

# 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

廣瀬 英雄<sup>†</sup>

(受付 2017 年 6 月 19 日；改訂 2018 年 4 月 14 日；採択 4 月 18 日)

## 要 旨

多様な学生を受け入れながら最終的には大学のディプロマポリシーに沿う学生にまで育成させる使命を負う大学が抱えている重要な課題の一つは、早期にドロップアウトリスクを持つ学生を発見し適切な手当を施し可能な限りそのリスクを低減させることである。このため、広島工業大学では新しく開始した基礎教育のフォローアッププログラムに、項目反応理論(IRT)を用いた評価法を取り入れたオンラインテストシステムを組み込み、2016年4月から運用することで対応してきた。対象学生は1年生全員で対象科目は数学(解析基礎と線形代数)である。開始から2年が経過した現在、システムは順調に動き、ラーニングアナリティクスが実施できるようなデータもそろいつつある。そこで本稿では、これを、大規模なテストシステムによって学習データを周到に蓄積するシステムの事例として紹介し、そこから得られた大量のデータを統計的に分析することにより、ドロップアウトリスクのある学生を早期に特定する可能性について述べる。これまで漠然と感じてきたリスク要因が、データを統計的に分析することで明確に示されるようになった意義は大きい。ここで調査した科目は数学であるが、他の基礎科目(例えば、統計や統計教育、あるいはSTEM教育のような、科学・技術・工学・数学の教育分野)にも容易に適用可能である。

キーワード：ラーニングアナリティクス，ドロップアウト，フォローアッププログラム，項目反応理論，習熟度確認テスト，アダプティブテスト。

## 1. はじめに

現在、日本の多くの大学では、多様な習熟度の学生を受け入れながら一定の水準にまで教育効果を上げるための新たな工夫が強く求められている。これは、大学のアドミッションポリシーに沿った学生を受け入れ、ディプロマポリシーに達するようなカリキュラム策定とその適切な運用が一層求められているからである(中央教育審議会, 2016)。そのような工夫がなければ、学生の多様性が増すにつれてドロップアウトする学生は増えると思われ、ドロップアウトの可能性を持つ学生を早期に発見し、適切な手当を施し、可能な限りドロップアウトのリスクを低減させることが一層重要になってくる(Waddington et al., 2016)。

これまで、このリスク低減のために、1)習熟度別クラス編成による教育体制、2)クラスでの適切な学生数、3)演習時間の確保、4)学習支援室の整備などの様々な対策がとられてきた。そして、それらの対策が効果的であったかどうかについて多く報告されている。しかし、効果を測

---

<sup>†</sup> 広島工業大学 環境学部：〒731-5193 広島市佐伯区三宅 2-1-1

るデータについては、いろいろな制限もあって、入学時と期末試験時、あるいは学年進行時などと時間間隔が長い事例や、クラス単位など限られた状況での特殊な場合の報告も多く、十分な質と量が確保されていなかった可能性もある。入学直後からのリスク早期発見のためには、データ収集の時間間隔をもっと短くして詳細な大規模データを周到に蓄積し、得られた大量のデータから内容の本質を統計的に分析する必要があると思われるし、また最近ではそのことが可能な時代に入ってきている。

そこで、本稿では、これまで行われてきたドロップアウトリスク低減策にあらたにセイフティーネット(ここではフォローアッププログラム(FP)と呼び、後述する)を加えるというだけでなく、本質的に質の異なる教育方法論への展開の可能性について述べる。それは、統計に限らず、数学、物理、英語など多くの科目での教育に、大規模なラーニングアナリティクスを実践することによって授業の教育効果を測り、またその結果を用いることで教育方法を改善する方法論を提案するという位置づけになる。大規模なデータを周到に蓄積するには組織的な理解や運用の仕組みも必要でありこれまで大学全体での取り組みの実例は少なかったように思われる。ここではその事例についても述べる。

次節では、ドロップアウトを減少させる仕組みの設計思想を述べ、それを実現させる具体的な方法論として毎授業時でのオンラインテストシステムによるデータ蓄積法の一例について述べる。続く3節では蓄積データを統計的に分析した結果の一例について述べ、これが早期のリスク発見につながる可能性があることを示し、最後に4節でまとめる。

## 2. フォローアッププログラムとオンラインテストシステム

### 2.1 フォローアッププログラムの基本姿勢

例えば、習熟度別クラス編成を行なうために入学直前／直後にプレースメントテストを行い、その結果を用いてリメディアルが必要と思われるクラスには特別なケアを施すということが行なわれている。クラス分けのしきい値には普通1回のテスト結果での得点が用いられる。多くの人に理解される方法であるが、テスト結果に確率的なゆらぎがあることにあまり配慮が届いていないようにも思われる。問題なのは、いったん特別なクラスに入ると、そのセメスターあるいはクォーターではクラスの入れ替えを行わず固定されたメンバーのままそのクラスでの授業を受けることになることである。

ところが、たまたま下位クラスに配置された学生でも上位クラスでの受講が十分可能なこともある。例えば、期末にすべてのクラスで同じ問題のテストを受けると、下位クラスの一部の学生の成績が上がったという報告を受けることがある。しかし、均質なクラスを成績によって上位と下位の2つに分けてもう一度クラス毎にテストすると下位クラスの平均点は一般に上昇することを考えると、下位クラスでの努力の結果そうなったのか単に確率的な変動でそうなったのかはわからない。

そこで、上位クラスは設けても下位クラスは基本的に作らずに均質な集団のまま授業を進ませながら、全クラスで毎回の授業での理解度の確認テストを行い、テストの結果理解度が不十分と考えられる学生については1週間以内に補習を行うことを繰り返すことで学生のドロップアウトのリスクを軽減させることを考えた。これがFPへの基本的な姿勢である。

### 2.2 オンラインテストシステム導入の動機

例えば数学系の授業では、丁寧な板書と演習との組み合わせによって理解度を深めることが伝統的に行われてきたが、インターネットやスマートフォンを駆使できる現代の若者には、黒板を書き写しながらの理解だけでなく、コンピュータを用いながらわかりやすい可視化画像や

ゲーム感覚のテストなどを組み込んだ授業の方が興味を惹きやすい。例えば、10分でできる5問程度の数学の問題をアダプティブに解いてもらうオンラインシステムを高校生に対して試したところ熱心に取り組んでいた(桑幡 他, 2012)。そればかりではない。高校生の両親や高校の教員にも興味を持って受け入れられたことがある。このことをヒントに、大学の数学系の教科書の付録に Web アシスト演習を組み込んだシリーズが出され(廣瀬・藤野, 2015, 桂 他, 2015, 2017)、それを使いながらの授業も進められている(Hirose, 2016b)。スマートフォンの操作に慣れた学生にはオンラインでのテストへの抵抗感もない。

オンラインテストの魅力は、入力されたデータ解析が迅速にできることに加え、データが蓄積されることにある。規模の大きいデータベースからは、経時変化の様相や分類されたグループ間での相違など、教員が単独にクラス単位で行っていたときのデータによるよりもはるかに多くの情報が得られる。定期的なオンラインテストの結果はラーニングアナリティクスへの重要な情報源になる。

そこで、これまで限られて利用されていたオンラインシステムを、全学的にさらに大規模に発展させたものを大学での教育に本格的に組み込むことができれば、詳細な教育データが大量に蓄積され、多様性に富む学生の教育改善に貢献できるのではないかと考えて、全学で取り組むオンラインテストシステムを構築した。これは、多様性が増加するときの学生の習熟度に合わせられる指導法を担保するため、1) 毎時間こまめに学生の理解度を確認し、2) 授業への補足が必要であればそれを促し、3) 補足授業を実施し、またその理解度の確認も行い、さらに、4) 授業の事前事後学習への体制を作っていくという学習支援プログラムである FP に対応させるための要となっている。ここでは、大学の初年次生すべてに適用されている必修の数学、物理、英語教育のうち、特に数学科目(解析基礎と線形代数)について、新しく構築した FP システムの概要を紹介する。

### 2.3 フォローアッププログラムシステムの構成

図 1 に、大学の初年次対象に全学で取り組んでいる FP システムの全体像を示す。図には、大きく、授業の理解度を確認する LCT (learning check testing)、オンライン演習を主な目的とした CWT (collaborative work testing)、補足授業の理解度を確認する FPT (follow-up program testing) によって運用されていることが示されている。このように、FP システムは主に 3 つのオンラインテストによって構成されている(Hirose, 2016a)。補足授業のクラスは LCT で理解度が不十分とみなされた学生に対して毎週クラスのメンバーが変わる動的なクラスになっており、これを FP クラス (FPC) と呼ぶ。

LCT は、授業の内容の理解度を確認するために 10 分程度で行なうオンライン小テストである。全員に同じ問題が出題されるが出題順はランダムである。採点は全学生からのテスト回答が大多数出揃った時点で一斉に行なわれる。問題と解答例は非公開としている。

CWT の問題は、演習の効果が上がるように、学生の習熟度に応じた問題が多量の問題プールの中から自動的に選ばれるアダプティブシステム(桑幡 他, 2012)によって出題される。採点と評価は回答直後に行われ、成績と解答例をその場で確認することができる。CWT の性格から、CWT は FP クラスを離れた事前事後学習にも使うことができる。

FPT は、当該学生の学習履歴(LCT や CWT によって得られた習熟度)に応じた問題が適切に選ばれるアダプティブシステムによって出題される。FP クラスにはさまざまな習熟度の学生が集まってくるため、FPT の成績評価は当該学生の習熟度が向上したかどうかで判定される。問題と解答例は非公開としている。

学生は自分のパソコンを大学に持参し正規の授業中や補足授業中に WiFi を介してオンラインテストに臨むことができるが、それだけでなく、CWT では休み時間など自由な時間帯でも、

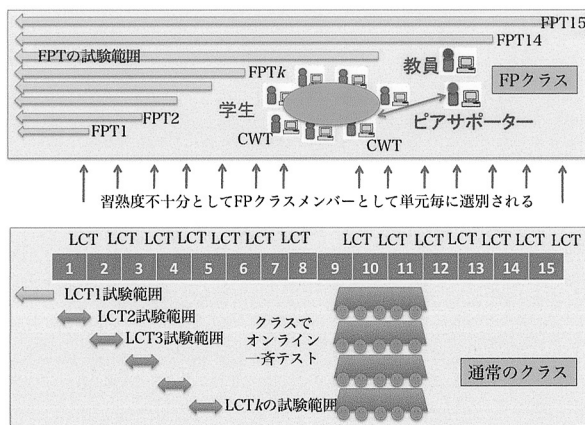


図 1. FP システムの全体像. LCT(learning check testing), CWT(collaborative work testing), FPT(follow-up program testing).

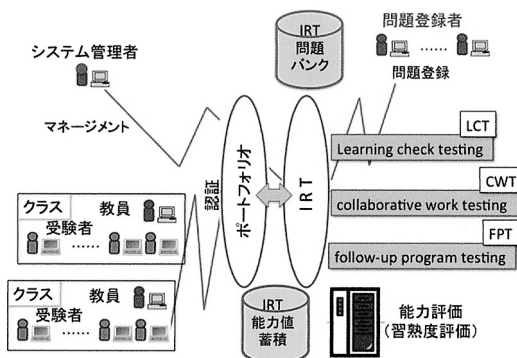


図 2. FP システムのシステム構成.

また自宅や通学途中でも自由にテストシステムにアクセスできる体制になっている。

図中に示されている, LCT, CWT, FPT, あるいはピアサポーターなどの詳細については付録で説明する。

また, FP システムのシステム構成を図 2 に示す. 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, 授業を担当している科目の教員, 学科で学生の支援を一括して執り行う学生支援アドバイザー(学科単位), 学科内で割り振られた学生の支援を行うチューター(学科内教員), 大学全体での学生の教育支援を行う学習支援センター(科目あたり数名の専門教員)に迅速に知らされるようになっているため, FP システムは科目の教員だけでなく学習支援に関係している人すべてに連携がはかられるような構造になっている。

FP で根幹になっているのは, LCT, CWT, FPT のテストシステムであるが, その他にも, テストに使う問題の登録システムや全体を運用するマネージメントシステムが付与されている. e-learning は作ってしまったらその後使われなくなるという評判をよく聞く. その一因はシステムの柔軟性の欠如にあると考える. 問題や解答の修正, 追加登録, 学生のレベルに

あった問題をきめ細かく選定することで、システムが生きてくる。そこで、システムには、問題作成担当教員なら誰でも自由に時間と場所の制限を受けずに問題登録を行うことができる機能が備えられている。この問題の登録システムの詳細については付録で説明する。また、問題の難易度が受験生とマッチするように問題の入れ替えを行なうなどのマネジメント機能も充実させている。問題の難易度分析や受験生の習熟度評価には、単元単位(授業単位毎と等しい)で項目反応理論(IRT)を採用している。

### 3. ラーニングアナリティクス

授業設計はクラスを担当する教員に委ねられていることが多いが、そこには自由度の大きい設計ができるというメリットがある。ところが、大人数を抱える学科になると一人の教員だけである科目の授業を行うことには教育効果の面で疑問があるため、複数の教員によるクラス設計が行われることがある。この場合、公平性を保つため、授業内容や評価方法についてある程度の共通化を図らなければならない。毎回のクラスにおける小テストなどの評価法についてもある程度の共通性があることが好ましいが、従来のような紙を使った方法では、その手間を考えると実現性に乏しい。一方、学生の普段の習熟度把握については指導上きめ細かい情報が得られることが好ましい。

そこで、これまでの各クラスの教員独自の方法に委ねてはなかなか得られなかったきめ細かい情報も得られるように、小テスト結果や補助的な指導結果についてもコンピュータを用いて着実に蓄積することで、教員の負担を軽減しながらも学生指導の適切な判断ができるようなラーニングシステムを構築することとした。本論文では上で説明したFPの中で実現できるようなオンラインテストを中心としたFPシステムを提案している。このようなシステムを使えば、きめ細かい情報が得られるため、それを統計的に分析することによって効果的で効率的な教育方法を探ることができる可能性がある。

#### 3.1 LCT アナリティクス

LCTは1年生の必修基礎科目を対象にしたオンラインテストで、数学の科目については、解析基礎と線形代数の2科目が対象になっている。少しの変則性はあるが、基本的には毎週の授業や演習時間を使って5問から10問を10分程度で解答させる。LCTの詳細については付録に示しているなのでそちらを参照されたい。ここでは、LCTから得られた結果について一つの分析を行なう。

図3に、一例として2016年度での解析基礎A(前期)の応答マトリクスを示す。図の右端に説明しているように行を学生idに列を問題idにとっており、正答と誤答を赤と緑で区別している。空欄は未回答か未受験を表している。全学生で1000人以上と学生数が多いので、ここでは約250人ずつに分けて4つのマトリクスで表している。授業が進むにつれてLCTの正答率は低くなっていることがわかる。

ここでは、この応答マトリクスを基礎データとして、期末試験で合格する学生と不合格となる学生を区別することができるかどうかを議論する。

まず、期末試験に合格した学生と不合格になった学生とを比較した応答マトリクスの一部を図4に示した。期末試験不合格の学生は合格学生よりもLCTの正答率が全体的に低いことが一目でわかる。更によく見ると、期末試験不合格の学生は授業が進むにつれてLCTの不合格の割合が大きくなっているようにも見える。

もう少し詳しく、期末試験に成功した学生と失敗した学生との間でLCTの成績のトレンドにどのような違いがあるか調べてみた。図5は、期末試験に成功した学生 $i_s$ のLCTの回数

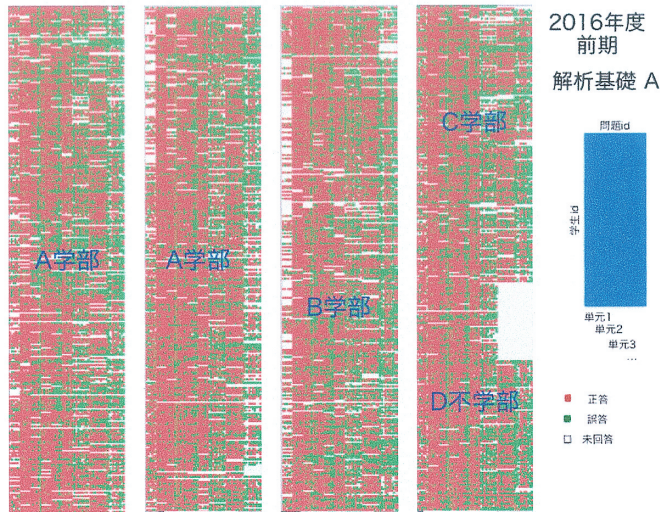


図 3. 解析基礎 A (2016 年度前期) の応答マトリクス.

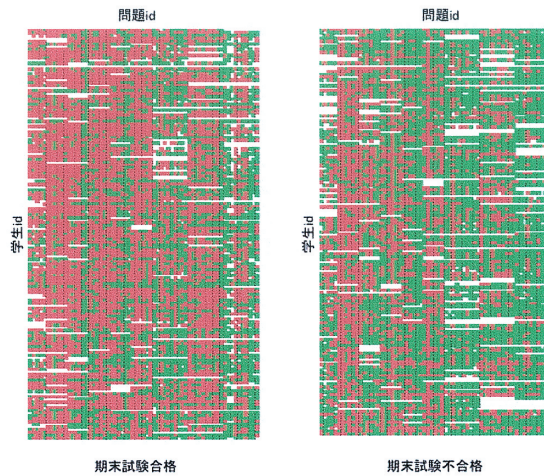


図 4. 期末試験合格学生と不合格学生の応答マトリクスの比較(一部分).

$k$  ごとでの正答率  $r_{i_s,k}$  の平均  $r_{s,k} = (\sum_{i_s=1}^{n_s} r_{i_s,k})/n_s$  と失敗した学生  $i_f$  のそれ  $r_{i_f,k}$  の平均  $r_{f,k} = (\sum_{i_f=1}^{n_f} r_{i_f,k})/n_f$  を比較したものである. 2016 年度前期と 2017 年度前期の解析基礎 A のものを示している. どちらの年度でも, 授業が進むにつれて, LCT の正答率の平均は期末試験合格グループと不合格グループともに低くなっていることがわかる. また, LCT の成績の両者の間には開きがあることもわかる. しかし, 両者のトレンドの間にどのような違いがあるかどうかはこれだけではわかりにくい.

そこで, 期末試験に成功した学生の正答率の平均  $r_{s,k}$  に対する失敗した学生の正答率の平均  $r_{f,k}$  の比  $r_k = r_{f,k}/r_{s,k}$  (Hirose, et al., 2017) を, 2016 年度前期と 2017 年度前期の解析基礎 A の各回の LCT について調べてみた. これを示す図 6 から, 比はどちらの年度でも下降傾向を示しているように見える. 2016 年度前期での回帰直線の傾きを求めるとその値は  $-0.0277$  で

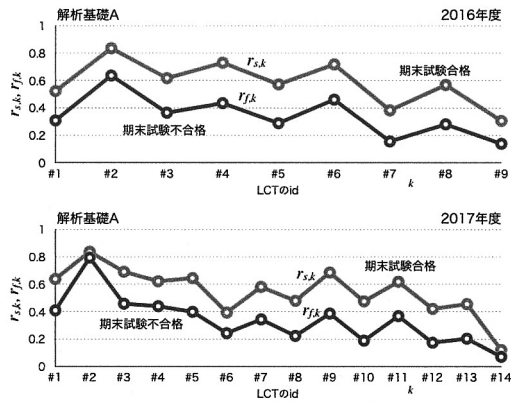


図 5. 期末試験合格学生の正答率の平均  $r_{s,k}$  と不合格学生の正答率の平均  $r_{f,k}$  の比較.

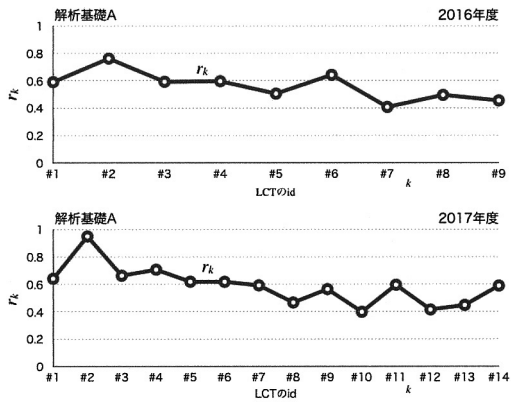


図 6. 合格学生の正答率の平均に対する不合格学生の正答率の平均の比  $r_k = r_{f,k}/r_{s,k}$ .

表 1. 正答率の平均の比  $r_k$  の回帰直線の傾きの推定値と標準偏差.

	推定値	標準偏差	推定値 + 2 × 標準偏差
解析基礎 A (2016 年度前期)	-0.0277	0.0010	-0.0257 < 0
線形代数 A (2016 年度前期)	-0.0062	0.0012	-0.0038 < 0
解析基礎 B (2016 年度後期)	-0.0273	0.0010	-0.0253 < 0
線形代数 B (2016 年度後期)	-0.0061	0.0013	-0.0035 < 0
解析基礎 A (2017 年度前期)	-0.0233	0.0007	-0.0219 < 0
線形代数 A (2017 年度前期)	-0.0152	0.0008	-0.0136 < 0

あった。また、ブートストラップ法によってその標準偏差を求めると 0.0010 であり、下降傾向があることが示された。つまり、授業が進むにつれて、期末試験に失敗した学生はそうでない学生よりも習熟度(能力値)の下降程度が大きくなっていることがわかる。

他のケースについても同様な傾向があるかどうかを計算した結果が表 1 である。すべての場合で、期末試験に失敗した学生はそうでない学生よりも習熟度(能力値)の下降程度が授業が進むにつれて大きくなっていることが確認できる。

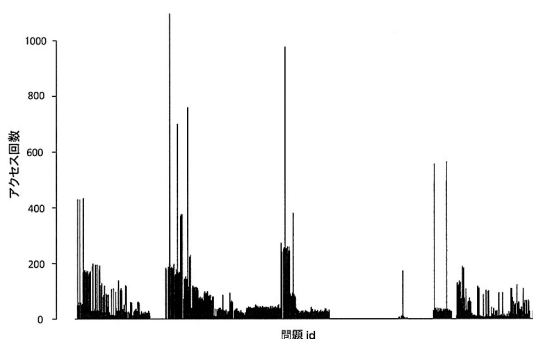


図 7. CWT の各問題へのアクセス回数の棒グラフ(問題総数は 2016 年 9 月時点で 3727 問である).

表 2. 各問題へのアクセス回数の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位数.

最大値	最小値	中央値	平均値	第 1 四分位数	第 3 四分位数
1096	0	19	38.16	0	38

このことから, できるだけ早期に期末試験に失敗する可能性のある学生を抽出し特別な手当を受けさせることには意味があることが再確認できる. この傾向は定性的にはよく知られているが, 実際到大規模なデータを収集して統計的に分析した例は少ないように思われる. このようにラーニングアナリティクスには大規模データそのものだけでなく, データを収集する方法まで考えることが重要であることがわかる.

### 3.2 CWT アナリティクス

FP クラスには LCT の成績によって習熟度不十分と判断された学生が毎週選別されるので, FP クラスの対象者は毎週変わる変則的なクラス構成になっている. そこでは共通問題でのテストを行っても相対的な評価結果は得られにくい. そこで, 個々の問題に注目した分析と, 個々の学生に注目した分析を行うこととした. FP システムに登録された問題数は, LCT, FPT, CWT あわせると 2016 年 9 月時点で 3727 問になる. LCT 用の問題数は 200 問以下であり FPT のそれも同じような数であるから, 多くの登録問題は CWT に使われていることになる. そこでは, 全くアクセスされない問題もある一方で多数回アクセスされる問題も出てくるし, 学生によっても CWT アクセスの頻度が異なってくる.

一例として, 2016 年度の 4 月初めから 7 月末までの CWT へのアクセス結果をしてみる. その期間, 問題への全アクセス数は約 50000 回であった. 1 回の CWT には 5 問が 1 セットとして出題されることから CWT へのアクセス回数は約 10000 回であり, 1 年生の数をおおよそ 1000 人と考えると, 一人あたりの CWT へのチャレンジ回数は 10 回となる. 実際には全く受験していない学生も 400 人以上いるので, チャレンジした学生の場合のアクセス回数はさらに増えこの 2 倍程度になると思われる.

図 7 に, 各問題にアクセスされた回数の棒グラフを示す. 横軸は問題の id を, 縦軸はアクセス回数を表す. 図にはすべての FP システムの登録問題へのアクセス回数を示しているが, LCT と FPT 問題は CWT には出題されないのので, LCT 問題をはじめとして CWT には全くアクセスされない問題も含まれている. 各問題へのアクセス回数の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位数の値を表 2 に示す. アクセス回数が 200 を超える問題もときに



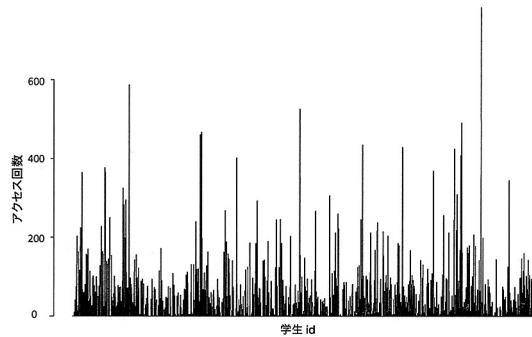


図 8. 各学生が CWT にアクセスした回数の棒グラフ (学生総数は 1160 である)。

表 3. 学生が問題にアクセスした回数の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位数.

最大値	最小値	中央値	平均値	第 1 四分位数	第 3 四分位数
784	0	20	45.82	0	65

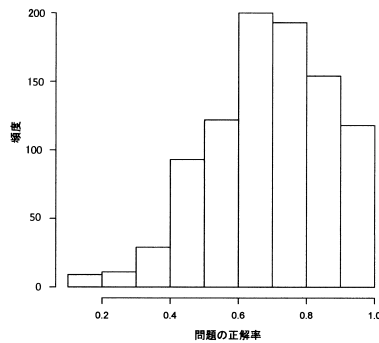


図 9. 正解率のヒストグラム.

あり半数以上の問題には 20 回以上のアクセスがあるので, CWT の問題アクセス頻度としてはそれほど不適切ではない。

次に, 各学生がどの程度問題を解こうとしているのかということについて述べる。図 8 は, 横軸は学生の id を, 縦軸はアクセス回数を表す。多くの学生がアクセスしているように見えるが, 何人かは一度もアクセスしていない。学生が問題にアクセスした回数の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位数の値を表 3 に示す。800 問近く解こうとしている学生や 500 回以上の挑戦者も 10 人以上いて, 自主的に問題を解いて力をつけたいという学生の姿が読み取れる。一方, アクセス回数が 40 問以下の学生も多く, 熱心な学生とそうでない学生の差は大きいように思われる。

CWT の問題のレベルと学生の能力がマッチしているときに能力推定のための情報を最大に引き出せるので, 問題のレベルはそのように設定されていることが好ましい。そこで, 学生が取り組んだ問題数に対して正答した問題数の比(ここでは正解率と呼ぶ)を調べた。図 9 に, 正解率のヒストグラムを示す。正解率の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位

表 4. 正解率の最大値, 最小値, 中央値, 平均値, および第 1, 3 四分位数.

最大値	最小値	中央値	平均値	第 1 四分位数	第 3 四分位数
1	0.11	0.70	0.70	0.58	0.83

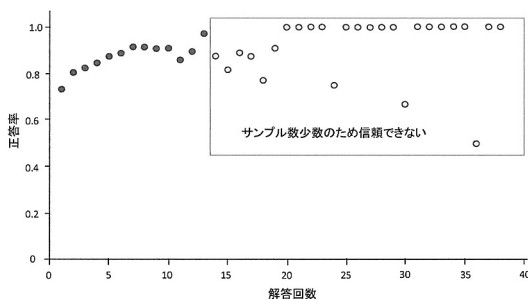


図 10. 解答回数と正答率の関係.

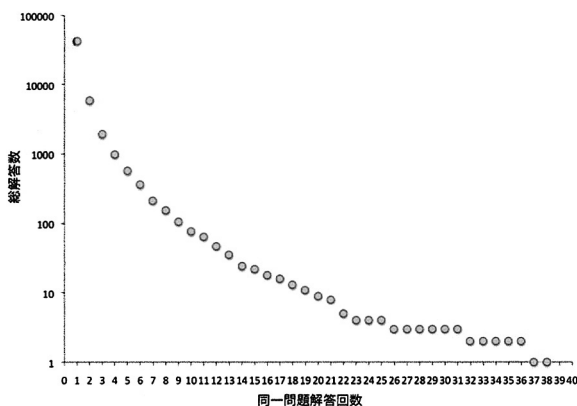


図 11. 同一問題解答回数と総解答の関係.

数の値を表 4 に示す. CWT の問題は受験生にとってやや易しい方に設定されていることがわかる.

同じ学生が同じ問題を真面目に複数回解いたとき, 問題に正解する割合は高くなると考えられる. 図 10 は, 横軸に何回目の解答にあたるかを示し, 縦軸は正答率を示す. 例えば, 1 回目に挑戦した場合の正答率は 0.73 である. このとき, 1 問挑戦した解答総数を調べておく必要があるため, それを図 11 に示した. 1 回の回答ケースは(1 回解答だけでなく 2 回以上解答した場合の学生も含む), 約 40000 ケースある(約 50000 件のうち大半がこれにあたる). このときの正答率が 0.73 であることは, 全体としても同様な正答率になっていることを示している.

図 10 を見ると, 同じ問題への解答を重ねるごとに正答率が上がっていることが分かる. 最初 0.73 だったものが 5 回目で 0.87 に, 10 回目で 0.91 に上がっている. この結果は, 習熟度が上がったのか解答を暗記していただけなのかは分からないが, 習熟度向上の可能性は否定はできない.

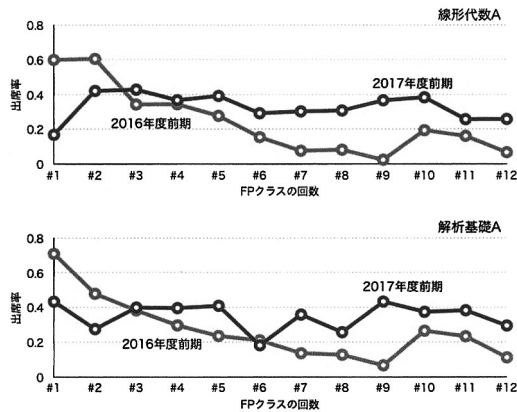


図 12. FP クラスへの参加率(2016 年度前期, 2017 年度前期)。

### 3.3 FPC アナリティクス

FP クラスでは FPT を行っているが、問題の困難度を個々の能力値に合わせるようなアダプティブな問題設定にしているため、個々の学生に対して全く異なった試験問題が設定される。それらを同じ土俵で取り扱うことはできない。そこで、ここでは FP クラスの参加率についてのみ述べる。

一例として図 12 に、2016 年度前期および 2017 年度前期での FP クラスへの参加率を示す。2016 年度は回を重ねるごとに参加率は段々小さくなる傾向にある。この原因は、FP クラスが 5 限目(16 時 45 分-18 時 15 分)と遅いため、学生の夕方の予定とぶつかること、FP クラスへの参加は単位取得とは無関係なのでただのサービスと受け止めて自分のものにする動機付けが薄いこと、まわりが行かないなら自分もという大衆心理などが働いていたものと思われる。2017 年度はこの取り組みを丁寧に説明して参加者が減らないように工夫した。その効果があって、2017 年度は参加率の下降傾向は治まっている。しかしながら、強いインセンティブがないためか全体的に参加率は低い。参加率が低いということは、FP クラスからのアナリティクス情報の信頼度が低くなると考えられるため、教育効果を統計的に正確に見るためにもある程度の出席率が確保されていることが好ましいと思われる。

### 3.4 ドロップアウト対象の特定

これまで、FP システムを支えている LCT, CWT, FPT それぞれのアナリティクスについて述べ、ドロップアウトの可能性のある学生は早くから LCT の成績が落ちていることを述べた。しかし、その対象を明確に個々に特定できるまでには至ってない。ここでは、更に別な情報も加味してドロップアウトの可能性のある学生を早期に特定できるかどうかについて議論する。加える情報とは、入学時のクラス分けのためのプレースメントテストの結果、毎授業の出席状況、FP クラスへの出席状況などである。

図 13 に、一例として 2017 年度前期での線形代数 A の LCT の各回での能力値の平均とプレースメントの成績との関係を示す。図には、期末試験の合格者(60 点以上)、D 不合格者(40 点未満)、E 不合格者(40 点以上 60 点未満)の 3 つの集団を色分けして示した。LCT アナリティクスでは、単元毎の能力値のトレンドからドロップアウトの兆候は示唆されることは示されたが、図を見ると、LCT の各回での能力値の平均をとると期末試験の合格者と不合格者の区別はつけにくくなっている。また、プレースメントテスト成績からも期末試験の合否を識別できる

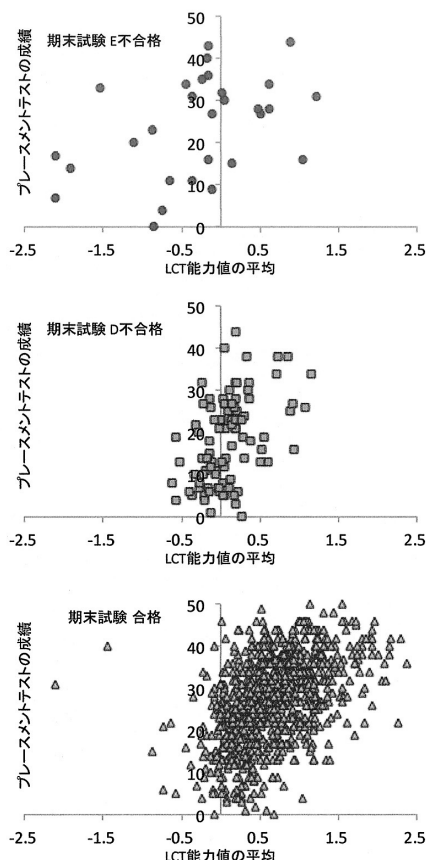


図 13. LCT 能力値の平均とプレースメントの成績との関係。

しきい値は見つけにくい。

図 14 には、授業への出席状況と LCT の受験状況を、期末試験の合格者(青枠で囲んだ)、D 不合格者(ピンク枠で囲んだ)、E 不合格者(赤枠で囲んだ)の 3 つの集団で区別して示してみた。これは、線形代数 A の 2017 年度前期の前半部分を一例として取り出したものである。出席状況は、電子カードによる情報で、1:出席, 2:遅刻, 3:在室(大幅遅刻), 4:届け出あり(出席とみなす), 5:欠席, 6:未判定(欠席扱い), 受験状況は、1:合格, 2:不合格(受験して), 3:不合格(受験していない), 4:免除(合格扱い), 5:未受講(合格扱い)である。この 2 つの数値から 1 桁目に出席状況を 2 桁目に LCT 状況とした(10 進数の)2 桁の数値を作り色分けしている。この図から、期末試験 E 対象者には授業を欠席している学生が多いことがわかる。また、D 対象者では LCT 不合格者数が多いことがわかる。ドロップアウト対象を特定するには授業への出席状況をまず見るということとはよく聞くことであるが、データによって明快に示されている。

プレースメントテストの結果、LCT の成績、毎授業の出席状況に加え、LCT の合否、FP クラスへの出席要請、FP クラスへの出席などのデータを説明変数に、期末試験での合否を目的変数にしたときの回帰分析を行なってみた。目的変数には、合格で 0、D 評価に 1、E 評価に 2 を与えている。その結果、FP クラスへの出席は不要(LCT 合格と同じ意味)という情報が有意

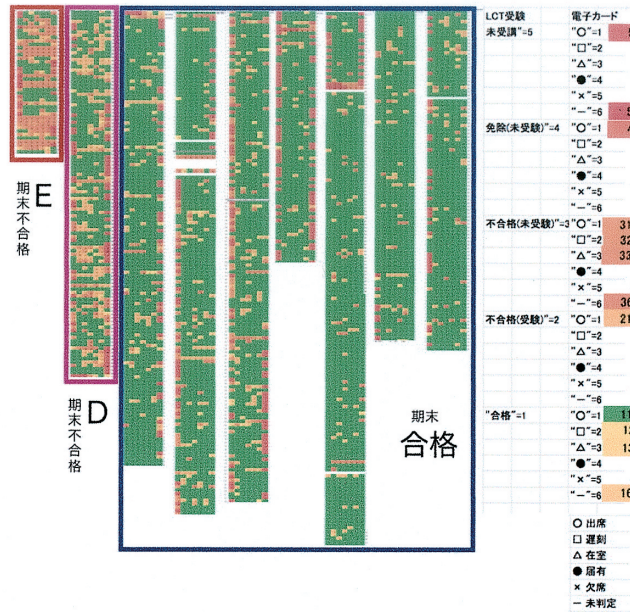


図 14. 授業への出席状況と LCT の受験状況。

であることが分かった。つまり、LCT に合格している回数が多い学生には目的変数の値 0 がよく対応し、そうでない学生には目的変数の値 1 以上がよく対応しているということである。

電子カードを用いた出席情報による出席数は実際に出席している学生数と合わないことがあり、極端な場合にはそれらの差は履修者数の数 10% にまでなることがあるが、LCT の受験履歴は着席している学生数と同じであるため正確な出席状況が把握できる。LCT を課す授業の場合、電子カードを用いた出席情報を FP を始める前と比較してみると出席状況が 90% から 95% へとよくなっていることが分かった。LCT 受験へのインセンティブによるものと思われる。なお、授業欠席者は期末試験 E 対象者が大半である。

以上の考察から、ドロップアウト対象者を早期に特定するには、まず授業への出席情報を正確に把握し、次に LCT の合否状況を把握することが重要であることがわかった。

#### 4. おわりに

多様な学生の中から早期にドロップアウトリスクを持つ学生を発見し、適切な手当てを施し可能な限りリスクを低減させるために新しく構築したテストシステムについて述べ、そのシステムから得られた大規模なデータからラーニングアナリティクスを試みた結果について述べた。その結果、期末試験に失敗する学生は期末試験に成功する学生よりも毎週の習熟度確認のテストでの正答率が授業が進むにつれて低くなっていること、また、授業への出席状況が悪いことが示された。

大学 1 年生の数学科目に適用したこのシステム(ここではフォローアッププログラムシステムと呼んだ)は、ラーニングシステムに搭載されるコンテンツを、物理、英語などの他の科目に変更することが容易であるため適用性や汎用性に富むものである。教育指針を統計的に与えることができる新しいシステムとして活用できるものとする。

大規模なオンラインテストによって得られたラーニングデータは今後も着実に蓄積されていく。そのデータを分析することによって得られるラーニングアナリティクスは、これまで見えなかった教育への新しい方法論を与えてくれる可能性があり、少なくともこれまで漠然と感じてはいても強く主張できなかった事柄をデータをもとにして統計的に主張できるものとする。ラーニングアナリティクス研究はこれからである。

今後は、テストシステムを演習に限定せず中間試験などの評価試験に適用できるようにオンラインシステムの信頼性を確保すること、一大学だけでなく大学間コンソーシアムなどによる協力体制を敷きながらテストに使用される良問を充実させることで問題データベースの利用度を更に高めることなどが課題である。

広島工業大学の数学グループには、問題登録、LCT実施など多大な協力をいただいた。謝意を表す。なお、本研究の一部はJSPS科研費17H01842の助成を受けたものである。

## 付 録

ここでは、LCT(Learning Check Testing)、CWT(collaborative work testing)、FPT(follow-up program testing)それぞれの設計の考え方と運用法について簡単に述べる。LCT、CWT、FPTでは、すべて多肢選択方式のテスト形式としている。ただし、1問につき複数の回答欄を設けすべてに合致した場合を正答とするようにしているため、ランダム選択による見かけ上の正答率は小さく抑えられるようになっている。従って、評価法に用いるIRTでのロジスティック分布関数には識別力と困難度の2パラメータを持つ分布関数を用いた。

### 付録 A. LCT(Learning Check Testing)

LCTは、通常授業の内容の理解度を確認するために10分程度で行なう問題数5問から10問のオンライン小テストである。直前に受けた授業の内容を確認する意味で当該の単元内容から出題される。従って、全クラスで同じ単元内容の授業を展開するスケジュールが学期の始めに与えられ、授業の進度もスケジュールどおりに進められることになる。もちろん、教員の都合等により多少のずれは容認できるようになっている。クラス数は、1000人程度の学生数に約20人程度の教員で対応するため50-60人程度のクラスが20クラス程度できることになる。

LCTでは、理解度確認の意味から全学生に同一問題を与えている。ただし、隣の解答を見ることを攪乱する意味から出題順は学生によって異なる(ランダム出題)。1問1問順に解答するようになっているが、解答が困難な場合にはスキップして後で戻ることができる。1問解答するごとに応答はシステムのデータベースに格納されていき途中のシステム不具合を回避しながら、また書き直しなどにより変更があれば最新のものが蓄えられていくようになっている。テストを始めるときには教員が始まりの指示を出し、その後学生が個別にテストに臨むようにして、テストの始まりから終わりまでの時間が全員同じになるように設定されている。

LCTの問題は、後に示す問題登録法(登録システム)によって、事前に適切な問題を各教員が寄せ合い、会議を経て選別するようにしている。問題の選定には正答率が50%程度になるように設定することが評価精度の面からは好ましいが、多様な学生に対応できるようにするため、すべて同一水準の問題にはせず、易しい問題から難しい問題を均等に配置する方法をとっている(Hirose and Sakumura, 2010)。こうすることで、正答できる問題数が学生の習熟度によって異なっても習熟度の評価の精度はあまり落ちないようにしている。LCTの問題や解答は公開されていない。

学生や教員に伝える能力評価には、全体の問題数に対して学生が解いた問題数の割合から計算する正答率ではなく、各問題の困難度も評価結果に加えることができるIRTを用いている。

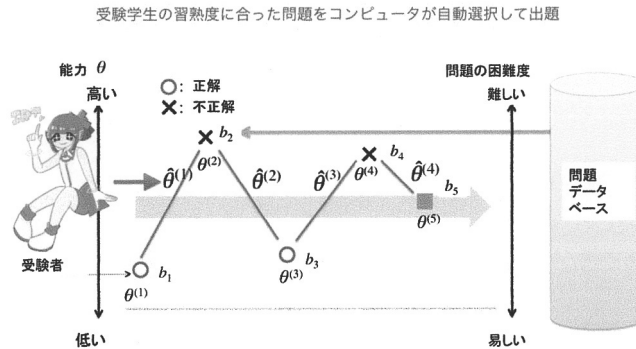


図 15. アダプティブシステムにおける自動問題選定.  $\theta^{(i)}$  はシステム側で設定した値を,  $\hat{\theta}^{(i)}$  は  $i$  番目の問題を解いた後で推定された値を示している. 従って  $\hat{\theta}^{(i)}$  は  $\theta^{(i-1)}$  と同じものである.

このように少ない問題でも評価精度ができるだけ高くなるようにしている.

LCTで成績が一定水準に達しない学生には、習熟度を補うための補習クラス(FPクラス)に出席するよう、FPクラス開講前にメールで連絡し、毎授業で着実に授業をフォローできるようにしている。対象者の選別は本来は絶対評価が好ましいが、FPクラスで収容できる物理的な学生数の制限およびFPの担当教員が指導できる学生数を考えて、相対的に全学生の10%以内を目指すようにしている。IRTの計算には全学生のLCT結果が揃っていることが望ましいので、LCTの採点は各週の一定時期を決めて全学生対象に一斉に行っている。

#### 付録 B. CWT(collaborative work testing)

CWTは、FPクラスで行なうオンライン演習を主な目的としているが、それに限らず、自宅や通学途中でも自由にシステムにアクセスすることによって通常授業の事前事後学習に備えられるようにしている。CWTの問題は、演習の効果が上がるように、学生の習熟度に応じた問題が問題プールの中から自動的に選ばれるようなアダプティブシステムとしている。このため、CWTにもアダプティブシステムを適用しやすいIRTを用いることにした。あらかじめデータベースに格納されている学生の習熟度に合わせて問題が最初に出され、その解答の成否によって次の問題が選定されながら、合計5問の問題が提示される(Hirose, 2016a)。全問回答後に、あるいは1問1問解答しながらの途中でも、解答の内容と正答および解説が確認できる。CWTはアダプティブシステムであるため、LCTや普通答案用紙を使ったテストのように、見直したり前後に行き来したりすることは許されていない。

付録図 15 に今回採用しているアダプティブシステムにおける自動問題選定法を示す。図では、受験者の習熟度がデータベースに記録されていれば最初にそのレベルの問題を、なければランダムに問題を選び、解答が正解であれば正解によって得られた情報と事前に得られている情報から受験者の習熟度をベイズ推定し、そのレベルの問題(回答した問題より難しくなる)を次に与え、正解でなければ逆により易しい問題を提示する、というようなアルゴリズムによって問題が出題される。図で  $\hat{\theta}^{(i)}$  は推定過程の受験者の習熟度の推定値を表している。能力値の推定精度を保つには出題数は10問以上が好ましいが(Sakumura and Hirose, 2017, Tokusada and Hirose, 2016)、10分という時間制約があれば5問から7問が現実的な出題数と考えられる。

FPクラスでは、FPC受講の学生5人に対し1人の上級生(サポーター)を配置するピアサ

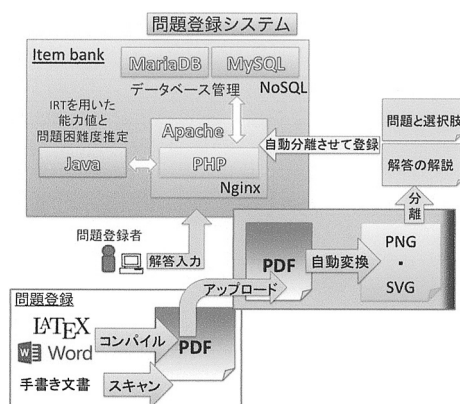


図 16. 問題登録の仕組み.

ポーターの制度を設けている。100 人が一堂に集まる会場で、5-6 人が一つのテーブルに集まって議論しながら演習を進められるような机の配置を作り、学生はパソコンに向かいながらコンピュータが与えてくれる自分に最適な問題を解いていくというシステムの形としている。FP クラス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クラスを進行させている。このシステムは、ピアサポーター自身の成長を促すシステムにもなっている。このように、学生、上級生、教員が協同で作業を行なうという意味で、CWT(collaborative work testing) という名称をつけている。

#### 付録 C. FPT(follow-up program testing)

FP クラスにはさまざまな習熟度の学生が集まってくるため、同じ問題を一齐に課しても学生の能力評価を確認するには正確さに欠けると思われる。そこで、基本的には当該学生の学習履歴(LCT や CWT による IRT 評価での能力値(習熟度))に応じた問題が適切に出されるように、CWT と同様にアダプティブなテストとしている。出題は、過去に LCT を受験した全単元の内容からシステムが個別的な出題を適切に選ぶようにしている。したがって、評価結果は他の学生との相対的な位置づけにはなっておらず、あくまで学生個人が FP クラス受講によって伸びたかどうかを見ている。

#### 付録 D. FP システムへの問題登録法

本論文で提案しているシステムではテストングが主体ではあるが、問題登録(Hirose, et al., 2016)も、問題データベースが柔軟性を持つという意味で、システムを支える大きな柱となっている。問題登録は、システムに精通した専任の担当者に委ねられるのではなく、すべての教員が自ら簡単に登録、修正、削除ができるようになっている。これは、システムへのアップロードファイルの形式を PDF ファイルとしたことで可能になっている。システムは、PDF ファイルを受け付けたのち、それを画像データに変換して蓄積するので、もはや問題作成の手段に制限はほとんど受けない。付録図 16 に問題登録の仕組みを示す。

提案するシステムによって実際に登録作業を行う際には、初期画面から登録画面に案内されると、まず登録する単元を選び、次に単元内での個別の問題登録を行う。このシステムでは多肢選択では記号や数値も選べるようにしている。



問題文の下にデリミッターとしての直線を1本入れ、その下に解答および解説部分がかかれた登録用のPDFファイルをアップロードすると、デリミッターを解読して問題文の部分が出題画面に、解答と解説部分が解答を終わった後で示される解説画面に表示される。登録時には、ファイルアップロード時に同時に正答を入力しており、アップロードファイルと関連づけられている。この登録システムでは、問題の登録だけでなく、修正、追加、削除も容易に行なうことが可能である。

## 参 考 文 献

- Hirose, H. (2016a). Meticulous learning follow-up systems for undergraduate students using the online item response theory, 5th International Conference on Learning Technologies and Learning Environments, 427–432.
- Hirose, H. (2016b). Learning analytics to adaptive online IRT testing systems “Ai Arutte” harmonized with university textbooks, 5th International Conference on Learning Technologies and Learning Environments, 439–444.
- 廣瀬英雄, 藤野友和 (2015). 『確率と統計—Web アシスト演習付』, 培風館, 東京.
- Hirose, H. and Sakumura, T. (2010). Test evaluation system via the web using the item response theory, The Thirteenth IASTED International Conference on Computers and Advanced Technology in Education, 152–158.
- Hirose, H., Takatou, M., Yamauchi, Y., Taniguchi, T., Honda, T., Kubo, F., Imaoka, M. and Koyama, T. (2016). 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, 433–438.
- Hirose, H., Takatou, M., Yamauchi, Y., Taniguchi, T., Kubo, F., Imaoka, M. and Koyama, T. (2017). Rediscovery of initial habituation importance learned from analytics of learning check testing in mathematics for undergraduate students, 6th International Conference on Learning Technologies and Learning Environments, 482–486.
- 桂利行, 池田敏春, 佐藤好久, 廣瀬英雄 (2015). 『理工系学生のための線形代数—Web アシスト演習付』, 培風館, 東京.
- 桂利行, 岡崎悦明, 岡山友昭, 齋藤夏雄, 佐藤好久, 田上真, 廣門正行, 廣瀬英雄 (2017). 『理工系学生のための微分積分—Web アシスト演習付』, 培風館, 東京.
- 桑幡隆行, 作村建紀, 廣瀬英雄 (2012). adaptive online IRT system について, 情報処理学会火の国シンポジウム, A-1-4.
- 文部科学省中央教育審議会大学分科会大学教育部会 (2016). 「卒業認定・学位授与の方針」(ディプロマ・ポリシー), 「教育課程編成・実施の方針」(カリキュラム・ポリシー)及び「入学者受入れの方針」(アドミッション・ポリシー)の策定及び運用に関するガイドライン.
- Sakumura, T. and Hirose, H. (2017). A bias reduction of abilities for adaptive online IRT testing systems, *International Journal of Smart Computing and Artificial Intelligence*, **1**, 59–72.
- Tokusada, Y. and Hirose, H. (2016). Evaluation of abilities by grouping for small IRT testing systems, 5th International Conference on Learning Technologies and Learning Environments, 445–449.
- Waddington, R. J., Nam, S., Lonn, S. and Teasley, S. D. (2016). Improving early warning systems with categorized course resource usage, *Journal of Learning Analytics*, **3**, 263–290.

## A Large-scale Testing System for Learning Assistance and Its Learning Analytics

Hideo Hirose

Faculty of Environmental Studies, Hiroshima Institute of Technology

One of the most crucial issues in universities, where a variety of enrolled students are educated to the level of universities' diploma policies, is to identify students at risk for failing courses and/or dropping out early, to take care of these students, and to reduce their risk. For this purpose, in April 2016, Hiroshima Institute of Technology implemented a newly developed online testing system to evaluate students' abilities into the follow-up program for fundamental undergraduate education; the system is based on item response theory. Since then, the system has been operating well. The subjects are basic analysis (calculus) and linear algebra. The accumulated learning data are sufficient for assessment of primary learning analytics. In this paper, we describe our case as a large-scale testing system for steadily accumulating learning data, and then explore whether we can identify students at risk by analyzing such data in the early stages. It is worth mentioning that risk factors that were originally ambiguous have revealed by statistical analysis of the data. Although the academic subject examined in this study was mathematics, this kind of system could easily be applied to other subjects, including statistics, statistics education, and STEM (science, technology, engineering, and mathematics).