

独立性の検定に関する HTML をベースとした教材を用いた シミュレーション学習における授業設計の影響の分析

東邦大学・薬学部 金子 真隆

Masataka Kaneko, Faculty of Pharmaceutical Sciences, Toho University

山口大学・教育学部 北本 卓也

Takuya Kitamoto, Faculty of Education, Yamaguchi University

東邦大学・理学部 野田 健夫

Takeo Noda, Faculty of Science, Toho University

1 はじめに

筆者の先行研究 [1] において、推測統計分野の中でもアンケート調査の結果分析などで応用範囲の広い「独立性の検定」に関する HTML をベースとしたシミュレーション用のコンテンツについて提案した。これは第2著者が開発を進めているシステム [2] を用いて作成されたものである。さらに筆者の先行研究 [3] では、このコンテンツを大人数の授業で利用するために改変した経緯について説明した上で、実際に利用した際の状況について報告した。特に、学習者がコンテンツを操作した際のログを取得し、これを Post Test (PT) の結果と照合して分類することにより、at risk の学習者を推定するための特徴量を操作ログから探索すると同時に、それに基づいて有効な教育的介入の方略を見出すスキームを提案した。しかし、結果的に、at risk の学習者を推定する最低限の指針が示唆される一方で、コンテンツを制約なしに操作させる授業フローでは、このトピックを学習する際に学習者にかかる高い認知負荷 [4] をなかなか低減できず、PT のパフォーマンスを期待したほど向上させられないことが判明した [3]。これに伴って、操作ログから抽出した特徴量によって PT のパフォーマンスやその背後にある認知負荷の程度を推定することが極めて難しい状況となった。カリキュラムの関係上、別の学習者を対象に同様の授業を実施する機会があったため、コンテンツを操作させる前後での授業設計を少し変更した。本研究では、これに伴って操作ログや PT のパフォーマンスがどのような影響を受けるか、分析することとした。加えて、学習内容の定着度を確認するため、授業実施から約1ヶ月後に Delayed Post Test (DPT) を実施し、その結果と操作ログとの照合も試みた。これは、やや長期にわたる学習活動のパフォーマンス評価に該当するが、ラーニングアナリティクスの観点でも、こうした場合の検証には難しさが伴うことが指摘されている。本稿でも、対象学生へのアンケート調査の実態から、そうした難しさの一面にふれてみたい。

2 教育工学的背景

ラーニングアナリティクスの分野で ICT の活用実態を検証する研究としては、少なくともその最初期において、LMS に残る学習活動のログを用いるものが主流であった。

目的は、初期の研究の代表例 [5] の論文タイトルにも見える通り、LMS 上での学習活動のログの分析をもとに at risk の学習者を早期に推定し、効果的な学習支援につなげるというものである。[5] でも採用されている典型的な手法は、LMS 上の複数種類の学習活動のログデータに対して重回帰分析を適用し、最終的な学習成果に関する predictive model を探索して、そこから予測される結果と実際の学習の成否を比較することにより、得られた model の predictive power を検証するというものである。これに引き続く先行研究 [6] では、2010 年頃までの当該分野における研究のメタレビューを行っているが、Web of Science, PsycInfo, ERIC といった大規模なデータベースから組織的な探索を行っているにもかかわらず、対象となった研究が 34 にとどまり、一定の基準が設けられていたとはいえ、当時までの研究事例がそう多くなかったことを示している。このレビューでは、ICT ツールの利用が教育効果を持つことを自明とはしない立場がとられている。実際、レビューした研究では LMS 上の各種ツールの利用頻度を変数として学習成果との関連を調べるものがほとんどであるのに対し、[6] ではツールの利用パターンによって学習者を分類した上で、分類した集団の間でツールの利用効果を比較するアプローチが採られており、その結果からは、的確な model を構築する上で、学習タスクからの要求・学習の文脈とツール利用との整合性、およびツール利用に関する学習者間のばらつきについて、詳細に分析する必要があることが示唆されている。さらに、discussion board における議論や、教室でのインタラクションなどの、Face-to-Face の要因を分析する必要がある、これが LMS ツールが想定したほど使われていない理由を解明する手掛かりとなる可能性にも言及されていて、その稀有な研究例として [5] が引用されている。実際その中では、discussion board 上のコミュニケーションを図 1 のように可視化し、学習成果とコミュニケーションのネットワークとの関係が分析されている。

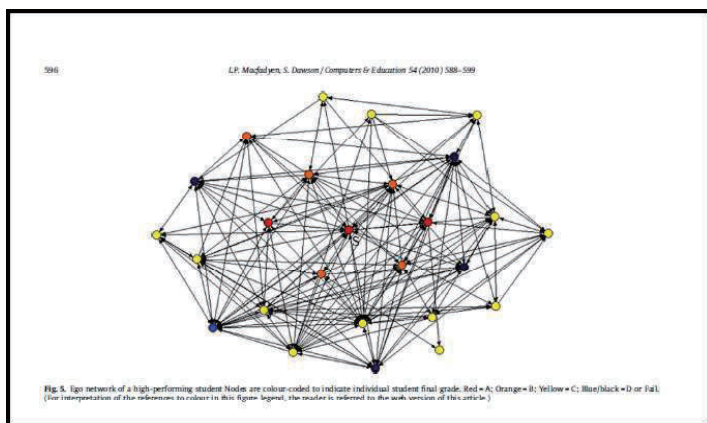


図 1: 学習者間のコミュニケーションの可視化

ただし、これはあくまでオンラインでのコミュニケーションの分析であり、対面でのものではない。当然のことながら、対面でのコミュニケーションについてはデータ収集自体が容易でなく、現在でもラーニングアナリティクス以外の学習科学の分野での研究が中心である。[6] の著者たち自身も、対面とオンラインの混成のスタイルの授業を対

象とした分析を行い [7], コミュニケーションのデータを学習支援のセッションを担当したサポートスタッフの記録から得た上で, LMS 上でのコンテンツの活用パターンと併せて学習者を分類することを試みている. しかし, この例のような長期にわたる学習活動の場合, こうした分類に立脚したとしてもなお, LMS 上のコンテンツの利用頻度から学習成果を予測しようとする, 大きな困難が伴うことを指摘している. その要因として, 学習者の素養やツールへの慣れが挙げられているが, 筆者の目からすると, 学習者間で行われるコミュニケーションが相当の影響を持ちうるものであり, 学習活動の duration が長くなればなるほど, その内容に関するデータの取得・分析が困難になるという事情が背景にあると考えるのが自然である.

その後, 2010 年代に入って LMS の機能がさらに向上し, 学習活動のログデータが大量かつ容易に入手できるようになると, predictive model の研究も多数行われるようになってくるが, その一方で, データのスケールの大きさのみに依存したアプローチには限界があることが意識されるようになってきた. 実際, 同一大学における 9 コースに登録される 4000 人程度の学生の大規模なログデータを用いた研究でも [8], コース横断で行った解析と各コース個別に行った解析とを比較しているが, ログデータとパフォーマンスの連関性がコースによって大きくばらつく結果となり, ツールの利用がコースや授業全体のデザインにどのように組み込まれているか, 慎重に検討する必要があることが指摘されている. 一例として, コースコンテンツへのアクセス頻度とパフォーマンスとの連関性がコース横断では有意であるのに対し, 調査対象となった数学や情報科学のコース個別ではこれが有意にならないという結果が報告されている.

以上の経緯をふまえて本研究では, LMS 上のコンテンツへのアクセス頻度などと比べ, より学習活動の内容に密着したコンテンツの操作ログをもとに, シミュレーションを行わせる際の設定を中心とした授業設計を変更した場合, 操作プロセスと PT におけるパフォーマンスの間の連関性がどのように変わるか, 分析することとした. これに加え, 以上の分析結果を DPT の結果とも対照し, 特に学習者間のコミュニケーションがどのように介在したか, 追究することとした.

3 実験授業のフローと授業設計の差異

2 回実施した実験のうちの 1 回目に関しては, [3] で紹介してあるが, 2 回目の実験に関しても, PT の内容を含めて基本的な流れは同じであったので, コンテンツや授業実施のフローの詳細はそちらに譲る. 対象となる学習者に関しても, 1 回目の実験とまったく同様で, 1 回目が女性 80 名・男性 29 名の計 109 名だったのに対し, 2 回目は女性 76 名・男性 36 名の計 112 名と, ほぼ同等の母集団での実施となった. [3] にもふれた通り, この試行は通常授業の一環として行われたが, 実施時の学習者の状況に関しても, 2 回の実験の間にまったく差異はない状況であった.

2 回目の授業の際に大きく変更されたのは, コンテンツを用いたシミュレーションを行わせる際の制約条件である. 常識的には薬剤群での治癒例数はコントロール群での治癒例数よりも大きな整数となることが想定されるが, 1 回目においては, [3] にも示した通り, 数理モデルとして幅広い探索を行わせたいという考え方から, このような「正の

相関」を示す事例だけでなく、薬剤群での治癒例数がコントロール群のそれに満たないような、「負の相関」を示す事例についても、探索を妨げなかった。これに対し2回目においては、「正の相関と負の相関」についてシミュレーションを行わせる前に明示的に説明し、検定の設定上、薬剤群での治癒例数は60から100の間の整数となることが常識的であるとして、入力値に事実上の制限を設けた。このような設定変更のもとで、2回目に関しても、入力値の時系列的な変化について1回目と同様の可視化 [3] を行い、視察による比較を行った。2回目に関しては、探索に際して一定の制約を設けたものの、自由度が引き続きかなり高く、ログデータの視察のみによって認知負荷に関する推定を行うことは困難だと判断された。そのため、1回目と2回目、およびPTの正解群と不正解群の間で差異が確認され得る特徴量を入力のログから抽出することを試みた。

DPTに関しては、1回目と2回目の対象者が同時に受験する定期テストの際に実施した。図2はDPTの問題の画像であるが、PTでは62と38としていた薬剤群における治癒例数と非治癒例数をそれぞれ63と37に変更しただけのものである。

1. 薬剤投与の効果を検証するため、病気が治癒したかどうかを薬剤を投与された人とされていない人各100人に尋ね、独立性の検定にかけたいことを考える。投与されていない人のうち、治癒例が60、非治癒例が40であるとし、投与された人の内訳を変えていったとき、得られるp値は次の表のようになる。

	治った	治らない	p値
60	40	1	
61	39	0.885	
62	38	0.7719	
63	37	0.6629	
64	36	0.5601	
65	35	0.4652	
66	34	0.3795	
67	33	0.3039	
68	32	0.2386	
69	31	0.1835	
70	30	0.1382	
71	29	0.1018	
72	28	0.0733	
73	27	0.0515	
74	26	0.0353	
75	25	0.0235	
76	24	0.0153	
77	23	0.0037	
78	22	0.0059	

この表を参考にして、投与されていない人のうち治癒例が63、非治癒例が37である場合に、p値が以下の各値と最も近くなるのは、それぞれ投与された人のうちの治癒例がいくつの場合か答えよ。解答は63から100までの整数値とすること。

(1) $p = 0.9$

(2) $p = 0.1$

(3) $p = 0.05$

(4) $p = 0.01$

ただし、下図のクロス集計表において、各 T_i を $S_1, S_2, \dots, S_m$ に比例配分して得る期待度数を $e_{11}, e_{12}, \dots, e_{lm}$ とおくととき

$$\chi^2 = \sum_{i,j} \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$$

は S が大きければ近似的に自由度 $(l-1)(m-1)$ の χ^2 分布に従い、独立性の検定ではこれを検定統計量としていることを用いてよい。

	B_1	$\dots$	B_m	計
A_1	a_{11}	$\dots$	a_{1m}	T_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_l	a_{l1}	$\dots$	a_{lm}	T_l
計	S_1	$\dots$	S_m	S

図2：DPTの問題画面

PTの終了後、正解やそれに至る考え方について解説したが、「独立性の検定」に関してはそもそも正規の授業内容に加えての補足的な内容としてアナウンスしており、定期試験に出題することもまったく予告していなかった。これに伴い、授業終了後から定期試験にかけて、Moodle上などで該当の内容に関する教材もまったく提示していなかった。このような状況から考えて、学習者側に試験対策として解法だけ記憶しようというインセンティブが働いたとは考えづらい面がある。このため、PTの結果に対してDPTでパフォーマンスに著しい改善がみられた学習者を対象に、改善の背景に関するアンケート調査を実施することとした。該当者は1回目の授業に関して女性27名・男性9名の

計 36 名，2 回目の授業に関しては女性 27 名・男性 8 名の計 35 名で，これら合計 71 名に対し，授業回の区別をせずに，一律に回答を求めた．主たる問いかけの内容は，学習者間でのコミュニケーションがパフォーマンスの改善にどの程度寄与したかという点である．具体的な問いかけの内容は表 1 の通りで，冒頭の DPT における解答方針に関する問いかけに対する回答（ア・イ・ウ）に応じ，以後の問いかけを場合分けした．

解答に際してどのような考え方に従ったか？	
ア	薬剤群とコントロール群とで治癒例数を同じだけ増やすと期待度数もその分だけ増えるため、カイ2乗統計量の分子が変化しないと考えた
	このような考え方を理解できたタイミングはいつか
	このような考え方を理解するにあたり、誰かの助けを借りたか
	借りたのであればそれは誰か
イ	アの内容までは理解していなかったが、薬剤群の治癒例数と同じだけコントロール群の治癒例を増やせばよいというやり方を記憶していた
	このようなやり方を知らせてくれたのは誰か
	アの考え方について現段階で理解できるか
ウ その他	
	その考え方を簡単に説明せよ
	その考え方をだれかに教わったか
	教わったのであればそれは誰か
	アの考え方について現段階で理解できるか

表 1：アンケートでの質問内容

回答結果をもとに，対象者集団の中で解法などに関するコミュニケーションがどの程度行われたか，可視化することを試みた．

4 実施結果

2 回目の授業時に，PT の解答の入力は確認できるものの，シミュレーションのログが確認できない対象者が 3 名あったが，出席状況に疑義があったため，該当者のデータは分析対象から除外した．前述の 109 名と 112 名というのはこれを除外した後の数字である．また，2 回の実験とも，シミュレーション時の入力 が 1 回にとどまり，その後の探索が進んでいない事例が複数（1 回目は 4 例・2 回目は 8 例）確認されたが，これは送信トラブルである可能性は小さく，周囲の友人のシミュレーションを観察していた可能性が高いため，データ分析の対象から除外した．これらの対象者も含めた PT の実施結果は表 2 に示す通りである．2 回目の授業で PT のパフォーマンスが大幅に向上していることから，シミュレーションの際に新たな制約条件を課すという授業設計の変更により，対象者にかかる認知負荷を軽減する効果が生じたと判断される．これ以後の分析では，完全正解群と部分正解群を合わせて正解群として扱い，不正解群との間で比較することとする．

分類	条件	1回目	2回目
完全正解群	完全正解 または p=0.01の場合のみ不正解であるもの	13	23
部分正解群	全体にフィットは良いが、p=0.9の場合に 62よりも小さい値を解答しているもの	8	13
不正解群	不正解と判断されるもの（表の値を そのまま読んだと考えられるもの・内数）	88(35)	76(17)
合計		109	112

表 2：PT の結果

1 回目の授業の対象者による操作プロセスについては [1] で分析した通りだが，シミュレーション実施時の学習者にかかる認知負荷の程度が時系列的な操作のゆらぎに反映されるのではないかという事前の想定に反し，探索の時間や操作数，入力値の総変動量の時間平均のいずれについても，2 群間での有意差が確認されず，これらの特徴量をもとに認知負荷の程度を推定することは困難な状況であった．この背景として，探索に際しての制約条件が弱く，操作の自由度がかなり高かったために，入力値の変動が生じやすかったことが考えられる．ただし，正解群にはほとんどみられない一方で，不正解群には一定頻度で見られるような操作パターンとして，正解ゾーン（60 から 80 までの間）を飛び越えて上下の領域を往復するパターンが見出され，これが at risk の学習者を推定する手がかりとなる可能性が示唆された．2 回目の授業に関しては，シミュレーションに制約条件が加わることで，これらの特徴量が 2 群間を識別する指標となりうる可能性を考慮し，対象者ごとに探索プロセスの可視化を行ったので，その結果のサンプルを以下に示す．

例えば，図 3 に示される不正解群の事例では，総変動量の時間平均は小さいものの，正解ゾーンにほとんどヒットしておらず，[1] の図 7 の事例と類似している．

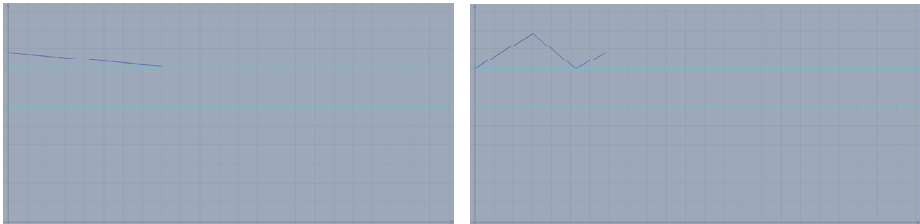


図 3：不正解群の操作プロセスの事例（I）

また，図 4 に示される不正解群の事例事例では，正解ゾーンの周辺が探索されている上に，総変動量の時間平均が大きくはないものの，すべての入力の中で正解ゾーン内にヒットしているものが極めて少なく，[1] の図 8 の事例と類似している．

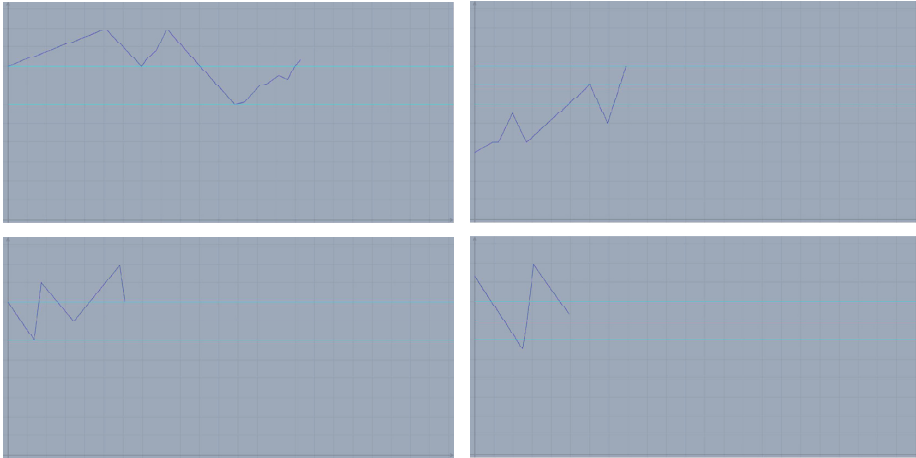


図4：不正解群の操作プロセスの事例（II）

これらを含めた2回目の可視化を1回目と比較すると、操作時間・入力数・総変動量のいずれも顕著な減少が確認される。これは、2回目に探索の自由度が抑制されたことによる当然の結果である。しかし、総変動量の時間平均をとってみると、予想に反して1回目と2回目の間、および正解群と不正解群の間のいずれについても有意差が確認されず、この指標から認知負荷の推定や授業設計の影響の検証を行うことは難しい状況であった。以上の結果に関し、これまで筆者が行ってきた動的コンテンツの操作ログに関する研究の観点からみて問題となるのが、特徴量の抽出にあたり経過時間のファクターが考慮されていない点であった。動的コンテンツの操作のように、図形要素を連続的に動かす場合であれば、各状態に経過時間の重みをつけることが考えられるが、本試行のように入力が散発的になる場合には、経過時間の意味づけ自体が難しい、という事情がその背景である。そこで、新たな特徴量を探索するにあたり、経過時間ではなく操作の順番によって重みづけを行うこととした。また、同一の変動量であっても、正解ゾーンに近いか遠いかで意味合いが異なると考えられるため、変動量のかわりに正解ゾーンからの距離を用いることとした。

以上のようにして見出された新たな特徴量の計算方法について、図5の事例を用いて説明する。

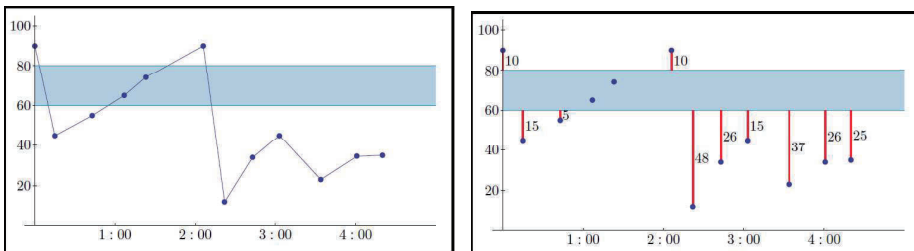


図5：操作プロセスの事例と正解ゾーンからの距離

まず，図5の右図にあるように，各入力値と正解ゾーンの間の最短距離を算出する．正解ゾーン内の入力に関してはこれを0とする．次に，図6でイメージされるように，操作の順番で重みづけする．時間的に後の操作ほど，独立性の検定に関するイメージが定まって，探索として有意義なものに近づくことを想定してのものである．

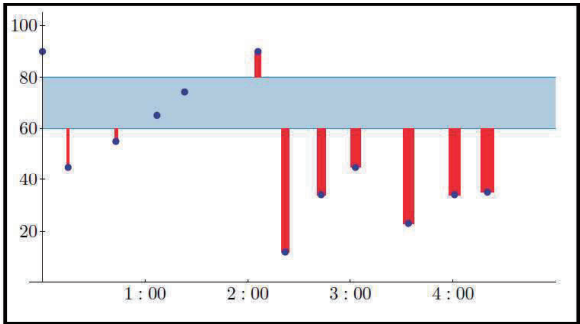


図6：操作の順番による重みづけのイメージ

具体的に図6の場合には，正解ゾーンとの距離に入力順を乗じた上で総和をとり，入力順の総和で除して特徴量を定義し，「重み付き域外探索量」と呼ぶことにする：

$$\frac{1 \times 10 + 2 \times 15 + 3 \times 5 + \cdots + 11 \times 26 + 12 \times 25}{1 + 2 + 3 + \cdots + 12}$$

この特徴量を対象者個々に計算し，授業回ごとにヒストグラムとしてまとめたのが図7である．ここで，左側の図が1回目，右側の図が2回目の結果である．いずれも，不正解群のデータをオレンジ色，正解群のデータを紫色で表示しており，2群のデータをオーバーレイ表示している．1回目の不正解群で2例，2回目の不正解群で4例が横軸の表示範囲におさまらず，表示から除外されている．

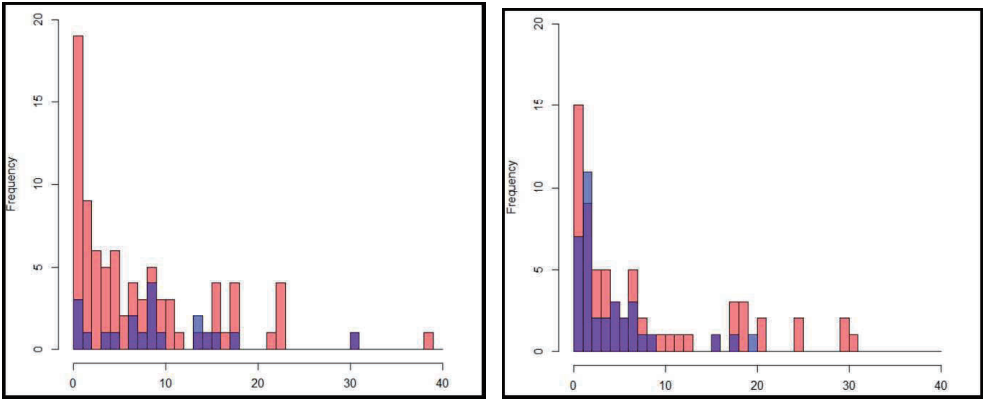


図7：重み付き域外探索量の分布

図7をみると，全体として2回目の特徴量の分布が，1回目と比べ，0に近い範囲により集中していることが観察される．これは，特徴量を総変動量の時間平均から重み付き

域外探索量に取り替えることで、授業設計の影響をよりの確に把握できる可能性を示唆している．さらに、不正解群の分布に対して正解群の分布が、1回目と比べて2回目に、0に近い範囲により顕著に偏っていることも観察される．実際、不正解群と比べて正解群の重み付き域外探索量が小さくなるという対立仮説を立てて Wilcoxon の順位和検定を実施すると、1回目に関しては p 値が 0.8864 となって、まったく2群の差を確認できないのに対し、2回目に関しては p 値が 0.0351 となって、2群の差が有意であることが確認される．この結果からは、2回目における制約を加えた授業設計の下で、重み付き域外探索量が高いとした場合、該当の学習者にかかる認知負荷が軽減できていないことを示し、 χ^2 統計量の数式表現に注目するように促す形の教育的介入が必要だと教授者が判断する上で、有効なシグナルとなりうる可能性が強く示唆される．

第3節に触れた通り、DPT については、PT と比べて正解者が 70 名程度増加する結果となり、パフォーマンスの向上が顕著であった．その一方で、前段にみたような操作ログの特徴量と DPT のパフォーマンスとの連関性がまったく確認できない状態となった．PT での正解者が DPT でもほぼ確実に正解していた結果をふまえると、DPT におけるパフォーマンスの改善が、コンテンツを用いたシミュレーションとは別の要因によってもたらされた可能性が高いと考えられる．第3節にとりあげた対象者間のコミュニケーションという要因は、このようにして浮かび上がってきたものである．アンケート調査に対しては 71 名中 63 名が回答したが、回答内容をまとめたのが表 3 である．

解答の根拠		教員以外の 助力あり	現段階での理解			
			YES	NO		
ア 概念レベル	13	0	13	0		
イ やり方レベル	36	8	13	23		
ウ その他	14	0	4	10	数式表現を意識	4
					根拠なし	10

表 3：アンケート調査の結果

図 8 は、定期試験の受験者全員を対象に、PT と DPT の結果にアンケートの回答を併せて集計したもので、1 段目が 1 回目、2 段目が 2 回目の授業の対象者のデータである．また、各回の上段は PT、下段は DPT の結果に対応する．さらに、各対象者のデータが縦 2 行にわたって示されており、左の列から右の列にかけて、学籍番号順に並べられている．対象者の所属する学部では、1 学年の定員が 240 名程度と多いため、入学当初から学籍番号の近い学生どうしでグループ化し、そこでコミュニケーションが取れるようにする仕組みが用意されており、過去の例でも、このグループが卒業までの集団行動の単位となるケースが多い．特に、数理統計のような座学の授業の際は、グループの学生どうしが近い場所に着席し、意見交換や質疑応答を行う場面がしばしば見られる．図の中で、水色のセルは、PT に関しては完全正解もしくは部分正解、DPT に関しては完

全正解だったことを示す。これに対し、赤色と黄色のセルはアンケート調査によるもので、前者は教えた側、後者は教わった側として名前が挙がったことを示す。黄緑色のセルは自力でDPTの解答に至ったとの申告があったことを示す。PTで赤色や水色の表示があるにもかかわらず、DPTで正解となっていない事例がみられるが、いずれも誤った箇所は限定的だったことが確認されている。

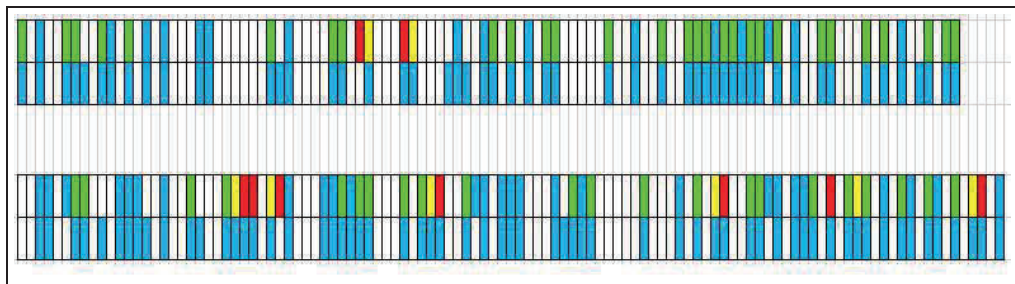


図8：PT/DPTの解答状況とアンケート結果の可視化

図中の赤色と黄色のセルの分布から容易に推察される通り、いずれのコミュニケーションも学籍番号が近い学生どうしの間で発生したことが確認されている。

5 総合論議と今後の研究の方向性

図7に明示される通り、操作ログの分析の観点で本研究から得られる中心的な知見の一つは、新たに加わった操作の制約条件に合わせて特徴量の算出方法を変更することにより、PTの結果に関する predictive power が上昇したという点である。これは、第2節でもふれた、学習タスク・学習の文脈とツール利用との整合性に十分な注意を払うべきとの主張 [6] に完全に符合するものである。

これに対し、DPTに関しては、第4節にもふれた通り、操作ログからの予測が極めて難しい結果となった。対象者間のコミュニケーションの影響を調べるためにアンケート調査を実施したわけだが、図8に示される結果からみる限り、PTからDPTにかけてのパフォーマンスの向上を説明する上で、十分な状況とはなっていない。ただし、図8の可視化を詳細にみると、この解釈には若干注意が必要だと考えられる。その根拠は、図8におけるDPT正解者の分布について、いくつかのクラスターが観察されるという点である。PTのセルが黄緑色で表示されている対象者からは、「自力で理解した」という回答が得られているわけであるが、授業の際に他者からの情報提供が全くなかったという保証は実はない。場合によっては、「教えられた」という自覚がないまま、実は重要な情報を周囲からもたらされていたという可能性も考えられるわけである。また、そもそもが大教室の授業であることから、仮に学籍番号が隣接した学生どうしの間で直接の知識伝達がなかったとしても、周囲を囲む別の学生集団の対話から間接的に情報が伝わるという可能性は否定できない。このような状況をふまえると、図8に示されるクラスターの中で、なんらかの情報伝達があった可能性は否定できないと考えられる。図1の事

例のように，学習者間の対話がすべてログとして残っている場合と比べ，本研究の場合は，基本的に対象者の自己申告によらざるを得ないことが，このような不十分な結果に結びついた面は否めないが，逆に，学習コミュニティの特性を事前に把握しておくことで，対面授業の環境下における学習者間のコミュニケーションを効率的に把握出来る可能性が示唆されているとも考えられる．

本研究で得られた知見を今後の当該テーマの教育に生かして行く上では，まだ課題も多い．そもそもラーニングアナリティクスは，そこから得られた情報が学習活動を深める結果に結びつくことで初めて意味をもつものである [9]．そうしたフィードバックのプラットフォームとして，筆者も Learning Dashboard[10] の活用を提案してきたところであるが [3]，教授者が学習者に対して具体的にどのようなアドバイスを行うかという点で，さらなる研究が必要だと考えている．

謝辞

本研究は JSPS 科研費 21K02752，22K02951 の助成を受けている．

参考文献

- [1] 北本卓也, 金子真隆, 野田健夫 : 独立性の検定に関する HTML をベースとした教材の作成と授業設計, 京都大学数理解析研究所講究録 2236, pp. 127–139, 2022.
- [2] T. Kitamoto, M. Kaneko, S. Takato : "E-learning system with Computer Algebra based on JavaScript programming language", Proc. of ATCM 2018, Yogyakarta, 2018, 123–133.
- [3] 北本卓也, 金子真隆, 野田健夫 : 独立性の検定に関する HTML をベースとした教材の操作ログの解析, 京都大学数理解析研究所講究録****, pp. –, 2023.
- [4] J. Sweller, P. Ayres, S. Kalyuga : Cognitive Load Theory, Springer, 2011.
- [5] L. P. Macfadyen, S. Dawson : Mining LMS data to develop an "early warning system" for educators: A proof of concept, Computers and Education 54, pp. 588–599, 2010.
- [6] G. Lust, N. A. J. Collazo, J. Elen, G. Clarebout : Content Management Systems: Enriched learning opportunities for all? Computers in Human Behavior 28, pp. 795–808, 2012.
- [7] G. Lust, M. Vandewaetere, E. Ceulemans, J. Elen, G. Clarebout : Tool-use in a blended undergraduate course: In search of user profiles, Computers and Education 57, pp. 2135–2144, 2011.
- [8] D. Gasevic, S. Dawson, T. Rogers, D. Gasevic : Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success, Internet and Higher Education 28, pp. 68–84, 2016.

- [9] C. Rose, E. MaLaughlin, R. Liu, K. Koedinger : Explanatory learner models: Why machine learning (alone) is not the answer, *British Journal of Educational Technology* 50-6, pp. 2943–2958, 2019.
- [10] J. P. Tan, E. Koh, C. Jonathan, S. Yang : Learner Dashboards a Double-Edged Sword? Students' Sense-Making of a Collaborative Critical Reading and Learning Analytics Environment for Fostering 21st-Century Literacies, *Journal of Learning Analytics* 4(1), pp.117–140, 2017.