

ラーニングアナリティクス： 授業確認テスト（LCT）の場合

廣瀬 英雄*

(平成28年10月24日受付)

Learning Analytics: in the Case of Learning Check Testing, LCT

Hideo HIROSE

(Received Oct. 24, 2016)

Abstract

In response to the extensive use of computer systems in learning, we are now in the age that we can do the learning analytics using the huge-size of database. We have recently established the follow-up program system (FP system) aimed at helping students who need basic learning and aimed at assisting teachers who have to engage in teaching a variety of educational students. The follow-up system consists of the learning check testing (LCT), follow-up program testing (FPT), and collaborative work testing (CWT). Since the FP system was first introduced, four months have passed, and we have accumulated the large scale of testing results. In the LCT we use the item response theory (IRT), which provides more accurate and fairer evaluations of individual abilities than classical test theory does, and thus the IRT has gradually been recognized as one of the proper evaluation methodologies in many testing fields. In this paper, we show the results of learning analytics focused on the LCT results. From the analytics, we have found that 1) five or six items are too small to evaluate the students' abilities, at least ten items are required, 2) although use of only one time testing results do not provide sufficient reliability to ability values, combination of multiple testing results provide ability values with a better reliability, which will give us a new measure to students' abilities, 3) the EMtype IRT worked in estimating the empty elements of item-user response matrix, 4) the learning analytics to LCT results showed us valuable information in leading us to a new direction.

Key Words: Learning analytics, Online testing, Follow-up program system, Learning check testing, Item response theory, Coefficient of correlation.

1 はじめに

高校を卒業した生徒が大学の入学試験に失敗する割合はわずかに5%という大学全入時代である。大学は多様な学

生を受け入れなければならない。そのため、高校までの知識が十分でないまま入学してくる学生には、彼らが学士力を身に付けることができるように、特別のケアができる仕組みを作っていく必要がある。広島工大ではこのことがで

* 広島工業大学環境学部建築デザイン学科, データサイエンスリサーチセンター

きるように2016年度からフォローアッププログラム (follow-up program, FP) を開始した。数学と物理と英語の基礎科目で重点的にそのことを行なっている。ここでは数学科目で行なっているFPの中のテストシステムから得られた分析結果を紹介する。

数学のFPは、授業時間でのオンラインテスト (LCT, learning check testing), フォローアップクラス (FPC) でのCWT (collaborative work testing), FPT (follow-up program testing) から構成されている。ここでは特に、LCTのラーニングアナリティクスについて述べることにする。

2 数学のフォローアッププログラム

図1に広島工大で取り組んでいる数学グループのFPの全体図を示す。図には、大きくLCT, CWT, FPTによって運用されていることが示されている。ここでLCT, CWT, FPTのそれぞれについて簡単に述べておく。

LCT: LCTは、当該授業の内容の理解度を確認するために授業後10分程度で行なう小テストである。全学生同一問題にオンラインで解答する。評価は項目反応理論 (IRT, item response theory) によって行なう。成績が一定水準に達しない場合には授業を補う意味でFPC出席の対象となる。LCT実施後直ちにFPCメンバーが確定する。

CWT: CWTは、FPCで行なうオンライン演習のことである。アダプティブオンラインIRTシステムを用いて学生が自ら演習を行う。この際、1テーブルあたり10人程度のFP対象学生が集まっているところに1人のピアサポーターが配置されている。FPCの担当教員は全体としてはこのピアサポーターを監督しながら授業を進行させる。つまり、上級生が1年生を指導するような形態である。このため、ピアサポーター自身の成長も促すように配慮されたシステムになっている。

FPT: FPTは、FPCの成果を確認するテストであり、

アダプティブオンラインIRTシステムを用いている。確認の範囲は、それまでの授業のすべてに渡っている。FPTに失敗した学生は翌週もFPCを受講することになる。

図2にFPシステムをハード的に支えているシステム構成を示す。このシステムは学内のポートフォリオシステムと一体になっているので、LCT, CWT, FPTの評価結果は、学生本人のみならず、科目担当教員、学科の学生アドバイザー、学科内で割り振られたチューター教員、学生支援センターに迅速に連絡できるように連携がはかられている。

the follow-up system in our university

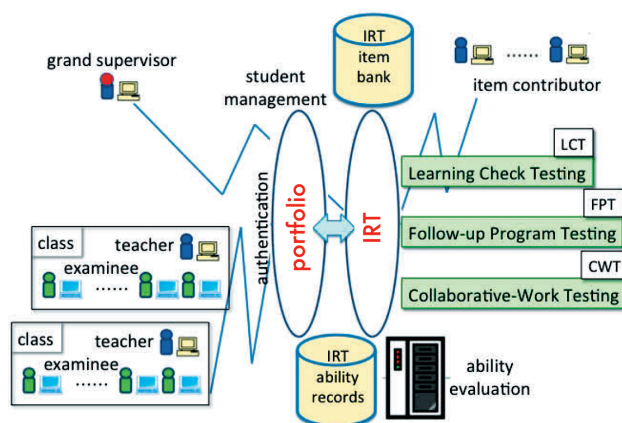


図2 フォローアッププログラムのシステム構成

3 LCTを実施し始めた状況

LCT実施については毎週実施ということでスタートした。始めた当初は、WiFi環境の不安定さ、学生のPC操作の不慣れなどで、わずか10分のテスト時間にもかかわらず30分以上を費やして試験することがあった。また、LCTの目的がFPC参加学生を選ぶためと位置づけてとらえていたため、LCT受験のインセンティブとして希薄であった。そこで、何回目かの授業時間に「FPの意味と意義、その中でのLCTの位置づけ」ということできちんと説明を行った。学生は理解してくれた。

5月になると、WiFi環境の不安定さと学生のPC操作の不慣れはともに安定してきて、テストも15分程度の時間内で済むようになってきた (ただし、トラブルに見舞われる学生が皆無になったということではなく、1クラスに数人は毎回何らかの不都合で受験ができない)。また、オンラインテストを教員も学生も負担に感じなくなった。学生の方は、かえって、毎週の授業開始前の10分間の緊張感と受験への準備からか、学生はLCTを前向きに受け入れてくれるようになっていた。そうしたこともあってか、LCTは自分の学習に役に立つツールであるという認識ができ上がってきた。アンケートにもその結果が表れている。中間まとめ以降は事前に決めてあった隔週でのLCT実施

the follow-up program in our university

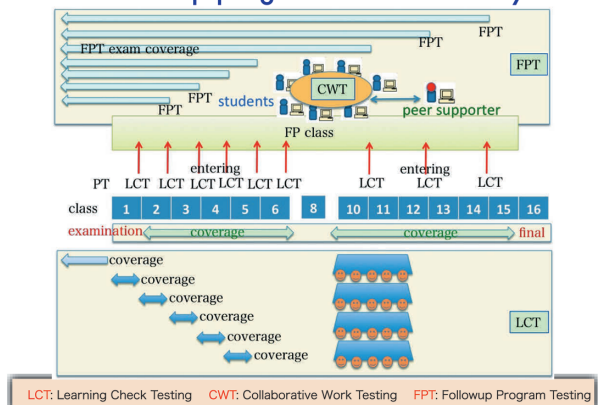


図1 フォローアッププログラムのテストシステム

とした。従って、中間まとめまでは6回、中間まとめ以降では3回の合計9回実施ということになる。解析では講義と演習がセットになっているので同じ科目で30週、一方、線形代数では演習を設けていないので15週の授業時間であるが、LCT実施の頻度については、解析基礎でも線形代数でも同じ扱いとした。

さて、LCTでは古典的なテスト理論に基づいた評価法を使っていない。現代テスト理論として位置づけられる項目反応理論(IRT, item response theory)を用いている。従って、一人だけの解答結果からその学生の評価値は得られない。クラス全員、あるいは対象学年全員の解答結果がそろった時点で、問題の困難度とともに学生の評価値を計算して、より公平で公正な評価を行おうとしている。従って、あるタイミングでないと評価値は得られない。このため、受験直後から評価値が返ってくるまでに少し時間がかかる。このため、FPCに参加する学生を選ぶのに少し時間を要すことになり、そのためFPC参加要請連絡がFPCの直前になることもあった。そういう意味では即座の応答ができない不便さはあったが、反面評価には正確を期すという面で今回のLCT実施によって多くの知見が得られた。

4 項目反応理論 (IRT, item response theory)

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、各受験者の評価値に加えて問題の難易度も同時に求めながら、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 $P_j(\theta_i)$ がロジスティック分布、

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

に従っていると仮定する。 a_j , b_j , c_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）、困難度（文字どおり、問題の難易度を表す）、当て推量（偶然に正答する確率を表す）、 θ_i は受験者 i の学習習熟度（ability）を表している。1.7は分布が標準正規分布に近くなるように調整された定数である。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、 $\delta_{ij}=1$ 、誤答なら δ_{ij}

$=0$ 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、 $c_j=0$ と仮定した場合、

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

と表される。これを尤度関数という。図3に、項目反応理論の概念図を示す。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。

ここで、なぜ古典的な評価法ではなくIRTを使った評価法が適切なのかについて考えてみる。

今、A, B両君が13問の数学問題に挑戦し、 δ の値が問題順に、

A 1111110001011

B 1011110011011

であったとする。2問目と9問目で正誤が入れ替わっているだけで他は同じ解答パターンなので、正答率はどちらも同じ値0.69となる。しかし、A, B以外の受験者も加えてIRTを使って問題の難易度を計算してみると、2問目では2.3、9問目では1.2なので、2問目を正答したA君の方が学習習熟度が高いと考えるのが自然であると思われる。実際、A, B両君のabilityを求めてみると、それぞれ1.70, 1.56であった。IRTは自然な配点を自動的にこなしていることが分かる。この例はIRTの方がよりふさわしい学習習熟度の評価値を与えていることを示唆している。

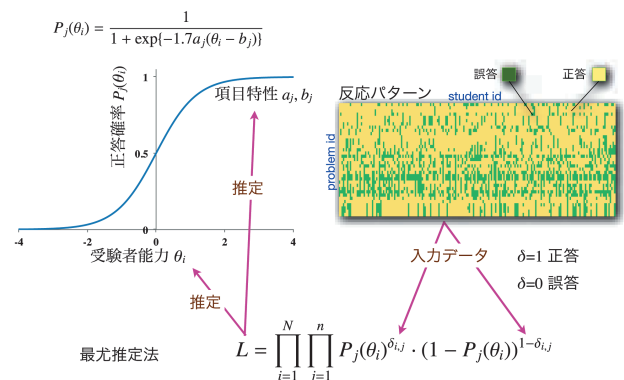


図3 項目反応理論の概念図

5 LCTの試験法の設定について

LCTを毎週実施する場合の試験時間はせいぜい10分であろう。それ以上になると講義や演習の時間を圧迫すると思われる。では、その10分で何問の問題を解答できるだろうか。これまで、線形代数や確率と統計のテキストに付随したオンライン演習「愛あるって」では5問を1セットとしていた。10分程度で問題を解くにはせいぜい5問程度が限度であろうということから設定していた。しかし、

「愛あるって」の結果を分析した結果、5問では ability の推定値にバイアスが出ることが分かってきた。この原因は、アダプティブテストを使う途中の推定を安定的に行なうためにベイズ法を取り入れているため、結果的に事後分布での推定値が事前分布（標準正規分布）に影響を受けている。これは、shrink 現象とも呼ばれている。この影響は結構大きく、問題数が5問程度であると推定精度が良くなく推奨できる問題数ではない。シミュレーションからは最低でも10問は必要であるという結果が出ている。そこで、10問実施を提案したが、やはり10分の試験時間の壁から譲っても6問ということになった。そこで6問、10分という設定でスタートした。実施してすぐに ability の推定精度が良くないことに気がつき、中間まとめ以降は、隔週実施としていたこともあり、15分で10問という設定に変更した。

6 LCT による結果について

シミュレーションでは受験者の能力値が分かっているの推定値と比較することで推定の正しさを検証できるが、実際の学生の受験結果から LCT の結果が正しい推定値を与えているかどうかをどのように測るかは難しい問題である。ここでは、入学時に全員一斉に実施するプレースメントテスト (PT) の結果と比較することを考えた。PT の試験時間は1時間程度であるが、基本的な問題が多く出題されているため、評価結果は一定程度の精度を持っているものと考えた。

図4に、1回目と2回目の LCT の解答パターンを示す。1は正答、0は誤答、空欄は未回答を表す。通常の IRT では未回答の空欄があると計算できない。そこで、図5に示すように、1回目と2回目の LCT 反応パターンの中の未回答部分はすべて0に変更して通常の IRT で計算してみた。最初に LCT01 (1回目の LCT) の結果を PT と比較した結果を図6に示す。この結果を見て驚いた。PT の評価値がある程度の精度を持っていると仮定した場合、LCT の結果をどのように評価したらよいのだろうか。まったく相関関係が見えない。まだ入学して間もないので教わっている内容は異ならない。LCT の評価結果にはあまり意味を持たないということだろうか。なお、解析基礎 (LCT_AB) と線形代数 (LCT_LA) との比較ではまったく相関が見えない。わずか6問ではあるが、LCT の結果のスコアはとびとびの値をとっていないことが示され、IRT を用いて評価している様子が分かる。同じ正解数でも ability の評価値は解いている問題によって異なるので微妙な変化を見せている。

student_id	LCT01						LCT02					
	1	2	3	4	5	6	1	2	3	4	5	6
01	1	1	1	1	1	0	1	1	1	1	1	1
02	1	1	0	1	1	0	1	1	0	1	1	1
03	0	1	1	1	1	0	1	1	1	1	1	1
04	1	1	0	1	1	1	1	1	1	1	1	1
05	1		1		1		1	1	1	1	0	0
06												
08	1	1	1	1	1	1						
09	1	0	1	1	1	1	0	1	0	1	1	1
11	1	1	1	1	0	0						
12	1	0	1	0	1	0	1	1	1	1	1	1
13	0	0	1	0	1	1	1	1	0	1	1	1
15	0	0		0	0	0	0	1	0	1	1	0
16	1	1	1	1	1	1	1	1	1	1	1	1

図4 1回目と2回目の LCT 反応パターン (オリジナル)

student_id	LCT01						LCT02					
	1	2	3	4	5	6	1	2	3	4	5	6
01	1	1	1	1	1	0	1	1	1	1	1	1
02	1	1	0	1	1	0	1	1	0	1	1	1
03	0	1	1	1	1	0	1	1	1	1	1	1
04	1	1	0	1	1	1	1	1	1	1	1	1
05	1	0	1	0	1	0	1	1	1	1	0	0
06	0	0	0	0	0	0	0	0	0	0	0	0
08	1	1	1	1	1	1	0	0	0	0	0	0
09	1	0	1	1	1	1	0	1	0	1	1	1
11	1	1	1	1	0	0	0	0	0	0	0	0
12	1	0	1	0	1	0	1	1	1	1	1	1
13	0	0	1	0	1	1	1	1	0	1	1	1
15	0	0	0	0	0	0	0	1	0	1	1	0
16	1	1	1	1	1	1	1	1	1	1	1	1

図5 1回目と2回目の LCT 反応パターンの処理法 (未回答部分はすべて0に変更)

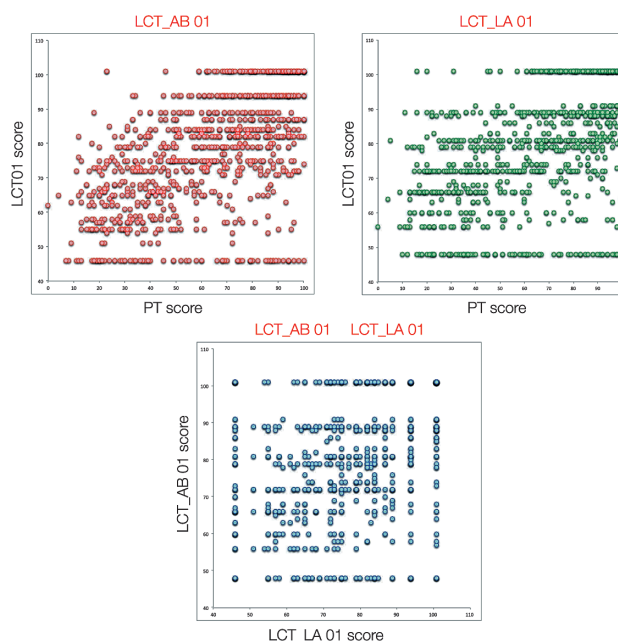


図6 1回目の LCT と PT とのスコア比較 (未回答部分はすべて0に変更)

1回目だけでは分からないので1回から4回までのLCTの結果をPTと比較してみることにした。それを図7, 8に示す。未回答部分はすべて0に変更して計算し, IRT計算結果後はそのスコアを0とした。やはり相関は見えない。全体的にLCTの評価精度が悪いことが分かる。初期には学生が戸惑ってabilityの推定結果がPTと合わない場合があるだろうということは考えられる。しかし, 回を重ねてもあまり変化は見られない。

奇妙な現象は, PTを標準とすると, PTで成績のいい学生はもちろんLCTでもいい成績を見せているが, PTで成績の思わしくない学生はLCTで最低から最高までほぼ一様に分布しているということである。途中のPTの成績の中頃の学生も自分のPTの成績を下回らない結果を見せている。これはどういう現象であろうか。LCTではパソコンを使っているが隣のパソコンを見ようと思えば見られないことはない。また, 隣どうして話し合って正解と思われる解答を入力することもできなくはない。LCT実施時には, 教科書は見てよいというような雰囲気もあり, 一部の学生には厳密な評価はできていないのかもしれない。チートはいけないことは知っているが, FPCに行きたくないためそのような行動に出ることは考えられる。LCTは学生にも評判がいい小テストとしての評価もできるので中間まとめの代用にするのかも考えられるが, LCTの結果を評価に使う際には試験環境を十分確認して評価する必要があると思われる。

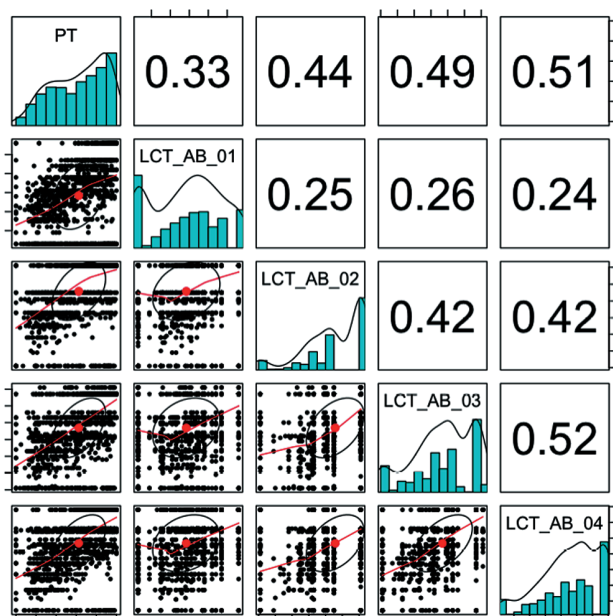


図7 解析基礎 LCT スコア比較 (未回答部分はすべて0に変更して計算, そのスコアも0とした)

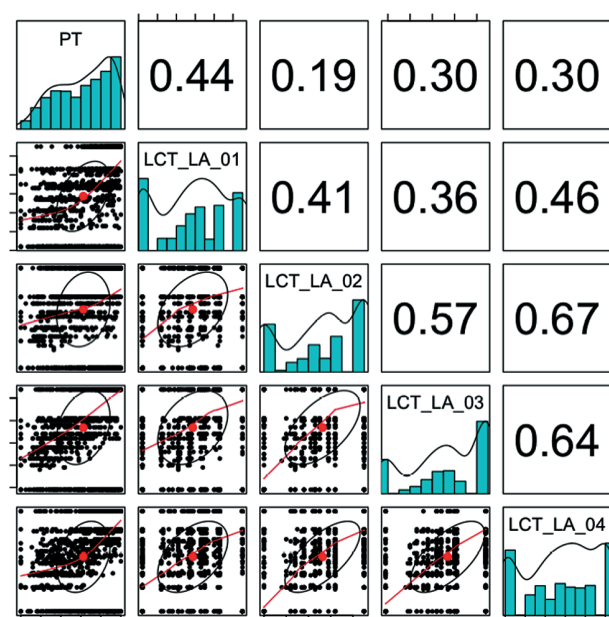


図8 線形代数 LCT スコア比較 (未回答部分はすべて0に変更して計算, そのスコアも0とした)

未回答部分をすべて0に変更して計算した後, そのスコアも0としたことで, 相関係数に偏りが出ているかもしれないと考えて, 未回答部分はすべて0に変更して計算した後で, そのスコアを空欄とした場合の結果を図9, 10に示す。LCTスコアの0の部分がある程度消えているが, 図7, 8は図9, 10とあまり変わらない。

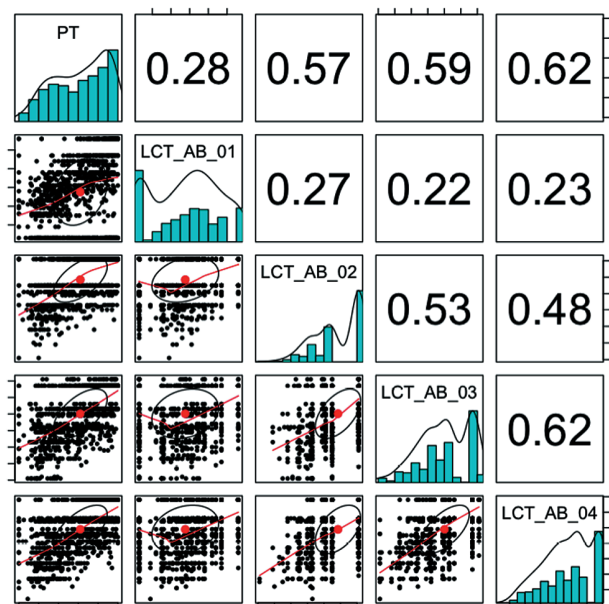


図9 解析基礎 LCT スコア比較 (未回答部分はすべて0に変更して計算, そのスコアは空欄とした)

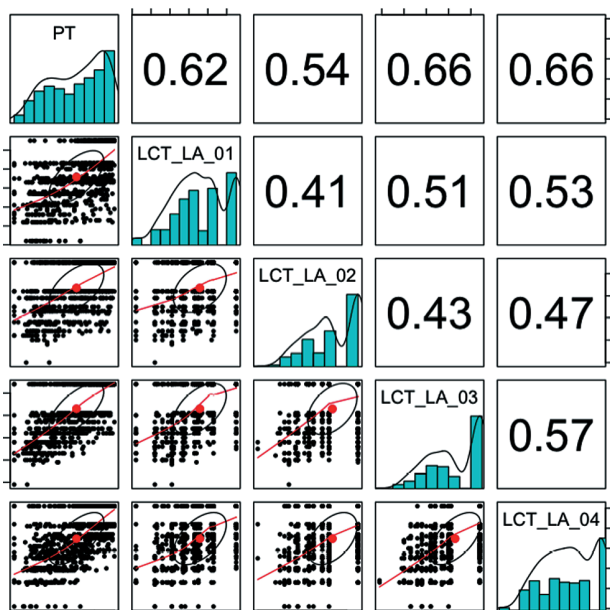


図10 線形代数 LCT スコア比較 (未回答部分はすべて0に変更して計算, そのスコアは空欄とした)

LCTの各回単独での評価ではあまり良い精度が出ないと思われるので、2回分をまとめて評価したらどうなるかを調べてみた。2回なら12問の解答ということになり回答数も多くなる。しかし、もっと他のメリットがある。1回だけの試験では、例えば図6のid008の学生は1回目出席で2回目は欠席している。2回目だけで評価すると0点になるが、1回目には全問に正解していることから2回目にもし受験していたらかなり良い評価結果が出るだろうと思われる。そこで、空欄を推定しながら埋めていくという方法をとった。このようにすればid006の学生の空欄も埋めて計算することができる。これにはEMtypeIRTを使えば計算が可能である。その様子を図11に示す。未回答部分の回答への推定値が得られていることが分かる。そのように処理した後のLCTのabilityの推定値は図12のようになった。各回単独での評価結果と比較してPTスコアとかなりよい相関を見せていることが分かる。同様な方法で、3回目と4回目をセットにしてabilityを求めたものを追加して相関図を求めたものが図13である。6問だけの評価から12問への評価に変えるだけでかなり改善されることが分かる。そこで、評価だけで考えると、毎週の結果を使うよりも2週間分をまとめて評価する意味で、隔週実施のLCTには意味があることになる。更に、LCTの評価をFPC参加に使うとか、あるいは中間まとめの代用として使うような場合には、できるだけ多くのテスト結果をまとめて用いることが好ましいということも分かる。

student_id	LCT01						LCT02					
	1	2	3	4	5	6	1	2	3	4	5	6
01	1	1	1	1	1	0	1	1	1	1	1	1
02	1	1	0	1	1	0	1	1	0	1	1	1
03	0	1	1	1	1	0	1	1	1	1	1	1
04	1	1	0	1	1	1	1	1	1	1	1	1
05	1	1	1	1	1	1	1	1	1	1	0	0
06	0	0	0	0	0	0	0	0	0	0	0	0
08	1	1	1	1	1	1	1	1	1	1	1	1

空欄推定値	2	3	4	5	6	7	8	9	10	11	12
	01	1	1	1	1	1	0	1	1	1	1
02	1	1	0	1	1	0	1	1	0	1	1
03	0	1	1	1	1	0	1	1	1	1	1
04	1	1	0	1	1	1	1	1	1	1	1
05	1	0.42	1	0.73	1	0.38	1	1	1	1	0
06	0	0	0	0	0	0	0	0	0	0	0
08	1	1	1	1	1	1	0.98	1	0.88	0.99	0.98
09	1	0	1	1	1	1	0	1	0	1	1

図11 1回目と2回目のLCT反応パターンの処理法2 (一部の未回答部分は空欄のままで反応パターンを未知数として推定し, すべての未回答部分はすべて0とした)

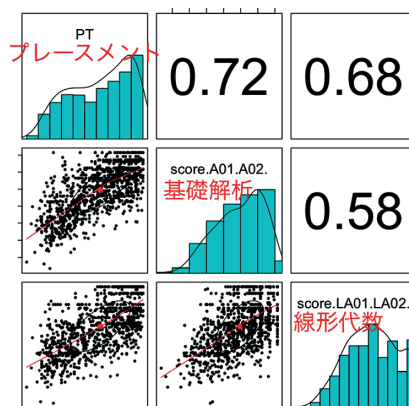


図12 1回目と2回目のLCT反応パターンを使ったLCTスコア比較1

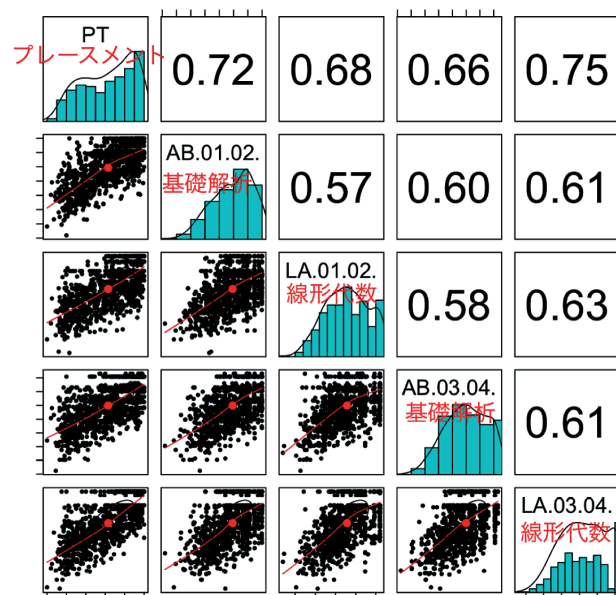


図13 1回目と2回目のLCT反応パターンを使ったLCTスコア比較2

そこで、LCTの結果を1回目から6回目までの中間まとめまでのものとその後の8、9、10回目（7回目は中間まとめに相当するので番号を付けていない）までのものについて、各回のLCTのabilityの平均値をとって、PT、解析基礎LCT、線形代数LCTの相関を見てみたものが図14である。ここでは、前述したような2回のデータをまとめてEMtypeIRTを使って処理してはならず、通常のIRTで計算している（従って、未回答の空欄はすべて0を入れて計算している）。しかしながら、その結果は驚くべきものである。PTとLCTとの間にはかなり明瞭な相関が見られる。解析基礎と線形代数とで比較すると、PTとの相関は解析基礎の方が高いようである。これは、解析基礎の内容が高校数学により近いためであろうと推測される。このことは、これまでの分析結果と同様である。更に、LCTだけについてみると、解析基礎と線形代数のLCTのabilityの平均の間にはかなり高い相関が得られていることが分かる。つまり、解析基礎ができる学生は線形代数もできているし、逆にできていない学生はどちらもできていない、ということで、どちらかが得意というような現象にはなっていない。これは、各回のLCTの結果の平均をとった操作から得られた知見である。平均をとるという操作は、確率論からは、大数の法則と中心極限定理から確率的な変動要因が除去されて真の姿が見えてくるというものと捉えられるが、図14はまさにこのことを示している。各回でのLCTではチートにより見せかけの良い成績を得ることがあるかもしれないが、それはいつも同じ状態で継続できるものではない。あるときには実力以上の結果は出せない。平均化はそのような雑音を除去してくれている。

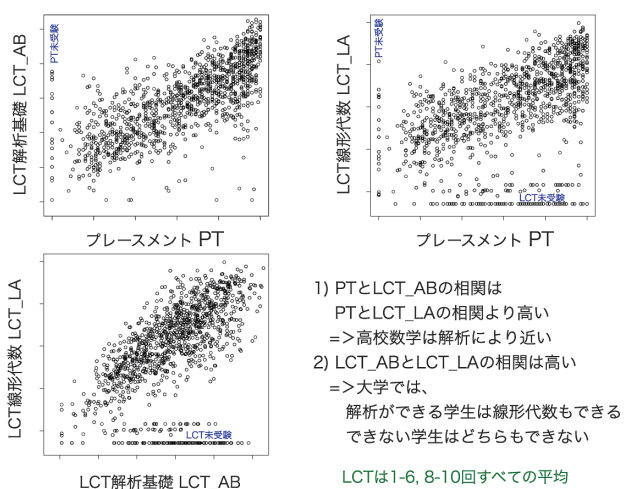


図14 1-6回目、8-10回目のLCTのabilityの平均とPTとの相関図（LCTで、未回答部分はすべて0に変更）

7 LCTでのIRT評価について

LCTではIRTを用いているが、FPTやCWTのようにadaptiveなテストではなく一斉テストなので、FPTやCWTと違って問題の困難度と受験生のabilityとを同時に求めることができる（FPTやCWTでは困難度はあらかじめ準備しておく必要がある）。これまで分析してきた内容ではすべて問題の困難度と受験生のabilityとを同時に求めたものを使っている。ここでは、LCTでもFPTやCWTのように問題の困難度を固定化して計算した結果と、ここまで用いてきた方法で求めた結果とを比較してみる。LCTでFPTやCWTのように問題の困難度を固定化して計算したものを使うことができれば、LCT受験後、即座にその結果を受験生に知らせることができる。ここではその可能性について述べる。精度が良ければ簡便法として使うことができるという意味である。

図15、16は、解析基礎および線形代数について、1-6回目、8-10回目について、両者で求めたLCTのabilityを比較したものである。横軸に、困難度とabilityとを求めた通常のIRTのability値、縦軸に、困難度を固定してabilityだけを求めたときのability値を示す。

図から、両者はほぼ対応した関係にあることが分かるが、困難度を固定してabilityだけを求めたときの方法では、ランクを求める際に不正確であることが分かる。例えば、図15の2-3回目のLCTではそれが顕著に現れている。1回目のLCTで簡便法でも大丈夫かもしれないと思われたが、2-3回目の結果を見て危険であることが分かった。これは、線形代数でも同じであり、6回目はその傾向が著しい。この結果から、簡便法は用いてLCTのabilityの近似値を求めることは行わないことにした。

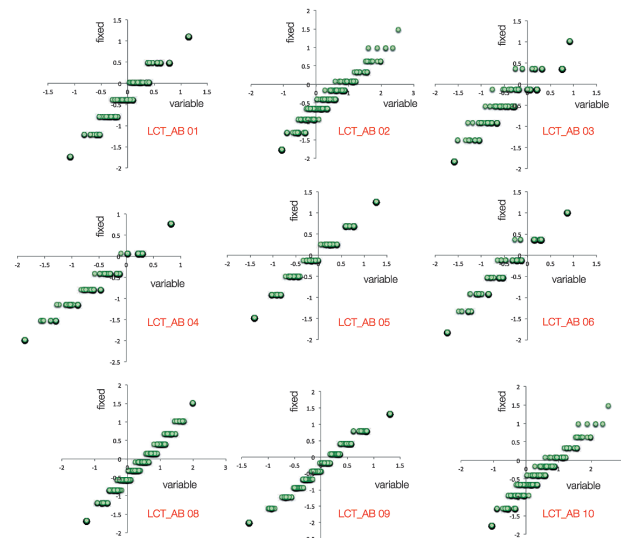


図15 IRT計算法の違いによるLCTのability値の比較（解析基礎）

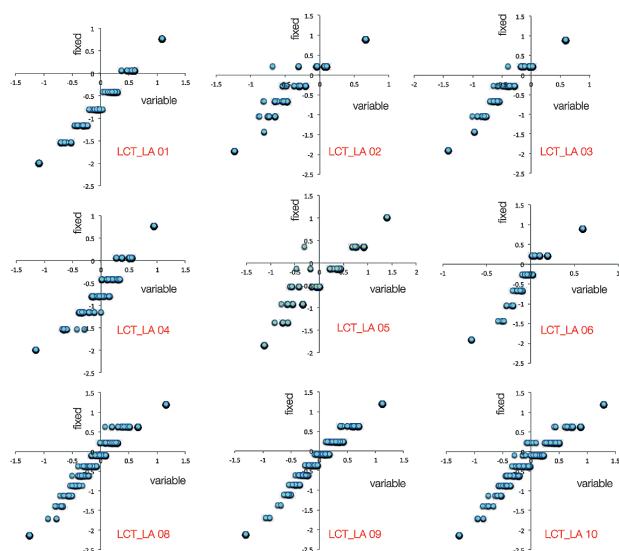


図 16 IRT 計算法の違いによる LCT の ability 値の比較 (線形代数)

8 まとめ

多様な学生の教育をうまく進める目的でフォローアッププログラム (FP) がスタートした。数学グループでは、このプログラムの中に、授業の理解度を確認する意味で LCT (learning check testing) というテストシステムを作り、2016 年の 4 月から毎週実施してきた。LCT は、最初は FP の対象となる学生を選ぶ意味が大きかったが、学生の受け入れの姿勢はかなり positive で、LCT 自体が意味を持つようになってきた。そこで、LCT の働きについて、統計的に分析しておこうというのが本論文の動機であった。

今回、LCT 結果の分析を行うことによっていろいろなことが分かってきた。それらを列挙すると：

1) 1 回の LCT は、問題数が 6 問と少ないので高精度の ability の推定結果は得られないかもしれないが、ある程度精度は得られると思っていた。しかし、プレースメントテスト (PT) と比較するとほとんど相関が得られていないことが分かった。

2) そこで、2 回分を一つにまとめて、EMtypeIRT を使って未回答部分を推定しながら ability の推定を行うと格段に推定精度が上がった。このことから、中間まとめ以降は隔週実施への方向を選んだ。

3) 前期でひととおり LCT が終わってすべての LCT の ability 結果の平均をとると、PT と LCT との間にきれいな相関が見られ、また、解析基礎と線形代数の間にも高い相関が見られた。つまり、毎週の LCT だけでは見えなかった真の姿が統計的に平均をとることで見えるようになり、各テストの間の高い相関が確認された。

4) 解析基礎 LCT と PT との間の相関は線形代数 LCT

と PT との間の相関よりも高い。また、解析基礎 LCT 線形代数 LCT と PT との間の相関はかなり高い、ということが分かった。

5) このことから、入学してくる学生の習熟度は明確に 3 つ程度に分けられる。LCT のトレンドは半期の中で変化を見せる可能性があるが、実際には 3 グループはあまり変動していない。このことから、特の下位のクラスには入学前と直後のリメディアル教育が必要であり、また重要であることを示唆している。

6) LCT で FPT や CWT のように問題の困難度を固定化して計算した (簡便法) 結果と、問題の困難度と ability とを同時に求める従来の IRT の方法で求めた結果とを比較してみた。その結果、簡便法で LCT の ability を出すには精度が不十分であることが分かった。

9 付録

LCT の評価値は IRT によったものをこれまでに示してきた。ここでは補足として古典的な評価法である正答率を示しておく。

未回答部分については、1) すべて誤答として処理した場合、2) すべて除外して処理した場合の 2 つの場合を併記した。学生がすべてを解答するまでに時間が足りなかったのか (この場合、誤答処理が適切)、欠席だったのか (この場合、誤答処理が適切)、未受験だったのか (この場合、除外処理が適切)、WiFi の不具合などの理由で解答できなかったのか (この場合、除外処理が適切) が混在しているからである。実際の正答率は両者の間にあると考えられる。

図 17 に、解析基礎 A の LCT についての正答率の棒グラフを示す。未回答 (空欄) はすべて誤答として処理している。また、そのヒストグラムを図 18 に示す。

図 19 に、線形代数 A の LCT についての正答率の棒グラフを示す。未回答 (空欄) はすべて誤答として処理している。また、そのヒストグラムを図 20 に示す。

図 21 に、解析基礎 A の LCT についての正答率の棒グラフを示す。未回答 (空欄) はすべて除外して処理している。また、そのヒストグラムを図 22 に示す。

図 23 に、線形代数 A の LCT についての正答率の棒グラフを示す。未回答 (空欄) はすべて除外して処理している。また、そのヒストグラムを図 24 に示す。

正答率は回答数に対する正答数から算出したもので古典的な評価法の代表的なものである。次に、IRT を使って計算した問題の困難度 b について求めたものを示す。

解析基礎 A の LCT についての正答率を棒グラフで表した図 17 に相当する IRT の困難度 b を表したものを図 25 に示す。また、線形代数 A の LCT についての正答率を棒グラフで表した図 19 に相当する IRT の困難度 b を表した

ものを図 26 に示す。

正答率は IRT の困難度 b に対応しているので、解析基礎 A について両者の関係をプロットしてみたものを図 27 に示す。また、線形代数 A について両者の関係をプロットしてみたものを図 28 に示す。IRT では正答する確率の分布を標準正規分布に対応させているので、困難度 b と正答率の 2 つは正規分布の分布関数を左右逆にしたような（生存関数のような）関係になっていることが分かる。

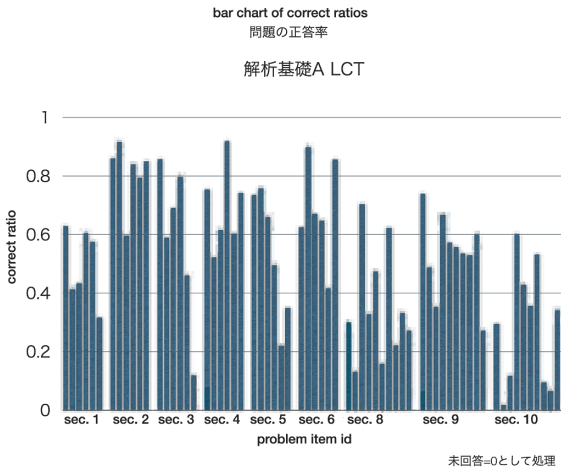


図 17 解析基礎 A の LCT の正答率の棒グラフ (未回答 (空欄) はすべて誤答として処理)

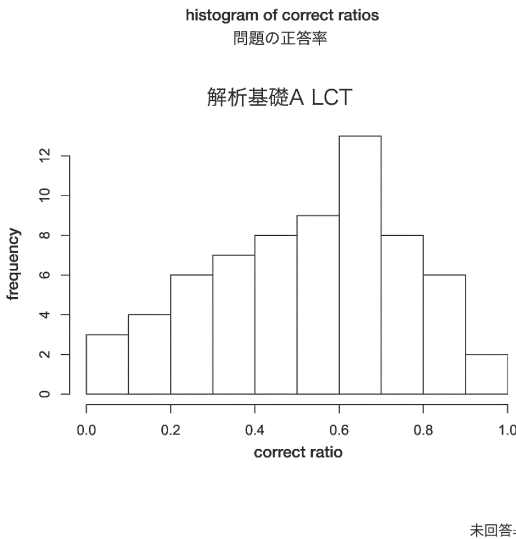


図 18 解析基礎 A の LCT の正答率のヒストグラム (未回答 (空欄) はすべて誤答として処理)

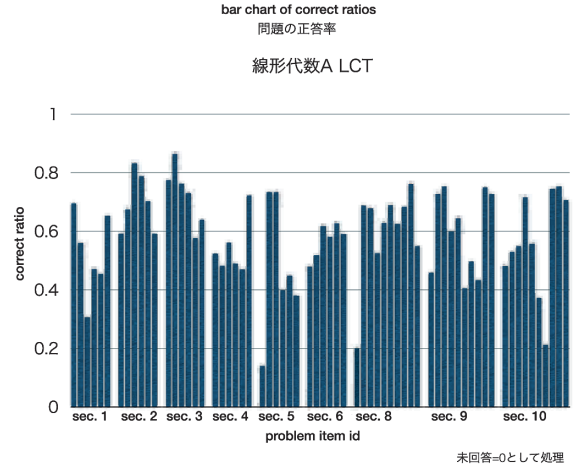


図 19 線形代数 A の LCT の正答率の棒グラフ (未回答 (空欄) はすべて誤答として処理)

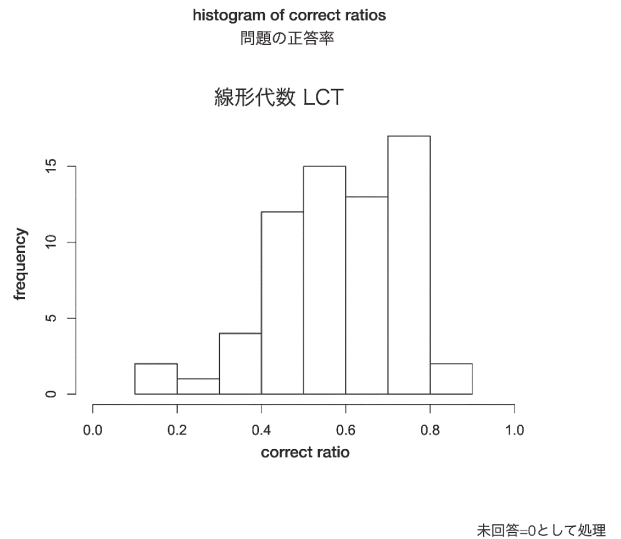


図 20 線形代数 A の LCT の正答率のヒストグラム (未回答 (空欄) はすべて誤答として処理)

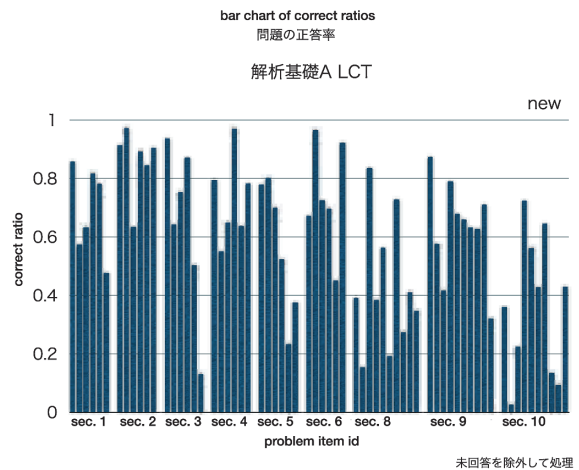
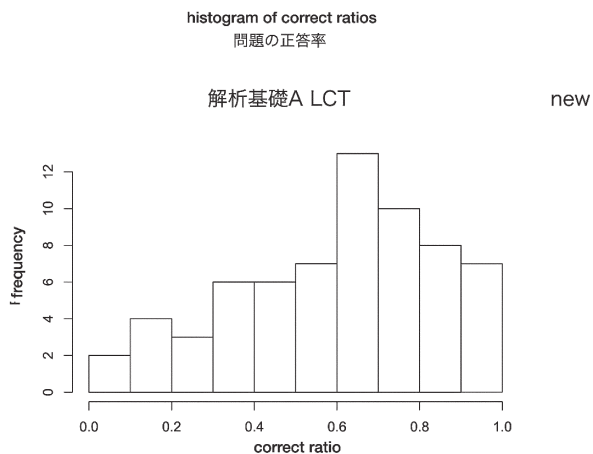


図 21 図 17 解析基礎 A の LCT の正答率の棒グラフ (未回答 (空欄) はすべて除外して処理)



未回答を除く処理

図22 解析基礎 A の LCT の正答率のヒストグラム
(未回答 (空欄) はすべて除外して処理)

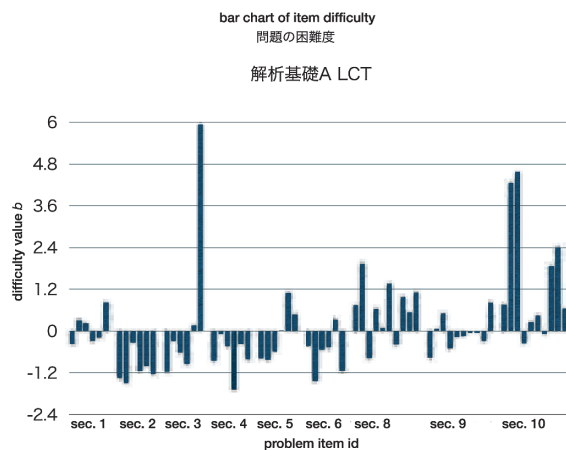
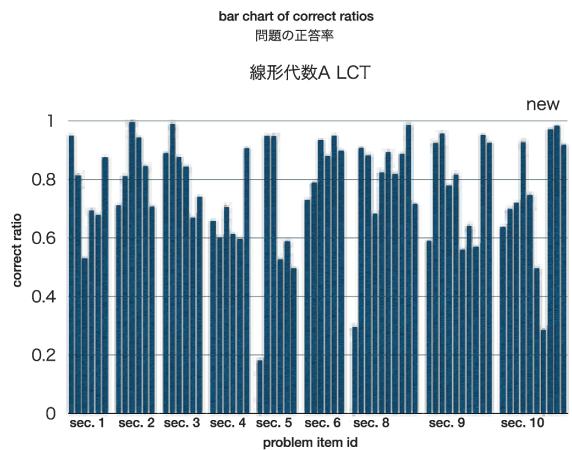


図25 解析基礎 A の LCT の IRT 困難度 b の棒グラフ



未回答を除く処理

図23 線形代数 A の LCT の正答率の棒グラフ
(未回答 (空欄) はすべて除外して処理)

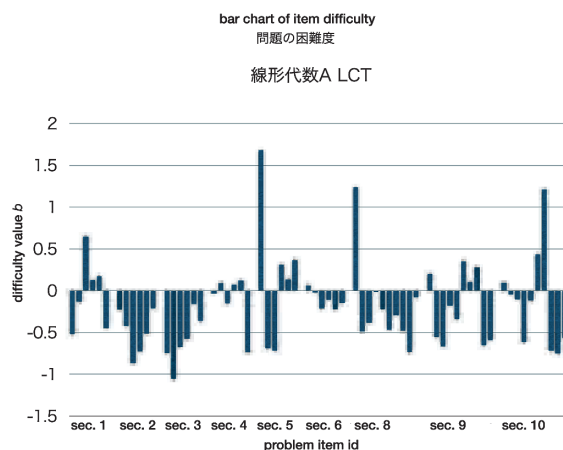
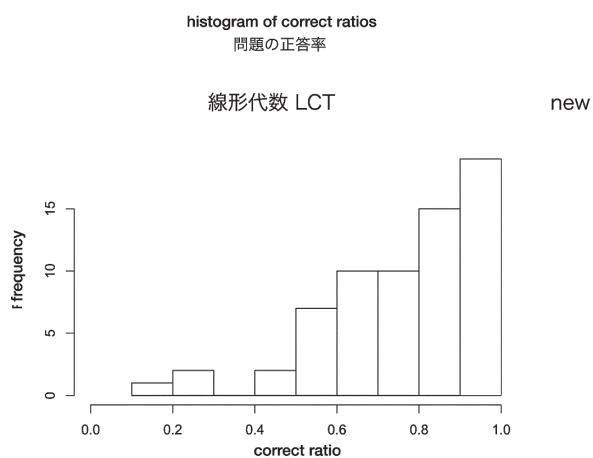


図26 線形代数 A の LCT の IRT 困難度 b の棒グラフ



未回答を除く処理

図24 線形代数 A の LCT の正答率のヒストグラム
(未回答 (空欄) はすべて除外して処理)

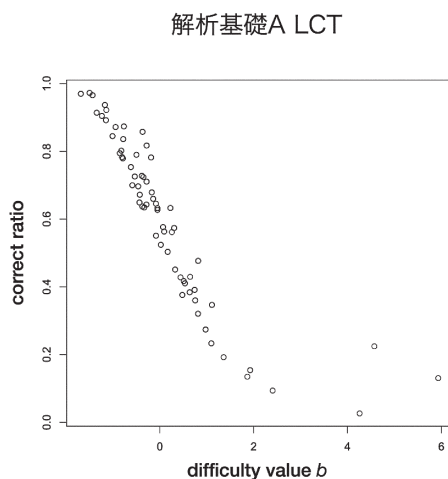


図27 解析基礎 A の LCT の IRT 困難度 b と正答率との関係

線形代数A LCT

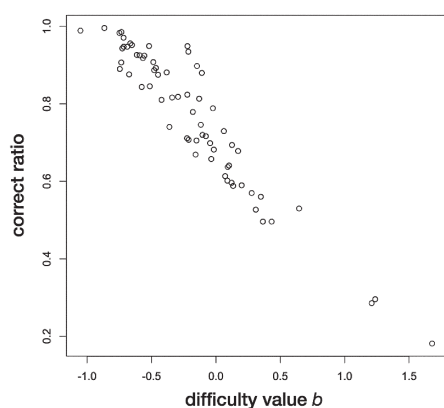


図 28 の LCT の IRT 困難度 b と正答率との関係

文 献

- 1) 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, July 10-14, 2016.
- 2) H. Hirose, M. Takatou, Y. Yamauchi, T. Taniguchi, T. Honda, F. Kubo, M. Imaoka, T. Koyama, Questions and Answers Database Construction for Adaptive Online IRT Testing Systems: Analysis Course and Linear Algebra Course, 5th International Conference on Learning Technologies and Learning Environments (LTLE2016), pp.433-438, July 10-14, 2016.
- 3) Y. Tokusada, H. Hirose, Evaluation of Abilities by Grouping for Small IRT Testing Systems, 5th International Conference on Learning Technologies and Learning Environments (LTLE2016), pp.445-449, July 10-14, 2016.
- 4) Hirose, Hideo; Aizawa, Yu, Automatically Growing Dually Adaptive Online IRT Testing System, IEEE International Conference on Teaching, Assessment, and Learning for Engineering 2014 (TALE 2014), 5C_5, pp.528-533, December 8-10, 2014.
- 5) Hirose, Hideo; Tokusada, Yoshiko; Noguchi, Kazuhisa, Dually Adaptive Online IRT Testing System with Application to High-School Mathematics Testing Case, IEEE International Conference on Teaching, Assessment, and Learning for Engineering 2014 (TALE 2014), 6B_1, pp.447-452, December 8-10, 2014.
- 6) Hideo Hirose and Yoshiko Tokusada, A Simulation Study to the Dually Adaptive Online IRT Testing System, IEEE International Conference on Teaching, Assessment, and Learning for Engineering 2014 (TALE 2014), 8E_3, pp.97-102, 2014.
- 7) 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 (TALE 2012), pp.8-12, August 20-23, 2012.
- 8) T. Sakumura, T. Kuwahata and H. Hirose, An Adaptive Online Ability Evaluation System Using the Item Response Theory, Education & e-Learning (EeL2011), pp.51-54, November 7-8, 2011.
- 9) H. Hirose and T. Sakumura, An Accurate Ability Evaluation Method for Every Student with Small Problem Items Using The Item Response Theory, Proceedings of the International Conference on Computer and Advanced TEchnology in Education (CATE 2010), pp.152-158, August 23-25 2010.

