	0.64
全学	LCT.A	全学	期末	正解率	0.58
工学	LCT.A	工学	期末	正解率	0.51
情報	LCT.A	情報	期末	正解率	0.70
全学	LCT.A	全学	期末	ability	0.58
工学	LCT.A	工学	期末	ability	0.50
情報	LCT.A	情報	期末	ability	0.69

ゆらぎ（確率的変動）には、問題によって受験者自身の応答が変動するもの（受験者ゆらぎ）と、記述式テストを複数の教員が採点するような評価者によって解釈が異なることによって変動するもの（評価者ゆらぎ）の二つが考えられる。本論文では、同じ問題の多肢選択式のマークシート方式の試験を実施して評価者の影響が出ないような環境を整え、ゆらぎがどの程度になるかを観察した。

これまで、記述式問題を複数の教員が評価することばかりを行ってきたため、評価者ゆらぎがかなり大きく、正当で公平で正確な評価から少し距離があるものと考えてき

た。しかしながら、そういう状況下でのゆらぎと、今回新しく試みた環境下でのゆらぎとの間にそれほど顕著な差異が認められなかった。このことは、受験者ゆらぎの大きさは評価者ゆらぎが隠れるほど大きいとも受け取れる。このことは予想外であり、言い方を変えると、それほど受験者ゆらぎの大きさのことを自覚していなかったとも言える。大きな発見であった。

8 まとめ

LCTを使って、できるだけ早めに学生のドロップアウトのリスクを少なくしようと、確率付きで期末試験の成否のアラートを流すことまで予測手法は整ってきたことをこれまで報告してきた。その際、ある一定のところまでは予測が可能であるが破られない壁があると感じ、それを破るには評価法をできるだけ公平公正にする必要があると考えてきた。今回、多肢選択式のCBTあるいはマークシート方式のテストを採用することによって、予測精度が上がるのではないかと期待したが、予想を覆して、評価者のゆらぎは受験者自身の正答力のゆらぎを下回るのではないかという結果的になった。

LCTと記述試験による期末試験の成績、LCTとCBT、LCTと多肢選択式による期末試験の成績など、いずれの場合でも、それらの間の相関係数は0.5から0.7の間にあり大きな変化は見られない。広島工業大学の学生という、あるいは線形代数科目というような、サンプリングデータの性格に依存する結果とも考えられないことはない。しかし、同じ受験生が同様な試験科目を続けて2回受験するような別の環境下でもまったく同様な結果が出ているため、上のように考えることはある程度正しいのではないかと考えている。

文 献

- 1) 廣瀬、ラーニングアナリティクス：LCT成績と期末試験成績の関係、広島工業大学紀要教育編、pp. 59-63, Vol. 18, 2019.
- 2) 廣瀬、大規模オンラインテストから得られるラーニングアナリティクス、広島工業大学紀要研究編、pp. 159-166, Vol. 53, 2019.
- 3) 廣瀬、新入生全員を対象としたオンラインテストの実際、広島工業大学紀要教育編、pp. 27-35, Vol. 16, 2017.
- 4) 廣瀬、フォローアップクラスにおける授業設計について、広島工業大学紀要教育編、pp. 37-41, Vol. 16, 2017.
- 5) 廣瀬、大規模授業支援テストシステムとそのラーニングアナリティクス、統計数理、Vol. 66, No. 1, pp. 79-96, 2018.
- 6) 廣瀬、ラーニングアナリティクス指向学習支援、コ

- ンピュータ & エデュケーション (CIEC)、Vol. 45, pp. 23-30, 2018.
- 7) Hideo Hirose, Meticulous Learning Follow-up Systems for Undergraduate Students Using the Online Item Response Theory, 5th International Conference on Learning Technologies and Learning Environments (LTLE2016), pp. 427-432, 2016.
- 8) Hideo Hirose, Difference Between Successful and Failed Students Learned from Analytics of Weekly Learning Check Testing, Information Engineering Express, Vol 4, No 1, pp. 11-21, 2018.
- 9) Hideo Hirose, Prediction of Success or Failure for Examination using Nearest Neighbor Method to the Trend of Weekly Online Testing, International Journal of Learning Technologies and Learning Environments (IJLTLE), Vol 2, No 1, pp. 19-34, May 31, 2019.
- 10) Hideo Hirose, Key Factor Not to Drop Out is to Attend Lectures, Information Engineering Express, Vol 5, No 1, pp. 11-21, May 31, 2019.