

文部科学省科学研究費補助金 萌芽研究 研究報告書

平成18年度～平成20年度

「グループ学力の評価法開発」

課題番号：18650252

平成21年3月

研究代表者：永岡慶三

早稲田大学 人間科学学術院・教授

Faculty of Human Sciences, Waseda University

## 研究組織

研究代表者	永岡慶三	早稲田大学・人間科学学術院・教授
連携研究者	竹谷 誠	拓殖大学・工学部・教授
連携研究者	北垣郁雄	広島大学・高等教育研究開発センター・教授
連携研究者	米澤宣義	工学院大学・工学部・教授
連携研究者	赤倉貴子	東京理科大学・工学部・教授
連携研究者	植野真臣	電気通信大学・情報システム学研究科・准教授
連携研究者	藤原康宏	岩手県立大学・ソフトウェア情報学部・講師
連携研究者	荘島宏二郎	大学入試センター・研究開発部・准教授

## 研究経費

平成18年度 1,000千円 平成19年度 1,300千円 平成20年度 900千円

総額 3,200千円

# 主な研究業績

## 1. 学会・研究会発表等

- ・ 永岡慶三, 植野真臣 (2007-08) グループ学力測定法とグループ受験, 日本テスト学会第5回大会, ポスターセッション発表.
- ・ 永岡慶三, 植野真臣 (2008-08) グループ学力におけるGuttmanスケールモデルと局所独立モデル, 日本テスト学会第6回大会, 口頭発表 (IRT1), 44-45.
- ・ 永岡慶三, 植野真臣 (2006-11) グループ学力の測定・評価 - 協調効果の計量評価 -, 日本教育工学会第22回大会講演論文集, 1p-B201-1, 469-470.
- ・ 藤原康宏, 西久保健太, 永岡慶三 (2008) ピア・アセスメント支援システムを利用した紙媒体のレポートの相互評価の実践, 電子情報通信学会教育工学研究会, ET2008-15, pp. 1-6.
- ・ 西久保健太, 赤倉貴子, 永岡慶三 (2008-9) オンデマンド授業における学習者の反応発信の検討 - 二種類の発信方法の比較 -, 教育システム情報学会 第33 回全国大会講演論文集, B6-6, 264-265
- ・ 西久保健太, 赤倉貴子, 永岡慶三 (2008-10) オンデマンド授業におけるリアルタイムチュータリングの実現, 日本教育工学会第24回全国大会講演論文集, 517-518
- ・ 米澤宣義, 野本竜哉, 平井和人 (2008-11) 理解構造と差異度に基づく学習者のグループ化アルゴリズム, 先進的学習科学と工学研究会, 54, 41-46, 6
- ・ 富田聡, 神保智広, 木原幸一郎, 杉林俊雄, 竹谷誠: 機械設計授業におけるグループ学習の学習効果の検証, 電子情報通信学会信学技報, ET2008-62, pp. 23-28 (2008)
- ・ 神保智広, 富田聡, 竹谷誠: ロジカルフローグラフを用いたグループ学習の学習効果と評価, 電子情報通信学会総合大会 (2009)

## 2. 学術論文・紀要等

- ・ 植野真臣, e テスティング: 先端理論と技術, 教育システム情報学会 (2009) 年後発刊予定
- ・ 荘島宏二郎・大津起夫・田栗正章 (2006) 自己組織化マップと完全情報最尤法から見た平成 17 年度センター試験の教科学力構造, 日本テスト学会誌, 2, 76-89.
- ・ 荘島宏二郎 (2007) 自己組織化マップによる平成 18 年度大学入試センター試験の教科学力構造, 日本テスト学会誌, 3, 162-178.

### 3. 国際学会発表

- M. Takeya, N. Yasugi, K. Nagaoka: Measurement and Evaluation for a Concept Mapping Test by Drawing Ordering Relations among Concepts, Proceedings of the Second International Conference on Concept Mapping, Vol.2, 188-191 (2006)
- Makoto Takeya, Masanao Kobayashi, Tatsuya Nakauchi, Ikuo Kitagaki and Keizo Nagaoka : A Navigation Method Dependent on Individual e-learning Strategies, IADIS International Conference e-Learning 2007, e-Learning 2007 Volume II , pp.17-20 (2007)
- Ikuo Kitagaki, Atsushi Hikita, Makoto Takeya, Yasuhiro Fujihara( 2007) Development of An Algorithm for Groupware Modeling for A Collaborative Learning, International Journal of Computers, Communications & Control Vol.II、 No.1,pp.66-73.



## 研究概要

### 【研究目的】

本研究の目的は、複数の人間により形成されるグループに対して「グループの学力」を定義し、その評価法を開発することにある。ここに言う「学力」は知識量だけでなく、「能力」や「コンピテンシー」という概念も含むものとする。

近年、複数の学習者が互いにコミュニケーションをとりながら学ぶことを目指した協調学習など、新しい教育・学習方法が研究、実践されている。ネットワーク情報基盤環境の普及にともないCSCL(Computer Supported Collaborative Learning)あるいはPBL(Project Based Learning)など複数の学習者がグループを形成して、協調しながら学習する考え方である。大学などにおいてもCSCLやPBLの方式による協調学習を授業に取り入れ実施されている先進的事例もあるが、最終的な成績評価は、結局のところ多くの場合、各学習者個人についての出席率や普段点、提出レポートの得点で行われている。協調学習が本来育成しようとしている”グループをなすことによって向上する”学力(能力)を妥当に評価できる手法の開発が望まれる。

研究目標はグループ学力の測定・評価に関わるデータの計量的な分析法およびその理論的基盤となる数理モデルの開発・提唱である。最近の教育評価に関する一つの流れとしての新しい評価の考え方として質的評価などが言われるが、趣旨にはおおいに賛同しその発展を期待するものであるが、現段階での研究成果では計量的・客観的手段を提供するには至っていない。近代のテスト理論における項目反応理論は、数理モデルに基づいたコンピュータの演算力とともに、個人の学力については成熟した技術となっている。このレベルの理論的方法をグループ学力についても開発・提唱を目指したい。

### 【研究の特色】

まず、評価対象をグループ(複数人の学習者集団)としたことである。従来、学力は「個人」の知識量や思考力に関するものであることは疑う余地さえ無かった。歴史ある教育・学習の測定/評価やテスト理論においても対象は学習者個人以外にはありえなかった。しかし、前述のように、社会の変容につれて学校教育をはじめとする現代教育システムに求められる機能は変化しつつある。特に、高度に発達しつつあるICT情報通信技術の活用により、急速に発達している協調学習などの新しい教育・学習支援システムに関する研究が活発であるが、評価法については相対的に進んでいない。それら新しい方法により、健全で正当な教育・学習が達せられるためには、早急な評価方法の開発が必要と考える。

協調学習/CSCLの研究開発は、国内の関連諸学会、日本教育工学会・教育システム情報学会・電子情報通信学会教育工学研究会などにおいて、近年活況を呈している。たとえば、日本教育工学会の論文賞、研究奨励賞は最近数年間の受賞のほぼ半分はCSCL/協調学習関連の研究である。他学会でも同様の状況にある。一方、テストに関する最近の研究は、これまでのテスト理論(項目反応理論)が円熟期を迎え、2003年に設立された日本テスト学会においても現状のテストを基準として標準化が進められている。いわば、パラダイムシフトを次段階のステップとして待望している感がある。

### 【受検者/学習者側の信頼性/内部一貫性】

テストを科学的に、計量的に扱うもっとも根本的な主要概念は信頼性である。信頼性(の推定方法)は考え方により多くの定義があるが、理論的明快さや実用性など種々の利点から最も多く用いられるのはCronbachの $\alpha$ 係数である。本研究においても信頼性といった場合、Cronbachの $\alpha$ 係数をさすものとする。

グループ学力を論じる場合、受検者/学習者側の信頼性を論じる必要があることが、本研究での一つ

の発見であった。グループ学力の概念化・定式化，測定データの整理・測定法開発，評価法のための分析法・数理モデル開発を研究目的として，グループ学力のモデルとして局所独立モデルと Guttman スケールモデルとを提唱した。それらは従来，テスト項目群の特性として用いられる概念であった。ただし，受検者側あるいは学習者側については信頼性は語感として相応しくない。同様の趣旨であるが内部一貫性や一次元性の方が用いる言葉として抵抗が無い。

なお Guttman スケールは実測データに対する記述的特性であり，局所独立は統計モデル的特性である。Guttman スケールを強いて IRT の 2 パラメータ・ロジスティックモデルの上で言えば，識別力のきわめて大きい項目群といえる。このとき項目群が Guttman スケールの特性を備えていれば，学習者／受検者側においても内部一貫性は高い値となり，必然的に Guttman スケールの特性を満たす。

ところが，項目群側に局所独立を仮定した場合，その項目群側の特性とは独立に学習者／受検者側に独立性の程度を仮定できる。すなわち，内部一貫性(信頼性)は大きい値も小さい値もとりうる。このことからグループ学力を大きくするためには，できるだけ内部一貫性の小さいグループ構成をすることであるといえる。

### 【グループ受検とグループ学力】

あるテスト項目  $j(a, b)$  が与えられたとき，グループ構成メンバー  $m$  人の潜在特性値を  $(\theta_1, \dots, \theta_m)$  とする。この際，一般的には多次元とされる能力(因子)潜在特性値  $\theta$  を一次元として議論する。

グループ学力は，そのグループのメンバーの個別学力  $(\theta_1, \dots, \theta_m)$  にて，あるいはそれらのみにて規定されるグループ学力は，代表値を以下に決めるかである。

$$\theta_G = f(\theta_1, \dots, \theta_m)$$

たとえば，

$$1) \text{ 平均値 } \theta_G = 1/m \sum_{i=1}^m \theta_i$$

また，Guttman スケール項目群を仮定すれば，

$$2) \text{ 最大値 } \theta_G = \max_i (\theta_1, \dots, \theta_m)$$

あるいは局所独立を仮定して，

$$3) \theta_G = P^{-1} [1 - \prod_i \{1 - P(\theta_i)\}]$$

ここに  $P^{-1}$  は逆関数である。

これは 2 人なら，1)~3) はそれぞれ，

$$\theta_G = (\theta_1 + \theta_2) / 2$$

$$\theta_G = \max(\theta_1, \theta_2)$$

$$\theta_G = P^{-1} [1 - (1 - P_1)(1 - P_2)]$$

となる。

たとえば， $N$  人の受検者集団から構成員数 2 のグループを構成するとする。すなわち二人ずつのペアを組むとする。このときグループ数(ペア数)は  $N/2$  である。 $N$  が奇数のときは最後を 3 人組にする等適宜に対応する必要があるが，ここでの議論には本質ではないのでこれ以上踏み込まない。

このときどのようなペアを構成していけばグループ学力の値が全体的に大きくなるか。2) が成り立つなら，各ペアの大きい方の  $\theta$  値でそのペア(グループ)の  $\theta$  値は決まるので，個別学力の大きい順に第 1 位から第  $N/2$  位までの  $N/2$  人を異なるグループに配し，それ以降の順位の第  $N/2+1$  位から第  $N$  位までの  $N/2$  人はどこに配されても同じで，平均グループ学力は，個別学力の大きい順に第 1 位から第  $N/2$  位までの  $N/2$  人の個別学力の平均値となる。ただ，この答えの方法論では，いささか自明解である。

しかし、ここに受検者側に内部一貫性の特性を導入すれば、話は別である。グループ(ペア)内の $\theta$ 値の大小だけでは決まらないのである。すなわちペアで考えれば、お互いがお互いの弱点を補い合うような組合せ、すなわちその2名の1,0得点パターンの排他的論理和が全体で最大化するような組合せを行うことで最大化が見込まれる。

さて、以降の詳細については、本文内の研究成果を参照いただきたい。

## 目次

1. グループ学力の提唱	
1-1 グループ学力の測定・評価 - 協調効果の計量評価 -	8
1-2 グループ学力測定法とグループ受験	10
1-3 グループ学力における Guttman スケールモデルと局所独立モデル	12
2. 教育・学習評価システム開発	
2-1 オンデマンド授業における学習者の反応発信の検討 - 二種類の発信方法の比較 -	14
2-2 オンデマンド授業におけるリアルタイムチュータリングの実現	16
2-3 e テスティング：先端理論と技術	18
2-4 A Navigation Method Dependent on Individual e-learning Strategies	31
2-5 Measurement and Evaluation for a Concept Mapping Test by Drawing Ordering Relations among Concepts	35
3. グループ学習	
3-1 機械設計授業におけるグループ学習の学習効果の検証	39
3-2 ロジカルフローグラフを用いたグループ学習の学習効果と評価, 電子情報通信学会総合大会	45
3-3 Development of An Algorithm for Groupware Modeling for A Collaborative Learning	46
3-4 理解構造と差異度に基づく学習者のグループ化アルゴリズム, 先進的学習科学と工学研究会	54
4. テスト	
4-1 自己組織化マップと完全情報最尤法から見た平成 17 年度センター試験の教科学力構造	60
4-2 自己組織化マップによる平成 18 年度大学入試センター試験の教科学力構造	74
4-3 ピア・アセスメント支援システムを利用した紙媒体のレポートの相互評価の実践	91

# グループ学力の測定・評価 - 協調効果の計量評価 -

Measurement and Evaluation for Group Ability

永岡 慶三 植野 真臣

Krizo NAGAOKA, Maomi UENO

早稲田大学, 電気通信大学

Waseda University, University of Electro-Communications

<あらまし> 古来, 教育の場において「学力」は個人に対するものでしかなかった。これまでも学力観については多くの議論がなされ, 新しい提案もされてきたが, やはり学力は個人の固有のものとして測定され, 評価されてきた。本研究では複数の学習者からなるグループに対しての学力を提唱する。この概念を「グループ学力」と呼ぶこととする。本稿では, 従来の個人学力を複数人の学力へ拡張した形での数理モデルを導入して, 測定・評価の方法を提案する。協調学習における協調効果の計量評価方法の可能性を開くものである。

<キーワード> テスト, 学力, グループ学力, 教育測定, 教育評価, 数理モデル, データ解析, グループ学習, 協調学習, グループ受験

## 1. 研究目的

本研究の目的は, 複数の人間より形成されるグループに対して「グループ学力」を定義し, その測定・評価の方法を開発することにある。なお, 以下で受験者または受検者の方が適切な語である場合もあるが学習者で統一する。

### 1.1 グループ学習・協調学習

近年, 複数の学習者が協調して学ぶことを目指したグループ学習や協調学習など, 新しい教育・学習方法に関する研究・開発されている。それらは ICT の進歩によるネットワーク情報基盤環境の普及にともない CSCL: Computer Supported Collaborative Learning あるいは PBL: Project Based Learning などで実践もされている。

### 1.2 新しい教育・学習への評価法においても個人の学力が対象

大学などにおいてそうした新しい教育・学習が授業に取り入れ, 実施されている先進的事例もあるが, 最終的な成績評価は結局のところ多くの場合, 各学習者個人について出席率や普段点, 提出レポートの採点により行われている。現行の大学の成績評価は知る限りすべて個人について行われているのは事実である。

### 1.3 実社会では共同作業

いうまでもなく, インターネットを始めとするネットワーク情報基盤環境の普及にともない, 実社会では多くの分野で人々は対面による以外に, 遠隔地間でもネットワークを介してコ

ミュニケーションを頻繁に行い, 共同・グループワークにより仕事を進めている。一個人が独立して作業を進める場面は非常に限られてきている。こうした状況下において, 今後の実社会に求められるのは, 単に個々人の学力が高いだけでなく, 目的に対応して形成された複数人からなるグループが有する学力であり能力であり, その評価法である。

本研究では, グループ学力について, テスト理論における項目応答理論 IRT: Item Response Theory を拡張する形で, 数理モデルを導入する。このモデルを用い, 協調学習における協調効果の計量評価方法の可能性を開くものである。

## 2. グループ受験とグループ学力

まず, 簡便のため通常のテストを想定する。テストというと答案用紙を配布して, 個々人が独力で回答する場合は普通であるが(他人との相談はむしろ不正行為), ここでは複数の学習者が相談など協調しながら回答する場面を考える。これを グループ受験 と呼ぶこととする。

以下, グループ受験により測定するグループ学力に関し, テスト理論における項目応答理論の自然な拡張を目指して数理モデルを構築する。グループ学習のいっそうの進展にも, 計量的/科学的評価法は欠かせないものとする。

### 2.1 項目応答理論

通常の学習者個人が受験する場合, 項目応答理論では, 学習者の学力を一般には多次元の潜在特性値(学力)  $\theta$  で表現するが, ここでは実用的に

も仮定されることの多い一次元の  $\theta$  とする。

学力  $\theta_i$  を持つ学習者  $i$  が識別力  $a_j$ , 困難度  $b_j$  を持つ項目  $j$  を受験した場合の正答確率は 2-パラメタロジスティックモデルでは次となる。

$$P_j(\theta_i) = 1/[1 + \exp(-a_j(\theta_i - b_j))] \quad (1)$$

## 2.2 グループ学力への拡張

簡単な例として、漢字のテストを考えた場合、学習者はその漢字を知っていれば正答でき、知らなければ正答できない。多肢選択式項目などのように偶然正答はない。(1)式で、 $a_j$  値の非常に大きな項目である。正答確率は、 $u(\cdot)$  を単位ステップ関数として、次のように簡単に書ける。

$$P_j(\theta_i) = u(\theta_i - b_j) \quad (2)$$

すなわち、学力  $\theta_i$  より難しい項目 ( $b_j > \theta_i$ ) は正答できず、易しい項目 ( $b_j < \theta_i$ ) は正答できる。

$N$  人の学習者 ( $\theta_1, \dots, \theta_N$ ) がグループ受験した場合、この種の項目は最低一人が正答できれば、そのグループは正答できることになる。実測データで表現すれば、4 学習者 ( $S_1, \dots, S_4$ ) が 5 項目 ( $I_1, \dots, I_5$ ) を受験し、正誤が下記の表 1 のようだったとすればグループ  $G$  の成績は最下行のようになる。グループで唯一正答を出せないのは全員が 0 点の項目 5 だけとなる。4 人の正答数が 3, 2, 1, 1 で平均正答率が 35% であるのに対し、グループでは正答数 4 で正答率が 80% と大きくなる。

表 1 実測データの例

	項目 1	項目 2	項目 3	項目 4	項目 5
学習者 1	1	1	1	0	0
学習者 2	1	0	0	1	0
学習者 3	1	0	0	0	0
学習者 4	0	0	1	0	0
グループ	1	1	1	1	0

グループ学力  $\theta_G$  は (1), (2) をもって、次のように既定値の  $\theta, a, b$ , により数式表現できる。

$$P_j(\theta_G) = 1 - \prod_{i=1}^N [1 - P_j(\theta_i)] \quad (3)$$

$$= 1 - \prod_{i=1}^N [1 - 1/[1 + \exp(-a_j(\theta_i - b_j))]] \quad (1)'$$

$$= 1 - \prod_{i=1}^N [1 - u(\theta_i - b_j)] \quad (2)'$$

ただし、上記では、 $P_j(\theta_i)$  が互いに独立であることの仮定が必要である。これは項目応答理論における (項目側の) 局所独立の仮定に相当する学習者側の局所独立の仮定であるが、項目応答理論に学習者側の局所独立に相当する概念がない。

## 3. 協調効果

上の例の漢字のテストでさえ、正確な知識も持

たない学習者どうしも、相談して不確かな知識を補い合い、触発し合って、正確な漢字を書くに至ることもありうる。まして、相談・協調が本質的に重要な問題解決型の課題では、一人一人が受験してできない項目も三人よれば文殊の知恵で、グループ受験によって正答確率が大きくなるのが通常期待できる。

### 3.1 協調効果の項を加えたグループ学力モデル

この相談・協調による正答確率の増分を Collaboration または Communication の頭文字をとって  $C_G$  書くこととし、次のように協調効果の項を (3) に加えてグループ学力を定義する。

$$P_j(\theta_G) = 1 - \prod_{i=1}^N [1 - P_j(\theta_i)] + C_G \quad (4)$$

数式 (3) に必要であった”学習者側の局所独立の仮定”が無く、従属的であった場合には (3) が成立しなくなるが、学力の従属性は協調効果をもたらしている要因の一つであり  $C_G$  に含まれる。この学力の従属性の他、静的な学力以外の、協調学習ではむしろ主題である協調能力やコミュニケーション能力などによる相互作用が協調効果の項にさらに寄与すると思われる。

### 3.2 実際の適用について

協調学習を意図して、グループを組んで課題解決などを行った場合、どのくらい協調効果があったか知りたいところである。そのための具体的手段として、次のような手順が想定できる。

そのグループの  $N$  人の学習者個々に、等質な  $n$  項目からなるテストを受験させ、得点一覧表 ( $x_{ij}$ ) を得る (表 1 の第 4 行までに相当)。このデータから計算によって得られる潜在学力値 ( $\theta_1, \dots, \theta_N$ ) をもって、各学習者の個人ベースでの学力とみなせる。

同じく等質なテスト項目よりなるテストをグループ受験させる。このときのグループ受験では十分に相談し合い、グループで一つの回答をするようにさせる。このとき得られる  $n$  項目に対する得点ベクトル (表 1 の最下行に相当) を ( $x_{.j}$ ),  $j=1, \dots, n$  とし、これより計算される  $\theta_G$  との差分を協調効果  $C_G$  の計量値として得ることができる。

## 4. おわりに

項目応答理論を拡張して、グループ学力の数理モデルを提案した。このモデルを用いて、協調効果を計量的に評価する可能性を示した。

まだ、多くの条件や仮定を明確にして洗練させる必要があると思われる。現在、このモデルに適合するデータを収集している

# グループ学力測定法とグループ受験

○永岡慶三\*, 植野真臣\*\*

\*早稲田大学 人間科学学術院, \*\*電気通信大学 大学院情報システム学研究所

<キーワード> テスト, 学力, グループ学力, 教育測定, 教育評価, グループ受験

## 1. はじめに

個人に対して定義される学力・能力を複数人のグループに概念拡張して「グループ学力」を定義する。

古来、学力は一個人について付与される概念であった。しかし近年、社会の各分野でのネットワーク基盤の普及もあり、人々はグループで協調して仕事にあたる場合は多くなってきており、一個人が独立して作業を進める場面は非常に限られてきている。

こうした状況下において、今後の実社会に求められるのは、単に個々人の学力が高いだけでなく、目的に対応して形成された複数人からなるグループが有する学力であり能力である。そして、その測定法・評価法である。

## 2. グループ学力の定式化

グループ学力の測定方法としては、特定モデルを導入して複数人の個人学力(通常の学力を以降あえて個人学力という)から合成計算する方法と、テストを複数人が協調して受験するグループ受験の方法の両者が考えられる。

個人学力からグループ学力を合成計算する方法としては、IRT 項目反応理論を複数受験者へ対応する定義拡張を行う。

一人の受験者に対する項目反応理論では(学力の潜在特性値  $\theta$  は一次元とする)、項目を 2-パラメータと規定して識別力  $a_j$ 、困難度  $b_j$  を持つ項目  $j$  を受験した場合の、個人学力  $\theta_i$  を持つ受験者  $i$  の正答確率は次となる。

$$P_j(\theta_i) = 1/[1 + \exp(-a_j(\theta_i - b_j))] \quad (1)$$

グループ学力  $\theta_g$  の正答確率への、 $N$  人の個人

学力  $[\theta_1, \dots, \theta_N]$  の反映のさせ方として、典型的な場合を、個人学力どうしが全く独立の場合と全く従属的な場合にとる。

### 2.1 受験者学力が局所独立である場合

グループ学力の正答確率が個人学力の和となる場合。(1)式は次のように拡張できる。つまり誰かが正答すればグループ全体で正答となるようなテスト項目である場合である。

$$P_j(\theta_g) = 1 - \prod_{i=1}^N [1 - 1/[1 + \exp(-a_j(\theta_i - b_j))]] \quad (2)$$

ただし、上記では、 $P_j(\theta_i)$  が互いに独立であることの仮定が必要で、これは項目応答理論における項目側の局所独立の仮定に相当する受験者側の局所独立の仮定といえる。

### 2.2 受験者学力が Guttman Scale 的な場合

受験者の個人学力が一次元 Guttman Scale 的であった場合、すなわち個人学力  $\theta_i$  が一次元的で  $a_j$  値の非常に大きな項目反応曲線を描く場合、正答確率は、 $u(\cdot)$  を単位ステップ関数として、次のように簡単になる。

$$P_j(\theta_i) = u(\theta_i - b_j)$$

したがって、

$$P_j(\theta_g) = 1 - \prod_{i=1}^N [1 - u(\theta_i - b_j)] \quad (3)$$

である。グループ内でもっとも高い個人学力  $\theta_i$  が  $b_j$  を上回れば  $P_j(\theta_g) = 1$ 、そうでなければ  $P_j(\theta_g) = 0$  となるような、テスト項目と受験者の組み合わせである。

一般にはテストとグループの組み合わせにより、上記のモデルの混合の状態になると思われるが、その程度を計数化できれば個人学力からグループ学力を予測計算できて便利である。

さらに実際に次に述べるグループ受験においては受験者どうしがコミュニケーションにより、いわゆる「3人よれば文殊の知恵」効果が生じることが観察される。

### 3. グループ受験

学校での期末試験など通常のペーパーテストを想定する。配布された答案用紙を個々の受験者が独力で回答する場合が普通であり、他人との相談は通常は不正行為であるが、ここでは複数の受験者が相談・協調などコミュニケーションしながら回答する場面を考える。これを**グループ受験**と呼ぶこととする。

同一のテストを個人受験とグループ受験の両方で実施し、グループ受験で得られる測定値と前節のモデルにより個人学力から合成計算されるグループ学力により想定される測定値とを比較し、検討する。

比較事例：漢字検定の4級, 3級, 2級, 準1級のテストより選抜した項目 60 問よりなる漢字の書き取り・読みのテストをとりあげた。

前節(3)式のモデルの検証を試みる意味で識別力  $a_j$  の大きいテストの例として、受験者がその漢字を知っていれば正答でき、知らなければ正答できないという偶然正答の少ない典型と考えたからである。

受験者は大学生 30 人で、同一のテストを個人受験した後一週間後に二人一組でグループ受験をした。個人受験後に採点フィードバックなどしていないので、同一テストの受験による学習効果の影響はないものと考えられる。

表1はそのうち2名でグループ受験を行った受験者の、個人受験得点とグループ受験得点である。9問をとりあげた。第一行目はグループ受験による得点で、2, 3行目は個人受験の得点で二人のうち高かった得点(最大値)を○で囲っている。4行目は [グループ受験得点] - [個人受験得点最大値] で、多くはグループ受験により得点上昇がみられる。

簡素この上ないと想定した漢字のテストでも、正確な知識も持たない受験者どうしが相談して不確かな知識を補い合い、触発し合って、正確な漢字を書くに至ることがあった。また逆に、個人受験測定値では正答だった受験者がグループ受験により他人の影響を受け、誤答してしまう例さえ観察された。

### 4. まとめ

以上より、(3)式にさらにグループ受験のコミュニケーション効果による増分  $C_G$  を加え、

$$P_j(\theta_j) = 1 - \prod_{i=1}^N [1 - P_j(\theta_i)] + C_G \quad (4)$$

と書きグループ学力の定義式とする。

(3)式に必要であった”学習者側の局所独立の仮定”が無く、従属的であった場合には(3)式自体が成立しなくなるが、学力の従属性は協調効果をもたらしている要因の一つであり  $C_G$  に含まれる。この学力の従属性の他、静的な学力以外の、コミュニケーション能力などによる相互作用が協調効果の項にさらに寄与すると思われる。

[謝辞] 本研究の一部は H18 年度三浦健介・加藤昌文卒論による。

[文献] 永岡 慶三, 植野 真臣(2006-11)グループ学力の測定・評価 - 協調効果の計量評価 -, 日本教育工学会第 22 回大会講演論文集, 1p-B201-1, 469-470.

グループ受験(2人)得点	52	39	42	49	44	47	45	47	49
個人受験者1 得点	25	<b>40</b>	25	<b>44</b>	34	38	<b>45</b>	37	<b>49</b>
個人受験者2 得点	<b>41</b>	30	<b>42</b>	37	<b>41</b>	<b>42</b>	25	<b>39</b>	42
[グループ受験得点] - [個人受験得点最大値]	9	-1	0	5	3	5	0	8	0

表1 グループ受験と個人受験の測定値比較



# グループ学力における Guttman スケールモデルと局所独立モデル

○永岡 慶三\*

植野 真臣\*\*

\* 早稲田大学人間科学学術院, \*\* 電気通信大学大学院情報システム学研究所

## 1. はじめに

複数の受検者よりなるグループに対して定まる一つの学力を“グループ学力”とここでは呼ぶ(永岡ら 2006, 永岡ら 2007). 最近の教育の場では CSCL: Computer Supported Collaborative Learning あるいは PBL: Project Based Learning といった協調学習やグループ学習など, 学習者がコミュニケーションをとり協同で学習する方法が重視されてきている.

当然, 現在および今後の実社会において, 多くの社会的活動はネットワーク基盤を通して, 多人数のグループにより行われる. こうしたことから, 個人の固有能力としての学力と同様, 複数の人間からなるグループについて一つの能力や学力を定義し, 測定または推定できることは有用であると考えられる.

## 2. グループ学力, グループ受検

以下では用語を次のとおりとする.

- ・ グループ学力: 複数人よりなるグループの学力
- ・ 個別学力: 従来とおりの個人固有の学力
- ・ グループ受検: テストを複数受検者が協同して回答すること. その回答を採点することでグループ学力の値とする.
- ・ 個別受検: 従来とおりの個人が回答・受検すること.

グループ学力をどのように測定あるいは規定するかは, 次の2通りが想定される.

- 1) 一つは, グループ受検による直接的測定である. 文字通り, 一つのテストを複数受検者が協同して回答し, その回答を採点することでグループ学力の値とする.
- 2) もう一つは, 受検者個人が通常の形式で 個別受検して測定された学力から, グループ構成員の受検者間の関係の情報を用いて合成し, グループ学力の推定値とする. この計算のためにモデルを設定する必要がある.

## 3. IRT の拡張によるグループ学力モデル

上記の 2) を計算するため, 項目反応理論 IRT をモデルとして採用し, 定義式をグループ学力に拡張する(永岡ら 2006).

受検者  $N$  人の潜在特性値(個別学力)を一次元とし, 学力  $\theta_i$  を持つ学習者  $i$  が識別力  $a_j$ , 困難度  $b_j$  の項目  $j$  を受検した場合の正答確率は 2-パラメータロジスティックモデルでは次となる(定数係数  $D$  等は省略)

$$P_j(\theta_i) = 1/[1 + \exp(-a_j(\theta_i - b_j))] \quad (1)$$

以下, 実測得点は正答 1 点, 誤答 0 点の 2 値とする. (実測データの表現としての)Guttman スケールモデルを仮定すれば, (1)式において  $a_j$  値の非常に大きな項目であり, 正答確率は,  $u(\cdot)$  を単位ステップ関数として, 次のように簡単に書ける.

$$P_j(\theta_i) = u(\theta_i - b_j) \quad (2)$$

このような項目に対し,  $N$  人の受検者 ( $\theta_G: \theta_1, \dots, \theta_N$ ) がグループ受検したとすれば, (1),(2)式より

$$P_j(\theta_G) = 1 - \prod_{i=1}^N [1 - P_j(\theta_i)] = 1 - \prod_{i=1}^N [1 - u(\theta_i - b_j)] \quad (3)$$

となる. ただし, 通常の項目反応理論では言われない受検者側における内部一貫性や局所独立性を仮定する必要がある.

## 4. 項目側と受検者側の $\alpha$ 係数比較

グループ学力を扱う場合, 受検者側の内部一貫性を扱わざるを得ないが, IRT にもとづき  $\alpha$  係数を推定する正確な議論でも, 受検者側は一次元  $\theta$  のだけであつかわれているようである(荘島ら 2002).

いくつかのテストデータに対する内部一貫性について, Cronbach の  $\alpha$  係数を計算した結果が表 1 である. 受検者側についても同データの逆行列について

表1 各テストの項目側と受検者側の $\alpha$ 係数  
(受検者種類はほとんどが同一大学の同一学部生)

テスト種類	正答率	項目数	項目側 $\alpha$ 係数	受検者数	受検者側 $\alpha$ 係数
数学計算問題・文章題	0.753	20	0.543	20	0.915
ドラマの関心度	0.670	60	0.862	89	0.929
漢字検定	0.627	60	0.821	16	0.903
世界遺産に関する知識	0.490	35	0.580	42	0.935
芸能人認知度テスト	0.574	20	0.792	50	0.949
アニメクイズ	0.504	25	0.877	44	0.951
2007年のスポーツ知識	0.468	50	0.851	32	0.877
マナー診断テスト	0.562	25	0.579	43	0.827
数理パズル	0.647	6	0.449	34	0.870
数理パズル(グループ受験)	0.764	6	0.210	17	0.869
漢検	0.452	30	0.826	44	0.954
漢検(グループ受験)	0.606	30	0.738	22	0.924

$\alpha$ 係数相当のものを計算した。データサイズが小さいので統計的なことは言えないが、案の定、受検者側の内部一貫性が項目側に比して大きい。

表1の結果から、さらに $\alpha$ 係数だけを取り出して、散布図に示したのが図1である。項目側 $\alpha$ 係数がかなり小さい値のものまであるのに対し、受検者側の $\alpha$ 係数はすべてのデータで項目側のそれより大きく、数値は最小でも0.827という大きさである。

表1にあるようなテストはほとんどが知識の有無で正誤がきまる(2)式のあてはまる種類のものである。しかも受検者はかなりの等質集団であり、数値からも内部一貫性はきわめて高い。

しかしながら、永岡ら(2007)にあげた事例のように、グループ受験により得られるデータに(3)式のモデルがかならずしもあてはまらない。このことはグループ受験におけるコミュニケーション効果を示唆する。

したがって、コミュニケーション効果による正答確率の変化分(事例では必ずしも増加にならない)の項 $C_G$ を補って、次をグループ学力のモデルとする。

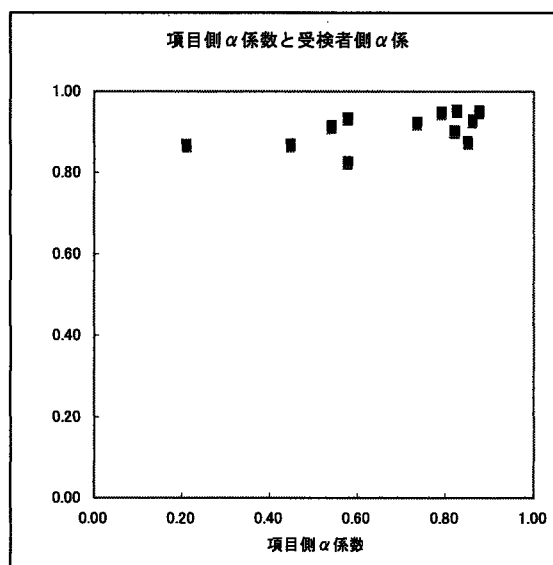


図1 散布図：項目側と受検者側の $\alpha$ 係数

$$P_j(\theta \mid \alpha) = 1 - \prod_{i=1}^N [1 - P_j(\theta_i)] + C_G \quad (4)$$

## 5. 適用

既存の個別受検データからグループ学力を推定できれば、協調学習CSCLの学習効果評価、共同作業やプロジェクト構成の最適化の指標などの道具となり便利である。あるいは逆にグループ受験を実施することで、個別学力による推定値との差異から、グループの相性を客観的に計量できる。テスト項目群の性質(一次元に限定すればパラメータ $a, b$ ですむ)とコミュニケーション効果 $C_G$ への影響が明確なテストを設計することも今後重要である。

### [謝辞]

本研究に有益な示唆をいただいた大学入試センター荘島宏二郎氏、ならびにデータ処理などの支援を受けた荻谷謙太君はじめ研究室の卒業研究生に謝意を表します。

### [文献]

荘島宏二郎, 豊田秀樹(2002)項目反応理論におけるCronbachの $\alpha$ 係数の推定, 心理学研究, 73, 227-233.

永岡慶三, 植野真臣(2006-11)グループ学力の測定・評価-協調効果の計量評価-, 日本教育工学会第22回大会講演論文集, 1p-B201-1, 469-470.

永岡慶三, 植野真臣(2007-08)グループ学力測定法とグループ受験, 日本テスト学会第5回大会, ポスターセッション.

# オンデマンド授業における学習者の反応発信の検討

## - 二種類の発信方法の比較 -

### Examination of Sending Learner's Reaction to the Teacher in On-demand Class -The Comparison of Two Methods-

西久保 健太<sup>†</sup> 赤倉 貴子<sup>‡</sup> 永岡 慶三<sup>†</sup>

Kenta NISHIKUBO<sup>†</sup> Takako AKAKURA<sup>‡</sup> Keizo NAGAOKA<sup>†</sup>

<sup>†</sup>早稲田大学 人間科学学術院

<sup>‡</sup>東京理科大学 工学部

<sup>†</sup> Faculty of Human Sciences, Waseda University

<sup>‡</sup> Faculty of Engineering, Tokyo University of Science

Email:nishikubo@ruri.waseda.jp, akakura@ms.kagu.tus.ac.jp, k.nagaoka@waseda.jp

あらまし：オンデマンド授業において、「学習者-教師」間のコミュニケーションを補完するために、学習者が聴講中に反応を発信できる機能を付加した。付加した機能は「いつでも反応を発信できる機能」「定期的に反応を発信させる機能」の2種類である。システム評価実験でのアンケートの5段階評価の結果より、どちらの機能も肯定的な回答が得られた。「いつでも反応を発信できる機能」の記述式のアンケート結果より、反応の発信に対する即時フィードバックの必要性が示唆された。

キーワード：オンデマンド授業，VOD，eラーニング，反応発信

## 1. はじめに

近年、情報通信技術の発達により、授業映像を配信するオンデマンド授業が普及した。しかしながら、オンデマンド授業では、「学習者-教師」間のコミュニケーションが欠落するという問題点が指摘されている[1]。学習者は受け身となり、教師はそのときどきの学習者の状態を把握することができないため適切な指導がしづらい状況にある。

これまで、教師が参考にする学習者の情報を調査し、聴講中に学習者の反応を発信するシステムを開発してきた[2]。しかし、どのように反応を発信するユーザインタフェースが学習者にとって有用なのか示されていない。

キーボードとマウス操作による学習環境を想定すると、聴講中に学習者が反応を発信するインタフェースは以下の2種類に大別できると考える。

- (1) 学習者が聴講中にいつでも能動的に自身の反応（質問・興味等）を発信できる機能。
- (2) 教員が設定したタイミングで学習者に反応（質問・興味等）を発信させる機能。

そこで本稿では、上記2種類の機能を実装し、利用状況と利用後のアンケートの結果を比較した。

## 2. システムの概要

開発した実験システムの画面を図1、図2に示す。

### (1) いつでも反応を発信できる機能（図1）

学習者は、授業に対し何かを感じたときに“画像を選択”“フォームに文章を入力”することにより反応を教師に発信することができる。

### (2) 定期的に反応を発信させる機能（図2）

教師が設定した授業のキリの良い所で、画面が切り替わり、学習者に“その時点までの授業の内容についてどう思ったか”を5つの画像から選択させる。次に、その時点までの授業についての質問内容と感

想をフォームに入力させる。反応を発信すると次の授業シーンに進む。



図1. いつでも反応を発信できる機能

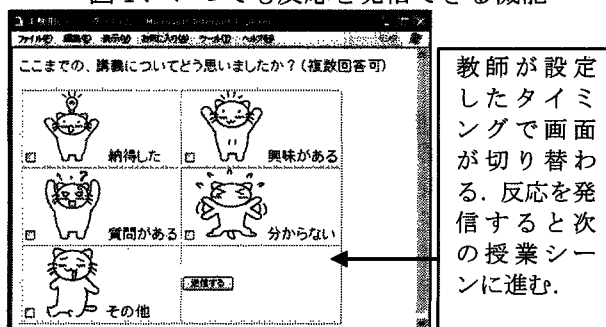


図2. 定期的に反応を発信させる機能

## 3. 評価実験

### 3.1 評価実験の概要

大学生（人間科学部3年生）6名を対象にシステム評価実験を行った。システムを使用して授業を受けてもらい、その後、理解度確認テストと事後アンケートを行った。授業映像は著者が作成したものであり、内容は3DCGについてである。授業時間は約12分であり、3分ごとに授業映像を区切り「定期的に反応を発信させる機能」を使用させた。実験手順を表1に示す。

表 1. 実験手順

事前アンケート	
事前理解度確認テスト	
システム使用 (図1)	授業シーン1 (座標について) 定期的な反応を発信1 (図2)
	授業シーン2 (モデリングについて) 定期的な反応を発信2 (図2)
	授業シーン3 (CSG表現について) 定期的な反応を発信3 (図2)
	授業シーン4 (スワイプ表現・パーティクルについて) 定期的な反応を発信4 (図2)
	定期的な反応を発信
事後理解度確認テスト	
事後アンケート	

3.2 理解度確認テストの得点

事前・事後テストの結果を図3に示す。テストは全5問である。事前・事後テストでは問題は異なるが、難易度は同程度になるように心がけて作成した。2名はテストの得点が変わらなかった。

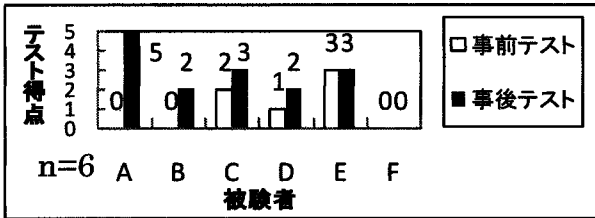


図 3. 理解度テストの得点

表 2. 機能の使用状況

被験者	A		F		C	
	いつでも反応を発信できる機能	定期的な反応を発信させる機能	いつでも反応を発信できる機能	定期的な反応を発信させる機能	いつでも反応を発信できる機能	定期的な反応を発信させる機能
授業シーン1	疑問		疑問		疑問	
定期的な発信1	興味	納得		分からない		興味
授業シーン2					疑問	
定期的な発信2		納得		その他	疑問	
授業シーン3					疑問	
定期的な発信3		納得		分からない	疑問	
授業シーン4					疑問	
定期的な発信4		納得		その他	興味	

3.3 システムの使用状況

「いつでも反応を発信できる機能」の使用回数は学習者によりばらつきがみられた。最も使用した学習者は15回反応を発信した(図3被験者C)。平均使用回数は4.5回であった(SD: 4.75)。

事後テスト(図3)の得点が最も高いAと、得点が最も低いFと、機能の使用回数が最も多かったCのシステムの使用状況を表2に示す。事後テストの得点が最も高かったAは、授業シーン1で疑問を感じたが、その後は授業内容を納得している。一方、得点が低かったFは、最初は授業に興味をもっていたにも関わらず、シーン1の途中で疑問を持った後、授業内容が分からないと回答している。今回の機能では、学習者の反応を発信させるだけで教師側からフィードバックを与えていない、そのためある地点で躓いてしまうと其後の授業が分からなくなってしまうことがある。一方Cは、授業の各箇所

を感じながらも最後は納得したと答えている。

3.4 事後アンケート

「いつでも反応を発信できる機能」と「定期的な反応を発信させる機能」についてそれぞれ“機能に満足したか”“機能を使用することにより授業の参加意識が高まったか”“機能を使用することにより授業に集中することができたか”を5段階評価で回答させた。さらに、その評価の理由について項目ごとに記述式で回答させた。

5段階評価の結果を図4に示す。いずれの項目も評価値の平均が3以上であるため肯定的な回答が得られたといえる。2つの機能の評価値の平均に有意な差は見られなかった。

「いつでも反応を発信できる機能を満足したか」の標準偏差が0.96であり(図4)他の項目と比べてばらつきが大きい。そこで「いつでも反応を発信できる機能を満足したか」の評価値が低かった被験者に、その理由を記述式で回答させたところ“質問をしても答えが出る訳ではない為”“クリックすると何が起きるのか謎だったから”という回答が得られた。このことから、反応を発信した際に、なんらかの即時フィードバックを与える必要があると考えられる。

「いつでも反応を発信できる機能」に対する肯定的な評価の理由として“聞いているだけにならない”“分からないところを探そうとした”という回答が得られた。このことから、学習者の授業への参加が促されたといえる。

一方、「定期的な反応を発信させる機能」に対する肯定的な評価の理由として“最後に(質問内容と感想を)書かないといけないという義務感があるので授業に集中できた”という回答が得られた。

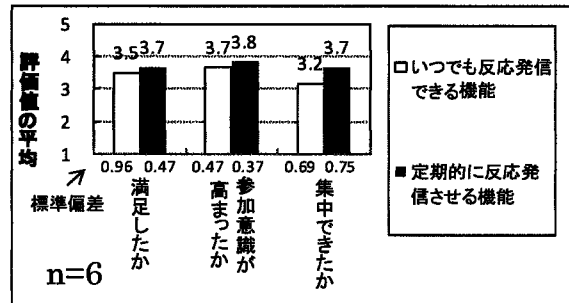


図 4. 事後アンケート結果 (5段階評価)

4. まとめと今後の課題

本稿では、オンデマンド授業において、学習者が聴講中に反応を発信できる2種類の機能を実装し、比較評価した。今後の課題として、学習者の反応に対し、教師側から学習者になんらかの方法で即時のフィードバックを与えることを検討することである。

参考文献

- (1) 赤倉貴子 (2005): “学習者の情報発信に視点をおいた非同期型 e-Learning System”, 日本教育工学会研究報告集 JET05-6, pp. 51-58 (2005)
- (2) 西久保健太, 赤倉貴子 (2006): “学習者が感情を発信できる VOD 型 e-Learning System の開発” 情報科学技術レターズ Vol.5, pp275-278 (2006)

# オンデマンド授業におけるリアルタイムチュータリングの実現

Implementation of the Real-time Tutoring in On-Demand Class

西久保 健太\*1 赤倉 貴子\*2 永岡 慶三\*1  
Kenta NISHIKUBO Takako AKAKURA Keizo NAGAOKA

\*1早稲田大学 WASEDA UNIVERSITY \*2東京理科大学 TOKYO UNIVERSITY OF SCIENCE

<あらまし> 本研究では、オンデマンド授業においてリアルタイムチュータリングを実現するシステムを開発した。評価実験のアンケートの結果より、リアルタイムチュータリングに対する肯定的な回答が得られた。また、リアルタイムでチューターが対応するという環境が用意されていること自体が、学習者にとってよい影響を与えるケースがあることが示された。

<キーワード> オンデマンド授業, VOD, チュータリング, eラーニング, システム評価

## 1. はじめに

近年、情報通信技術の発達により、授業映像を非同期に配信するオンデマンド授業が普及した。しかしながら、通常、オンデマンド授業では、聴講中に学習者は教師に質問・反応を発信できない。学習者は受け身となり、このことはモチベーションの低下の要因になりうる。従来、オンデマンド授業において、学習者が教師に対して質問をする場合、電子掲示板に質問内容を書き込むか、教師にeメールを送信することになる。これらの場合、学習者は授業映像を見終わった後に質問内容をよく吟味してから送信する。そのため、授業のある特定の場面で感じるような“ちょっとした質問”がしばらく状況にあると考えられる。また、電子掲示板やeメールによる指導では、学習者は教師からリアルタイムでフィードバックを得ることはできない。そのため、数学など内容に連続性がある授業では、学習者は教師からのフィードバックを得るまで、疑問が生じた以降の授業内容を理解できないケースが考えられる。一方、対面での教室授業では、授業中に学習者は自発的に挙手をして、教師やTAに質問することにより、その場でフィードバックを得ることができる。そこで、オンデマンド授業においても、学習者が聴講中にいつでも質問でき、リアルタイムでフィードバックを得ることができる環境があれば役立つと考えた。

以上をふまえ、本研究では、オンデマンド授業においてリアルタイムでのチュータリングを実現するシステムを開発した。

## 2. 開発システム

開発したシステムの概要を図1に示す。開発システムの内、リアルタイムでのチュータリングを行う機能を以下「チュータリング機能」と呼ぶ。

次に「チュータリング機能」の使用手順を説明

する。学習者は、授業について何かを感じたときに、4種類のキャラクタ画像(図2左上)から自分の状態に近いものを選んでクリックする。画像をクリックした時点で、学習者の状態はチューターに送信される。キャラクタ画像を選択させる理由は、学習者が質問内容をすぐ言語化できない場合であっても、何かを感じた時点でまず行動させることを狙っている。選択させるキャラクタ画像は、鈴木(2006)、西久保ら(2007)が実施した調査結果を参考に決定した。学習者は、キャラクタ画像をクリックした後、画面左下のフォームに質問内容を入力する。チューターはシステム上で学習者の授業ビデオの再生時間を把握することができ、リアルタイムで質問に対するフィードバックを与える。チューターとのやり取りのログは画面右下に表示される。

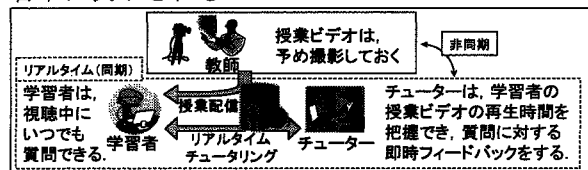


図1. システムの概要



図2. 学習者システム画面

表1. アンケート結果 (5段階評価)

質問項目	3DCG(n=7)		基礎統計学(n=8)	
	平均	標準偏差	平均	標準偏差
1「チューターからのフィードバック」は役に立ちましたか?	4.43	0.73	4.50	0.71
2「チューターからのフィードバック」により貴方の学習意欲は高まりましたか?	4.00	0.00	4.25	0.66
3「チューターからのフィードバック」の内容に満足しましたか?	3.29	1.39	4.00	0.71
4「チュータリング機能」に満足しましたか?	4.00	0.76	3.75	0.66
5「チュータリング機能」があることで講義への参加意識は高まりましたか?※1	4.33	0.47	4.00	0.71
6「チュータリング機能」があることで講義を最後まで集中して聴くことができましたか?	3.71	0.45	4.13	0.78

※1 質問5の3DCGの授業の評価について、1名が無回答であったため(n=6)として処理した。

### 3. 評価実験

#### 3.1 評価実験の概要

2008年5月29日、大学生(人間科学部3年生)10名を対象にシステム評価実験を行った。システムを使用して2つの授業を受講させた。各授業の後、理解度確認テストとアンケートに回答させた。

授業内容は3DCDおよび基礎統計学(t検定)である。3DCDの授業時間は約12分であり、授業ビデオは著者が作成した。3DCGの授業のチューターも著者(西久保)が担当した。基礎統計学(t検定)の授業時間は約17分であり、授業ビデオは博士課程の学生1名に出演を依頼した。基礎統計学(t検定)の授業のチューターは博士課程2名の学生に依頼し、その内1名はビデオに出演した者である。

#### 3.2 チュータリング機能の使用状況

3DCGの授業では、10名中7名の学習者が「チュータリング機能」を使用した。基礎統計学(t検定)の授業では、10名中8名の学習者が「チュータリング機能」を使用した。

#### 3.3 事後アンケート調査

それぞれの授業の後に、「チュータリング機能」を使用した学習者を対象に“チューターからのフィードバック”“チュータリング機能”の満足度に関する6項目を5段階評価で回答させた。表1示す結果より概ね肯定的な回答が得られたといえる。しかしながら、3DCGの“「チューターからのフィードバック」の内容に満足したか”の項目は評価値の平均が3.29であり若干低い(表1項目3)。そこで、チューターに求めることを記述式で回答させたところ“質問者の理解のレベルがどの程度かの見極め”“質問に対する返答がわかりやすいこと”という回答が得られた。

次に、“3DCG”と“基礎統計学(t検定)”では、どちらの授業のほうが「チュータリング機能」を活用できたか5段階評価で回答させ(表2)、さらにその評価の理由を記述式で回答させた。3DCGのほうが活用できた理由として、“自分に

とって興味がある内容だったから”“t検定は何を質問していいのかわからなかったの”という回答が得られた。一方、基礎統計学(t検定)のほうが活用できた理由として、“t検定のほうが難しく疑問が多かったから”という回答が得られた。このことから、「チュータリング機能」を学習者に活用させるには(1)適度な難易度の授業であること、(2)学習者の興味を惹くような内容であることが重要であると考えられる。

表2. どちらの授業が機能を活用できたか(n=10)

3DCG	どちらかという と3DCG	どちらでもない	どちらかという と基礎統計	基礎統計
2	1	3	3	1

今後また「チュータリング機能」を有したオンデマンド授業を受講してみたいと思ったかどうか5段階評価で回答させた(表3)。概ね肯定的な回答が得られたといえる。なお、“3DCG”と“基礎統計学”のどちらの授業においても1度も機能を使用しなかった学習者は“また受講したい”と回答しており、さらにその理由を訊ねたところ“(チュータリング機能は)助かる、向こうから呼びかけがあってもいいかもしれないと思った。”と回答した。このことから、たとえ機能を使用していなくとも、リアルタイムでチューターが対応するという環境が用意されていること自体が、学習者にとってよい影響を与えるケースがあることが示された。

表3. また受講してみたいか(n=10)

受講し たくない	どちらか うとしたい	どちら もない	どちらか うとしたい	受講 したい
0	1	2	3	4

### 4. まとめと今後の課題

本稿では、オンデマンド授業においてリアルタイムでのチュータリングを実現するシステムを開発した。今後の課題として(1)長期運用による評価、(2)チューターの対応の検討が挙げられる。

【参考文献】  
鈴木雅也(2006)講義中に教師が参考にする学生の非言語的行動・感情の分析, 東京理科大学工学部卒業論文  
西久保健太, 赤倉貴子(2007)学習者が反応を発信できるVOD型e-Learning Systemの検討, 電子情報通信学会総合大会 情報システムソサエティ総合大会特別号, pp.137

キーワード：e テスティング、適応型テスト、テスト・マネジメント・システム、真正な評価

## 1. はじめに

Web 上でテストを実施することを「e テスティング (e-Testing)」と呼び、近年、実用化が進むとともに注目されてきている<sup>(1)</sup>。ただし、e テスティングはオンライン・テスト(Online Testing)<sup>(2)(3)</sup>やオンライン・クイズ(Online Quiz)<sup>(2)(3)</sup>、IBT(Internet Based Testing)<sup>(4)</sup>と呼ばれることもある。

e テスティングが普及する前、パーソナルコンピュータが一般に普及し始めた 1980 年代からコンピュータ上でテストを行うことはすでに実用化していたが、この時代はインターネットが普及していなかったので、スタンドアロン型コンピュータ上でテストが実施され、このようなテスト形態を一般に“CBT(Computer Based Testing)”と呼んできた<sup>(5)(6)</sup>。当時は、フロッピーディスクなどの物理媒体にテスト実施システムと項目ファイルをコピーして学習者に配布し、学習者の回答をそれらの媒体に電子的に記録してテスト終了後に回収するという方法をとっていた。1990 年代よりインターネットが普及し、テストの配信、回収もインターネット上で行えるテスト・システムの開発が盛んになった。さらにその少し後に e ラーニングが大きく注目され始めたことにより、e テスティングは e ラーニングのツールとしても注目され始め、急速に実用化してきた。

CBT が開発され始めた 1980 年代—1990 年代冒頭までは、CBT の欠点としてコンピュータ・リテラシーのない受検者が不利益を被ることが懸念されていた。例えば、赤倉、永岡(1993)<sup>(7)</sup>では CBT を受けた場合とペーパー・テストを受けた場合の点数差がコンピュータへの興味と相関があることを示している。その後、

1990 年代後半よりコンピュータの一般家庭への普及が進み、上のような差が減少したため、大きな問題として捉えられなくなってきた。このことも e テスティングへの普及を促進させたと考えられる。

例えば、TOEFL や TOEIC で有名なアメリカの ETS(Educational Testing Service)も 2006 年でこれまで実施してきた CBT を全廃し、Web 上で受検する e テスティングに完全移行を行っている。

このように e テスティングは e ラーニングより古くから独立に発展してきた歴史を持ち、最近では以下のような様々な専門分野を統合した分野となっている。

- (1)心理学：評価理論、テスト理論
- (2)統計学：項目反応理論の推定
- (3)計算機科学：テスト構成・出題の高速アルゴリズム
- (4)人工知能：自動作問や論述式項目の自動採点
- (5)データベース：アイテム・バンクの構築
- (6)教育工学：e ラーニング、e ポートフォリオとの融合
- (7)ヒューマン・インターフェース：テスト画面の設計、障害者などへの適応的インターフェースの設計

テスト理論は統計学と心理学(計量心理学)の融合分野であり、CBT の開発はテスト理論研究者を中心に進められてきた。近年、コンピュータ・サイエンスの研究者が参入し、Web 技術のみならず、最新計算アルゴリズムや AI 技術が実装されるようになり、「e テスティング」という名称が用いられるようになってきた。

e テスティングは従来のペーパー・テストにはない以下のような利点を持つ<sup>(1)</sup>。

1. マルチメディア技術を用いることにより、ペーパーテストでは実現できなかった真正な評価環境を実現できるようになる。

テスト採点が、自動化もしくは半自動化される。小論文方式のテストでさえ、自動採点できる。さらに即時的にテスト結果をフィードバックできる。

---

\* 電気通信大学 大学院情報システム学研究科  
Graduate School of Information Systems,  
The University of Electro-Communications

1. アイテム・バンクの領域からランダムに項目の抽出が可能であり、テスト出題バイアスを減少できる。
2. コンピュータによるテスト構成支援機能は、信頼性と妥当性の高いテスト構成を可能にする。
3. コンピュータによる自動テスト構成により、等質のテストを複数自動構成することができる。
4. 適応型テストなど受検者の能力を随時推定しながらその能力に最適な項目を選択出題するような動的テストを実現し、テスト時間を短縮し、測定精度も向上させることができる。
5. 様々なテスト目的に応じたテスト出題方略を持つことにより、目的に応じたテスト構成や適応型テストを実現できる。
6. 回答所要時間や回答書き直し回数、回答見直し回数などペーパー・テストでは測定することができなかったデータが測定することができるようになり、質問項目評価の多様性をもたらし、補助的に用いることにより測定精度を向上させることができる。
7. 障害を持つ受検者への対策が容易に行える。例えば、字を大きくしたり、ヘッドホンをつけさせて音を大きくしたり、特別な入力装置（手以外で入力できる入力装置）、出力装置（点字など）を用いることにより障害者の受検を可能にすることができる。
8. 遠隔地にいる受検者の評価を行うことができる。
9. 遠隔地に分散するテスト作成者が、共同でアイテム・バンクを構築し、テスト構成を行うことができる。
10. ペーパー・テストに比較し、テストの印刷、運搬のコストが減少する。
11. 大量のペーパー・テストの束を管理することに比較し、電子ファイル上のテストは管理が容易でセキュリティも高い。
12. テスト実施に際して、テストの配布、回答の回収が自動的にできる。

本論文では、e テスティングの最新の理論と技術について紹介する。2章ではe テスティングの基礎となる数

学モデルである項目反応理論を紹介し、3章ではe テスティングの情報システムである TMS (テストニング・マネジメント・システム)の構成と自動テスト構成や適応型テストなどの各モジュールの最新研究をレビューする。さらに4章では自動作問技術、5章ではe ラーニングやe ポートフォリオにおけるe テスティング技術、6章ではe テスティングにおける標準化の研究動向をレビューし、7章で筆者らが開発している IPA 情報処理技術者試験の大規模 TMS を紹介する。

## 2. 項目反応理論 (IRT)

e テスティングの発展は、受検者の能力潜在変数と各項目への正答確率を数学モデルで定義した項目反応理論 (Item Response Theory: IRT) (例えば、文献(8)参照) が寄与する部分が多い。IRT は以下のような利点を持つ。

- IRT を用いて推定される能力推定値は、正答数などの尺度に対して信頼性が高い。
- 異なるテストを受けた複数の集団を同一尺度上で評価することができる。
- IRT を用いて各項目の特性を示すパラメータが独立に得られることにより、アイテム・バンクのメタ・データとして適している。
- テスト・データで欠側データがある場合にもデータを容易に扱うことができる。
- 受検者の出題項目への反応履歴から次に出題する項目を適応的に選択する適応型テストの基本原理になる。

IRT には様々なモデルがあるが、ここでは最も一般的な2パラメータ・ロジスティック・モデルを紹介する。このモデルは*i*番目の受検者の能力母数を $\theta_i$ として、*j*番目の項目へ正答する確率を以下のように仮定する。

$$P(x_{ij} = 1 | \theta_i) = \frac{1}{1 + \exp(-1.7a_j(\theta_i - b_j))}$$

ただし、 $x_{ij} = \begin{cases} 1: \text{被験者 } i \text{ が項目 } j \text{ に正答} \\ 0: \text{被験者 } i \text{ が項目 } j \text{ に誤答} \end{cases}$

パラメータ  $a_j$  は反応曲線で  $P(x_{ij} = 1 | \theta_i) = 0.5$  と

なる点での反応曲線の傾きを示しており、項目 *j* が受検者の能力をどれほど識別できるかを示す



パラメータと解釈される。パラメータ  $b_j$  は反応曲線

で  $P(x_{ij} = 1 | \theta_i) = 0.5$  となる能力の値を示しており、

それが大きいほど正答するのに必要な能力が高くなることから難易度パラメータと解釈される。n 人の受検者の m 個の項目を持つテストへの反応データ

$X = \{x_{ij}\}, (i = 1, \dots, n, j = 1, \dots, m)$  を所与としたとき

の尤度は、局所独立構造を仮定し、

$$l(a_j, b_j, \theta_i | X) = \prod_{i=1}^n \prod_{j=1}^m P(x_{ij} = 1 | \theta_i)^{x_{ij}} [1 - P(x_{ij} = 1 | \theta_i)]^{1-x_{ij}}$$

と示される。パラメータ推定手法には、(1)同時推定法、(2)周辺推定法、(3)EM アルゴリズム、(4)ベイズ推定法、(5)MCMC 法、などが挙げられる。前述の CBT は IRT を基礎として統計学、計量心理学を中心に研究が進められてきている。

### 3. TMS(テストィング・マネジメント・システム)

当初開発された CBT は、アイテム・バンクと適応型テストやテスト実施システム、客観式項目への自動採点システム、のみからなる単純な構成であったが、e テスティングではテストィング・マネジメント・システム(Testing Management System: TMS)と呼ばれるテスト構成・実施・採点・フィードバックなどを統合的にマネジメントできる情報システムを用いる。図 1 は、著者らが開発している TMS<sup>(1)(9)(10)(11)</sup> の構成である。以降、各モジュールとその周辺最新研究について解説する。

#### 3.1. アイテム・バンク

アイテム・バンクはコンピュータでテスト項目を管理するための項目データベースであり、当初の CBT の時代より存在した e テスティングの基本的な概念の一つである。アイテム・バンクは、コンピュータ・データベース上に細分化された教育目標と各目標に対応した項目、それに関するテスト実施後の統計データ等を管理する。これにより、能力測定型テスト、選抜型テストではテスト構成時に領域から項目をランダム・サンプリングやより高度な手法を用いて受検者の能力推定の予測効率を向上させたり、診断型テストや形成

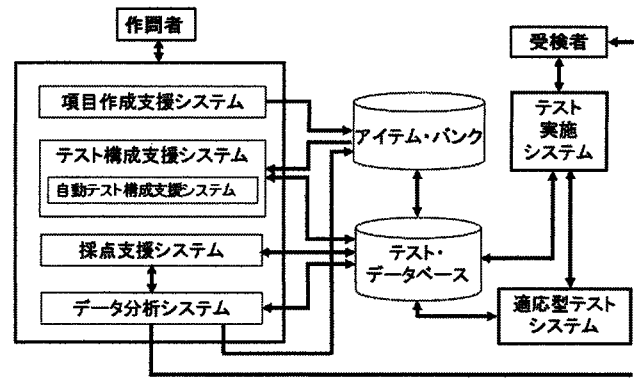


図 1. TMS の構成

- 項目 ID
- 教科名
- 領域目標名
- 目標タイプ (知識、理解、応用、分析、結合、評価)
- 項目出題形式 (多枝選択式項目、真偽式項目、単答式項目、論述式項目)
- 項目メディア (文章、図、音声、ビデオ)
- 項目内容
- 正答もしくは正答例
- 関係する URL
- 下位項目 ID
- 項目正答率
- 項目正答率の標準誤差
- 項目平均所要時間
- 項目所要時間標準誤差
- 項目反応理論におけるパラメータ

図 2. アイテム・バンクの構成例

的テストで、各目標との項目の対応より各学習者がどの目標を達成できていないのかの分析などが容易になる。e ラーニングで用いる場合は、学習者が達成できなかった目標に対応したコンテンツに自動的にジャンプさせたりもできる。

基本的なアイテム・バンクの利点としては、1.項目のアクセス性、2. 共有性、3. 再利用性が挙げられる。1.アクセス性は、コンピュータで項目を管理することにより、簡単に必要なテスト項目を検索できる利点、2.共有性は複数のテスト開発者によってテスト項目を共有することが非常に容易になる利点、3.再利用性は、

大規模なデータベースの場合、良問項目は何度も用いることができる利点を示している。

図 2 に著者らが開発しているアイテム・バンク<sup>(1)(9)(10)(11)</sup>の構成を示す。図中の項目 ID は、項目を識別するために割り当てられる数字やアルファベットによって構成される文字列であり、一般に教科番号、領域番号、目標タイプ番号、出題形式番号、出題メディア番号などを組み合わせて設定されることが多い。また、アイテム・バンクでは、教科名、領域名、目標タイプを記述することが多い。図ではブルームの細目表<sup>(12)</sup>を簡易化して目標タイプを記述している。さらに、項目出題形式(多枝選択式項目、真偽式項目、単答式項目、論述式項目)とマルチメディア・テストにも対応した項目が用いているメディア(文章、図、音声、ビデオ)を記述する。最も重要な項目内容と正答(正答例)の記述は当然として、項目の解説になるような URL を記述しておくことでフィードバックが容易になるし、診断型テストなどで必要となる対象項目の下位項目(その項目に正答しなければ対象項目に正答できない項目で、対象項目に隣接したもの)を記述している。

テスト実施後に得られる項目に関する統計情報も項目改善のための分析材料としてだけでなく、構成されたテストの予測得点や予測時間を計算してテスト構成を支援したり、受検者の能力を逐次推定しながら最も適応した項目を自動出題する適応型テストの実現に用いられるのである。

### 3.2. テスト構成支援システム

実際のテスト構成では、テストの得点予測や所要時間予測が重要となることが多い。例えば以下のような状況が挙げられる<sup>(13)(14)</sup>。

- (1) 資格試験などにおいて、合格点をあらかじめ決めてあるような絶対評価の場合、毎回のテストで合格の難易度が異ならないようにテスト得点分布及び所要時間分布はほぼ一定でなければならない
- (2) 入試問題などの選抜テストにおいて、センター試験の数学や理科などに代表されるように教科選択が行える場合がある。選択した教科によってテストの難易度が異ならないように得点分布及び所要時間分布はほぼ同一でなければならない

- (3) 入試などの選抜テストにおいて、テスト中に受検者が大問を選択できる選択式テストが用いられることがある。この場合、選択した大問によって難易度が異ならないように大問ごとのテスト得点分布及び所要時間分布はほぼ同一でなければならない

上のような場合、作問者は得点分布および所要時間分布を予測しながらテスト構成をしなければならず、多くの場合、過去を参考にしながら経験と勘に頼りながら作問する。しかし、近年、e テスティングの普及に伴い、アイテム・バンク方式のテスト構成が一般化して、テスト得点および時間の予測の実現が可能になりつつある。アイテム・バンク中のすべてのテスト領域に項目が入力されるとテスト構成が可能になるシステムが開発されている<sup>(15)</sup>。

通常、テスト構成はネットワーク上で一人もしくは複数のテスト開発者がテスト構成支援システムを用いて行う。特に、e テスティングではテスト構成がネットワーク上でブラウザを用いてどこからでも行えることが特徴である<sup>(16)</sup>。Songmuang and Ueno<sup>(16)</sup>は、協働でテスト構成する場合に協働者数が構成される妥当性を増加させるが、一方で信頼性を減少させることが多いことを実験で示した上で、協働でのテスト構成でのアドバイスをを行う知的エージェントを持つ協働テスト構成支援システムを開発している。

さらに最近ではテストを自動的に複数生成するための研究も進んできている。たとえば、文献(16)(17)は、貪欲アルゴリズム(Greedy Algorithm)を用いてテストの制約付情報量最大化を提案しているし、文献(18)では遺伝的アルゴリズムを用いて同様のことを行っている。テストの複数生成では Armstrong<sup>(19)</sup>らがネットワークフローを用いて信頼性を最大化するテスト生成法を提案しているし、van der Linden<sup>(20)</sup>が、数理計画法を用いて、情報量を等質化するテストを複数生成する Big Shadow 手法を提案している。しかし、これらの手法は、平均得点や所要時間の等質性を無視しており等質テストといえるものではなく、計算時間も膨大で、生成されるテスト数を最大化できないという問題点があった。この問題を解決するために、著者らは数理計画法に分枝限定法 (Branch and Bound, BB)<sup>(21)</sup>を用いて同一のテスト領域、テスト情報量、テスト得点分

布、テスト所要時間分布を持つ等質テストをアイテム・バンクより自動的に最大数構成するシステムを開発している。最新研究では、石井ら<sup>(22)</sup>が、最大クリーク法を用いてより複雑な制約条件のもと等質のテストを高速でアイテム・バンクより最大数自動的に構成するアルゴリズムを提案し、システム開発を行っている。この手法では、等質テストの生成数を最大にできることが数学的にも証明されている。

以上のように、従来のテスト理論のみでなくコンピュータ・サイエンスの最先端技術がeテストング技術を支えていることがわかるであろう。

### 3.3. テスト実施システム

テストデータベースにテスト情報が格納されるといつでもテストの配信が可能になる。テストデータベースシステムでは、テストの質問項目 ID 集合とともに配信に関する期間や配信先の制約なども書き込まれている。テスト実施システムは、テストデータベースの情報を読み取り、アイテム・バンクよりテスト期間や受検者制約に応じて質問項目を抽出してテストを配信するシステムである。さらにeテストングでは、ビデオや音声を用いた質問項目を配信することができ、これまでペーパー・テストでは実現できなかったような真正な評価環境を実現することができる。例えば、語学テストでは動詞の概念などはビデオを用いて表現すると非常に明確になるし、科学実験や音楽などもビデオを用いることにより多様な評価が行えるようになっている。受検者はテストデータベースに格納されたテスト順序に従ったテスト項目を順に受けることができ、先に飛ばしたり、あとに戻ったり、回答を書き直したりもできる。テスト実施システムは、受検者の項目への回答を逐次、テストデータベースに保存するだけでなく、所要時間や見直し回数、回答の書き直しプロセスなど詳細な回答履歴データを蓄積していく。このことも、これまで正誤データのみを情報としてきたペーパー・テストとは異なり、これまで得られなかったような回答履歴データより多様な評価を実現できるのである。

### 3.4. 採点支援システム (自動採点システム)

テスト項目は大きく、多枝選択式項目、真偽式項目、

単答式項目、論述式項目に分類できる。当初、コンピュータの得意な自動採点は、多枝選択式項目、真偽式項目のみで行われてきた。しかし、近年、自然言語処理の技術が発達し、単答式項目や論述式項目の自動採点が導入されつつある。例えば、ETSでは英語小論文の入試採点のための自動採点システム“e-rater”が開発されている<sup>(23)</sup>。このシステムでは、人間の採点を目的関数とする重回帰分析に評価観点「構造」、「組織化」、「内容」から得られる57の説明変量として評価得点を求めている。Burstein and Wolskaは、e-raterと専門家との採点の一致率が89%であると報告している<sup>(23)</sup>。日本でも、石岡らによって日本語論文の自動採点システム“JESS”が開発されている<sup>(24)</sup>。最近出版された自動エッセイ採点のためのテキストでは、洋書では(25)、和書では(26)があるので、詳細はそれらに譲る。

### 3.5. データ分析システム

テスト終了後、各学習者の各項目への正誤データ回答所要時間、回答書き直し回数、回答見直し回数などがテストデータベースに書き込まれる。データ解析システムは、これらのデータを解析し、即時的にフィードバックを返すシステムである。フィードバックは、(1)受検者、(2)テスト実施者、(3)アイテム・バンク、の三方向へ行われる。(1)受検者には得点データや偏差値、診断結果などテスト目的に応じたフィードバック、(2)テスト実施者には(1)のフィードバックに加えて、テスト自体の信頼性などを計算した指標や項目別の情報をフィードバックし、(3)アイテム・バンクには項目に関する様々な統計解析の結果が自動的に書き込まれる。

eテストングの特徴はこれまでペーパー・テストでは測定できなかったようなデータ(例えば、回答所要時間データや回答書き直し回数など)が容易に得られるようになったことが挙げられる。これらのデータは、これまでではできなかった項目分析や受検者分析が期待される。しかし、現状はこれらのデータをどのように有効活用するかについて十分に研究・活用されているとはいえず、早急に研究すべき課題であろう。

これまでに、レスポンス・アナライザやコンピュータ・テストングより得られる回答所要時間データ

の解析手法として集団学習応答曲線 (Response Curve)<sup>(27)</sup>は古くから研究されている。集団学習応答曲線とは、学習者集団に演習問題などの課題を与え、課題終了までの時間、すなわち回答所要時間を何らかの手段で測定し、時間経過に沿って回答率をグラフに描くというものであり、様々な数理的手法が適用されてきた。これらのうち、藤田 (広) らにより初期に研究されたもの<sup>(28)</sup>としては、実際の授業中においてレスポンス・アナライザーにより得られる回答所要時間データへの寿命曲線・ワイブル分布のあてはめが知られている。また、その後、藤田 (恵)、成瀬によって回答所要時間への McGill 分布のあてはめ<sup>(29)</sup>が提案されている。しかし、ワイブル分布や McGill 分布の教育分野での所要時間分布へのあてはめには以下のような問題が存在した。

- 1) パラメータ推定が比較的難しく、リアルタイムでの計算には時間コストが高い
- 2) パラメータの解釈が難しく、特に教育分野で得られる所要時間分布として捉えた場合のパラメータ解釈が経験的である

以上の問題を考慮し、永岡ら<sup>(30)</sup>は、所要時間分布としてガンマ分布を提案している。このアプローチでは、それまで集団学習応答曲線が、実際の授業中にレスポンス・アナライザーより得られる反応時間データを対象としていたのに対して、各学習者が様々な場所や時間で受験したコンピュータ・テストにおける所要時間を集積して解析するといった方法がとられていたことも特徴であった。また、このアプローチはパラメータ推定が非常に容易であり、ガンマ分布中のパラメータ  $\alpha$  については、数学的な意味づけから非常に明快な解釈を得ることができるという利点もあった。ただし、ガンマ分布は2つのパラメータにより構成されており、もうひとつのパラメータの解釈については明確な解釈を得ることができておらず、また実用的な解釈にも用いられてこなかったという問題もある。教育分野の数理モデルによるデータ解析アプローチを、大量にあるデータを少数のパラメータに圧縮して有用な意味を引き出すデータ圧縮の立場として捕らえると、数理モデル中のすべてのパラメータが目的に対して有意義であることが望ましい。植野、永岡<sup>(31)</sup>は、上の問題を解決し、ガンマ分布におけるパラメータが、「思考

過程の複合度」を示すパラメータ  $\alpha$  と「単純な思考過程の平均所要時間」を示すパラメータ  $\beta$  であり、これらは数学的に陽に導かれる拡張ガンマモデルを導出しており、実際に活用されている。また、データマイニングにおける異常値検出技術の応用として、Ueno<sup>(32)</sup>は受検者の e テスティングにおける回答所要時間データを用いて、受検者の不正行為を検出するシステムを開発している。永岡<sup>(33)</sup>は、CBT において、先駆的に切斷指数分布を用いた予測得点分布とガンマ分布による予測所要時間分布を提示するテスト構成支援システムを開発している。Ueno<sup>(34)</sup>は、インターネット上における e テスティングのアイテム・バンク統合システムを開発し、二項分布、混合二項分布を予測得点分布とする Web ベースのテスト構成支援システムを開発している。一方、テスト得点分布のみに焦点化した研究は、テスト得点分布としてベータ二項分布を当てはめてきた研究が知られている<sup>(35)</sup>が、その他にはほとんど存在しない。それに対し、テストの所要時間分布に関する研究は多い。前述の集団応答曲線の研究のほかにも、Thissen<sup>(36)</sup>が対数正規分布モデルのテスト所要時間データへの当てはめを行っているし、Verhelst ら<sup>(37)</sup>は心理テストにおけるスピードテストにガンマ分布を用い Roskam<sup>(38)</sup>はテスト所要時間にワイブル分布を当てはめている。さらに、正規分布、対数正規分布、ワイブル分布を実データに当てはめて比較した結果、ほとんどの場合、対数正規分布が最もあてはまりがよいことが報告されている<sup>(39)(40)(41)(42)(43)</sup>。最新研究では、ソナムアンらが、上のすべてのモデルを比較して、データへのあてはまりは得点分布では IRT、テスト所要時間モデルでは対数正規分布モデルが最もよく、予測分布としては IRT による得点分布と拡張ガンマ分布による所要時間分布が最も良いことを示し、これらのモデルを搭載したテスト得点・所要時間予測システムを開発している<sup>(44)</sup>。

また、e テスティングにおける回答書き直しプロセスや回答見直しプロセスを解析する手法も提案されており<sup>(44)(45)</sup>、e テスティングでしか得ることのできないさまざまなデータ処理法が提案されてきている。

### 3.6. 適応型テスト

e テスティングの技術で現在、最も注目されている

のが、適応型テスト(CAT: Computerized Adaptive Testing)<sup>(46)</sup>である。適応型テストとは、受検者のテスト項目へのそれまでの回答履歴から逐次的に受検者のIRTにおける能力パラメータを推定しながら、その能力に最も適した項目をアイテム・バンクより抽出し出題するというものである。TOEIC や TOEFL で有名なETS(Educational Testing Service)をはじめ多くのテスト業界で実用化されている。当初は、受検者の能力に最適な項目を出題することにより、能力の推定効率を上げることやテスト採点が瞬時かつ自動的にできることが主な利点とされてきたが、最近では、異なるテストを受検者ごとに自動的に生成した上で同一尺度上で評価できるという利点が注目されてきている。当初の適応型テストのハードウェアは、スタンドアローン型コンピュータを用いて実現されてきたが、近年では高速サーバーを用いたWeb上での適応型テストが主流となりつつあり、より管理コストを減少させている。

適応型テストの歴史は、コンピュータが普及する以前の、ペーパーテストの時代には、能力推定のためのテスト実施フェーズと推定された能力に応じたテスト実施フェーズを持つ二段階テスト<sup>(47)</sup>、項目反応理論を簡易化した近似的ベイジアン項目選択<sup>(48)</sup>、受検者の各質問項目への反応に応じて出題項目の難易度を上下させる項目上下選択法<sup>(49)</sup>、各質問項目への正誤に対応した次に出題する質問項目を構成した樹上図に従う樹状型適応型テスト<sup>(50)</sup>などが用いられていた。

強力なコンピュータが出現してからは、計算量のより多い情報量最大化という出題原理<sup>(51)</sup>に基づく大規模なハイステーク・テスト(入試や検定など結果が受検者に大きく影響するテスト)の適応型テスト・プログラムが実用可能となってきた。この分野の最初のパイオニアはアメリカ国防省の語彙テストASVAB(Armed Services Vocational Aptitude Battery)であろう。なぜならば、大規模なASVABの適応型テストは、1980年代半ばにはすでに動作していたからである。しかし、真の意味で大規模なハイステーク・テストの最初の適応型テストは、アメリカ合衆国やカナダの通常の大学院へ進学へ必要な共通試験であるGRE(Graduate Record Examination)のCATバージョンである。その後も、アメリカ医師資格免許試験の適応型が実施されるなど、ほとんどの大規模なハイステーク

テストCAST(Computer-Adaptive Sequential Testing)

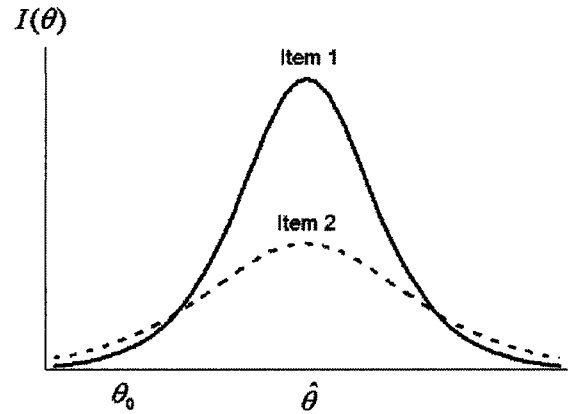


図3. フィッシャー情報量パラドックス

ク・テストは適応型テストへ移行しつつあるのである。

一方、これらのシステムでは未だにフィッシャー情報量が用いられるのであるが、以下のような問題点もある。

1. 識別力の高い項目しか出題されない傾向にあり、アイテム・バンクを効率よく活用できない
2. 適応型テスト開始直後には、受検者の能力推定の誤差が大きいが、誤差の大きい暫定的な能力推定値において情報量を最大にする項目出題は、実は真の能力にとって情報量が小さいことになる。例えば、図3は二つの項目について横軸を能力パラメータ  $\theta$ 、縦軸を各項目のフィッシャー情報量としている。真の能力を  $\theta_0$

とし、能力推定値を  $\hat{\theta}$  としたとき、Item1 と Item2 では識別力の低い Item2 のほうが真の能力に対して情報量が大きくなってしまふ。すなわち、能力の推定誤差の大きい初期には識別力の低い項目を出題するほうが情報量が大きくなる。これをフィッシャー情報量パラドックス<sup>(52)</sup>と呼ぶ。この問題を解決できる手法の一つとして、植野ら<sup>(53)(54)</sup>は、ベイズ統計のアプローチによる意思決定理論において定式化されている「データの価値(EVSI: Expected Value of Sample Information)の枠組みを用いて  $\theta$  の分布について周辺化する項目情報量を提案している。Chang and Ying は、受検者の最初から  $k-1$  番目の項目への反応から推定された能力パラメータ推定値  $\hat{\theta}_{k-1}$  の尤度と真

の能力パラメータ  $\theta_0$  の尤度とのカルバック・ライブラー情報量の期待値を  $\theta$  の分布について周辺化する項目情報量を提案している<sup>(55)</sup>。Veerkamp and Bergerは、尤度重みフィッシャー情報量を提案している<sup>(56)</sup>。さらにVan der Lindenは、完全ベイズ情報量<sup>(57)</sup>を提案し、ベイズ事後分布重み情報量の期待値として定式化している。これらはいずれも情報量を  $\theta$  推定値の経験分布で周辺化することで  $\theta$  推定値の誤差を考慮した情報量を提供するものである。そのため、これらの基準を用いた適応型テストの予測精度は大きく改善されるのであるが、すべて  $\theta$  を広範囲にわたって数値積分しなければならず、計算時間が非現実的にかかり実用に至っていない。そのため未だに現実の適応型テストではフィッシャー情報量やオーウェンの近似的ベイジアン項目選択が用いられてきているのである。筆者らが開発している後述の e テスティング・システムでは、これらの情報量と受検者の反応を元に木構造を生成し、分枝限定法 (Branch and Bound、BB)<sup>(21)</sup>を用いて高速に適応項目を選択する手法を開発している。

#### 4. 自動作問技術

テストにおいて最も大きなコストがかかるのが作問である。近年、注目される e テスティングの技術の中で自動作問技術<sup>(58)</sup>がある。自動作問は、単に問題を作るのではなく、問題のメタデータ (例えば、単なる足し算の計算問題を作る場合、足し算の桁数、繰り上がりの有無、など)をも自動作成する試みである。平嶋<sup>(59)</sup>は、「問いに答えるために必要な情報：所与情報」、「問いに答えを導く過程で得られる情報：中間情報」、「問いの答えとなる情報：解情報」を定義した上で、問題を「所与情報+問い」、問題解決過程を「所与情報から解情報を導く過程」と定義し、所与情報、中間情報の中での情報間の関係を「制約構造」と呼んでいる。自動作問技術は、(1)テキスト中に空欄を設けるテキスト空欄化の方法<sup>(60)</sup>、(2)制約構造を満たすエピソードを生成する制約ベースの方法<sup>(60)(61)</sup>、(3)問題解決過程を部分化・特殊化する問題解決ベースの方法<sup>(62)(63)(64)(65)(66)</sup>、(4)選択枝問題の誤答選択枝の自動生成<sup>(66)(67)(68)(69)</sup>に分類できる。主に人工知能分野で研究されており、これからの進展が期待される分野である。

#### 5. e ラーニングにおける e テスティング

e テスティングは、心理測定分野や統計学を中心に発達してきたが、2000 年代以降急速に普及してきた e ラーニングの評価システムとしても注目されるようになってきている<sup>(1)(70)</sup>。e ラーニングとは、学習者が Web 上で相互に協力しながらコンテンツを学習したり、e テスティングで演習しながら学習する仕組みである。このような新しい研究分野は、これまでのテスト研究ソフトウェアとは独立に教育工学、認知科学、コンピュータ・サイエンス分野、経営学分野を中心に広がってきたといえる。

e ラーニング分野では、知識とは真のものが存在するわけではなく、個々の人間がそれぞれの中で知識を構成していくと考える「構成主義」の立場をとり、真の知識を学習者に一方的に注入しようとする「教授主義」から学習者自身の自律的学習 (知識構成) を支援する「学習者中心主義」への移行を狙っている。そのために、学習評価の考え方も「真正な評価 (Authentic Assessment)」と呼ばれる、真正な文脈に基づく、現実的な課題解決による学習評価が中心となってきた。レポートなどの学習成果物 (もしくは、より広義にテスト結果や学習履歴データなどを含めて「学習ポートフォリオ」<sup>(71)</sup>とも呼ばれる) を自己評価したり、学習者同士で評価しあう相互評価が中心となる。このような場合にもテストや演習問題は e ラーニングに埋め込まれている。

しかし、前節で述べたような大量のアイテム・バンクを担当教員が作成するのは大変な労力となることが多い。片手間で作成された質問項目や演習問題は不完全なものが多く、信頼性や妥当性のないものを作成してしまう。そこで、e ラーニングでは学習者自身に質問項目を作問させて実際にテストや演習問題に用いる方法が注目されている。例えば、文献<sup>(72)(73)(74)(75)(76)</sup>では「問題作り」が問題解決能力や問題分類能力を向上させ、学習者のメタ知識を発達させる働きがあると示している。e ラーニングの各コンテンツごとに各学習者に質問項目を作成させ、それらを集積し、テストを実施して評価を行い、良問をフィルタリングしてアイテム・バンクを充実させる方法は、学習者主体の e ラーニングにおける e テスティングとして非常に注目されている<sup>(69)</sup>。

## 6. 標準化技術

一般に、e テスティングなど情報処理の分野における標準規格は扱う対象によって二種類に大別できる<sup>(77)</sup>。第一の分類は、装置やシステムそのものに関するもので、システム間の互換性・相互運用性を狙いとしたものである。TCP/IP, HTML などはこの分類に即する規格の例であり、e ラーニングの分野では WBT (Web-based Training) コンテンツの標準規格 SCORM (Sharable Content Object Reference Model) (Advanced Distributed Learning Initiative 2006) が代表例である。第二の分類は、システムやサービスの開発・運用のマネジメントに関するものである。マネジメント規格の代表的な例としては ISO 9000 や ISO/IEC27001 が挙げられる。ISO 9000 はよく知られているように品質管理に関する規格、ISO/IEC 27001 は情報セキュリティに関する規格である。これらはいずれも装置やシステム自体に関する規格ではなく、企業などが安定的に品質を保って製品を供給する仕組みや、情報セキュリティを確保する仕組み(マネジメントシステム)を確立し、維持・改善していくための要求事項を定めたものとなっている。e ラーニングの分野では ISO/IEC SC36 で e ラーニングの開発プロセスにおける品質保証に関する規格が策定されている<sup>(78)</sup>。

e テスティングにおいても上記の二種の分類に属する標準規格が存在する。前者に属するものとして QTI (Question and Test Interoperability) 規格<sup>(79)</sup>、後者に属するものとして「アセスメント配信における情報技術 (IT) 利用の実践のための規範 (A code of practice for the use of information technology (IT) in the delivery of assessments)」規格(以下、delivery of assessments の頭文字を取って DoA 規格と呼ぶ)<sup>(80)</sup>を挙げることができる。

前者の QTI 規格は、項目や評価結果データの標準化に関するものである。e テスティングの導入により、項目の電子化が可能となり、過去に作成した項目や他組織で作成した項目を含めて多くの項目が再利用可能になるというメリットが考えられる。また、テストの実施結果も電子化されて蓄積が容易になり、学習者特性の把握、設問特性の把握が過去の実施結果も含めて容易に行えるようになる点も e テスティングのメリットである。しかし、このようなメリットを十分に活か

すためには、項目、評価結果データが e テスティング・システム間で流通・再利用可能なようにデータフォーマットを標準化しておく必要がある。QTI 規格はこのような項目や評価結果データの標準化を目指したものであり、QTI 規格に準拠した e テスティング・システムの実装も行われている<sup>(81)</sup>。

一方、後者の DoA 規格は、e テスティングの実施・運営に関するものである。e テスティングの実施においては、従来のペーパー・テストにおける留意事項と同様、試験が公正・公平に実施されること、不正行為を防止すること、などが重要な要件になる。また、受検者の個人情報や問題・回答のセキュリティが保護されること、受検者の IT 習熟度が試験結果に影響を与えないことも重要な要件である。DoA 規格は、このような要件を満足して、品質の高い e テスティングをスムーズに実施するために留意すべき事項をガイドラインとして示したものである。

## 7. 情報処理技術者試験での開発例

本章では e テスティングの情報システムの一例として筆者らが開発している IPA 情報処理技術者試験用 TMS を紹介する。このシステムは、3章で紹介されたすべてのモジュール、すなわち図1の構成を持つTMSである。本システムの特徴は、1. 応理論のパラメータ推定に解析アルゴリズムを搭載し、高速に自動的に図4の3000を超える項目の属性を自動更新することができること、2. 正答率、平均所要時間や情報量、領域ごとの最小出題数などを図5の自動テスト構成システムに入力すると、最大クリーク抽出法<sup>(21)</sup>に分枝限定法 (Branch and Bound、BB)<sup>(20)</sup>を用いて同一のテスト領域、テスト情報量、テスト得点分布、テスト所要時間分布、信頼性を持つ等質テストを重複率 $\alpha$ %の制約条件でアイテム・バンクより自動的に最大数生成することができること、3. 適応型テストの完全情報量を最大化する木生成アルゴリズム<sup>(2)</sup>を提案し、テスト領域の制約を満たしながらフィッシャー情報量パラドックスを解決した完全情報量を最大化する項目を高速に選択し、オンラインで出題できる適応型テストの実相、4. 本論3.5で紹介された様々なデータ解析手法の実装、が挙げられる。世界的にも最も進んだ技術を持つTMSであり、2011年からの情報処理技術者機構IPAでの実用化が計画されてい

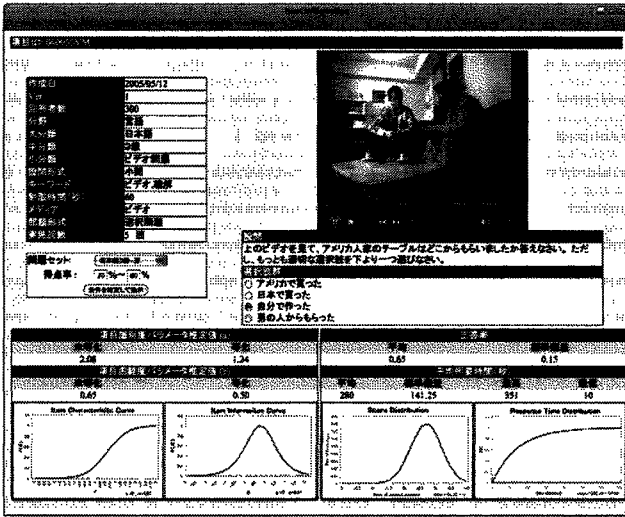


図4. アイテム・バンクの開発例

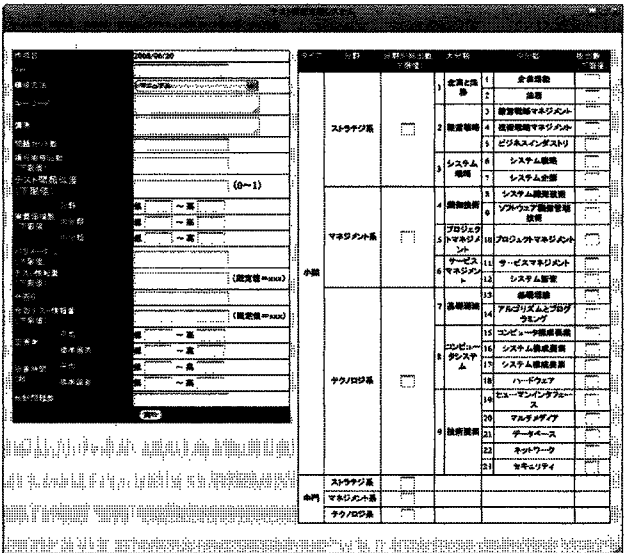


図5. 自動テスト構成システム

る。テスト理論だけでなく、コンピュータ・サイエンスの最新技術が多く利用されていることが特徴である。

### 8. おわりに

本論では、近年、注目されるeテストの理論および技術について紹介を行った。心理学、統計学を中心に発展してきたテスト理論にコンピュータ・サイエンス分野の最新計算アルゴリズム技術が加わることにより、eテスト技術は誕生したといえる。さらに論述自動採点技術や自動作問技術など人工知能分野での研究も盛んになりつつある。また、テストイン

グ技術は、従来から教育分野だけでなく、アンケートなどの調査を主とするマーケティング分野や社会学の分野でも基礎をなしてきた経緯があり、インパクトは教育分野に留まらない。今後、eテストの分野は多くの分野と融合していくことにより拡大していくと考えられ、産学ともにマーケットとしての将来は明るいと考えられよう。

最後に、本論で挙げたeテスト技術をより詳細に知りたい読者は、文献(1)を参照してほしい。

### 参考文献

- (1) 植野真臣、永岡慶三 編著：“eテスト”，倍風館、東京 (2007)
- (2) Hricko, M. and Howell, S. L. : "Online Assessment and Measurement: Foundations and Challenges", Information Science Publishing, Hershey (2006)
- (3) Williams, D. D. , Howell, S. L. , and Hricko, M. : "Online Assessment and Measurement: Emerging Practices", Information Science Publishing, Hershey (2006)
- (4) Bartram, D. and Hambleton, R. K. : "Computer-Based Testing and the Internet", John Wiley & Sons, Ltd, New Jersey (2006)
- (5) Mills, C. N. , Potenza, M. T. , Fremer, J. J. , and Ward, W. C. : "Computer-Based Testing: Building the foundation for future assessment", Lawrence Erlbaum Associates, Publishers, New Jersey (2002)
- (6) Parshall, C. G. , Spray, J. A. , Kalohn, J. C. , and Davey, T. : "Practical Considerations in Computer-Based Testing", Springer-Verlag (2002)
- (7) 赤倉貴子、永岡慶三：“学習者特性とコンピュータテストの関心の数量化分析に基づくテスト実施方法の検討”，日本教育工学会論文誌, 16(4), 223-231(1993)
- (8) Baker, F and Kim, S. H. : "Item Response Theory", Marcel Dekker Ltd, New York (2004)
- (9) Ueno, M. : "Web based computerized testing system for distance education", Educational Technology Research, 28, 59-69 (2005)
- (10) ソンムアン・ポクボン、植野真臣：“統合型eテストの開発と実践”，日本テスト学会誌, Vol. 4, No. 1, pp. 54-64 (2008)
- (11) ソンムアン・ポクボン、植野真臣：“eテスト



- における得点・時間予測システムの開発”、電子情報通信学会論文誌、Vol. J91-D, No. 9, 2225-2235 (2008)
- (12)ブルーム, B. S. (梶田 叡一訳, 渋谷 憲一訳, 藤田 恵壘訳): “教育評価法ハンドブック: 教科学習の形成的評価と総括的評価”、第一法規出版、東京、(1973)
- (13)植野真臣: “テスト理論入門”, 大学入試研究の動向, vol. 23, . 63-70 (2006)
- (14) 植野真臣: “大学入試センター試験「情報関係基礎」の統計分析,” 大学入試センター研究紀要, vol. 36, 71-99 (2007)
- (15) Songmuang, P and Ueno, M. : “Collaborative e-Test construction using predicted response-time and score distributions to improve reliability”, Knowledge Management for Educational Innovation, Springer Boston (2007)
- (16) Luecht, R. D. : “Computer-assisted test assembly using optimization heuristics”, *Applied Psychological Measurement*, **22**, 224-236, (1998)
- (17) Sanders, P. F. and Verschoor, A. J. : “Parallel test construction using classical item parameters”, *Applied Psychological Measurement*, **22**, 212-223 (1998)
- (18) Swanson, L. and Stocking, L. : “A model and heuristic for solving very large item selection problems”, *Applied Psychological Measurement*, **17**, 151-166, (1993)
- (19) Armstrong, R. D., Jones, D. H. and Wang, Z. : “Automated Parallel Test Construction Using Classical Test Theory”, *Journal of Educational and Behavioral Statistics*, **19**, 73-90 (1994)
- (20) van der Linden, W. J. : “Linear Models for Optimal Test Designs”, *Statistics for Social and Behavioral Science*, Springer Verlag (2005)
- (21) Land, A. H. and Doig, A. G. : “An automatic method of solving discrete programming problems”, *Econometrica* **28**(3): 497-520 (1960)
- (22) 石井隆稔、植野真臣、 “最大クリーク抽出法を用いた等質テスト生成数の最大化”、JSISE 研究会論文集 (2009)
- (23) Burstein, J. and Wolska, M. : “Toward evaluation of writing style: Finding overly repetitive word use in student essays”, In Proceedings of the 11<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary (2003)
- (24) 石岡恒憲: “記述式テストにおける自動採点システムの最新動向”、行動計量学会、**31**(2), 67-87 (2004)
- (25) Williamson, D. M. , Mislevy, R. J. , and Bejar, I. I. : “Automated Scoring of Complex Tasks in Computer-Based Testing”, Lawrence Erlbaum Associates, Publishers, New Jersey (2006)
- (26) 石岡恒憲: “論述式項目の自動採点”、植野真臣、永岡慶三 編著、”e テスティング”(pp. 95-118), 培風館、東京 (2009)
- (27) 永岡慶三 : “集団学習応答曲線”、教育工学辞典、実教出版、265-266 (2003)
- (28) 藤田広一、永岡慶三、長谷川周、佐藤隆弘 : “コンピュータによるアナライザーのデータ処理 (口) ”、電子情報通信学会教育技術研究会資料、12-20, (1973)
- (29) 藤田恵壘、成瀬正行 : “テスト項目の反応時間-正誤および自信との関連について”、教育工学雑誌、**1**(1). 12-24 (1976)
- (30) 永岡慶三、呉亜棟 : “コンピュータ・テストングにおける回答所要時間についての分析”、日本教育工学雑誌 **12**(4), 129-137, (1989)
- (31) 植野真臣、永岡慶三 : “ガンマ分布による e ラーニング所要時間データのオンライン解析”、日本教育工学会論文誌、**29**(2), 107-117 (2005)
- (32) Ueno, M. : “System for online detection in aberrant responses in e-testing”, Proc. Eighth IEEE International Conference on Advanced Learning Technologies, 824-828, Spain, (2008)
- (33) 永岡慶三 : “予測機能を有する実用コンピュータ・テスト・システムの開発研究”、日本教育工学会論文誌、**24**(1), 63-72. (2000)
- (34) Ueno, M. : “Web-based computerized testing system for distance education”, *Educational Technology Research*, vol. **28**, 59-69, (2005)
- (35) Keats, J. A. , and Lord, F. M. : “A theoretical distribution for mental test scores”, *Psychometrika*, vol. **27**, 59-72 (1962)
- (36) Thissen, D. : “Timed testing: An approach using item response theory, in New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing”, ed. D. J. Weiss, Academic Press, New York, 179-203 (1983)
- (37) Verhelst, N. D. , Verstralen, H. H. F. M. , and Jansen, M. G. H. : “A logistic model for time limit tests, in Handbook of modern item response theory”, eds. W. J. van der Linden and R. K. Hambleton, Springer-Verlag, New York, 169-185

- (1997)
- (38)Roskam, E. E. ."Models for speed and time-limit test", in Handbook of modern item response theory, eds. W. J. van der Linden and R. K. Hambleton, Springer-Verlag, New York, 187-208 (1997)
- (39) Schnipke, D. L. ." Assessing speededness in computer-based tests using item response times", Paper presented at the meeting of the National Council on Measurement in Education, San Francisco (1995)
- (40) Schnipke, D. L. , and Pashley, P. J. ."Assessing subgroup differences in item response times", Proc. the Annual Meeting of the National Council on Measurement in Education Chicago, IL. (1997)
- (41)Schnipke, D. L. , and Scrams, D. J. ."Representing response-time information in item banks", in LSAC Computerized Testing Report 97-09, Law School Admission Council, Newtown, Pennsylvania, (1999)
- (42)Scrams, D. J. , and Schnipke, D. L. ."Making use of response times in standardized tests: Are accuracy and speed measuring the same thing?," Proc. the Annual Meeting of the National Council on Measurement in Education Chicago, IL. (1997)
- (43)van der Linden, W. J. .;"A lognormal model for response times on test items," Journal of Educational and Behavioral Statistics, vol. 31, 181-204, (2006)
- (44)植野真臣、永岡慶三："ベイジック的アプローチによる回答選択過程モデル"、信学技報、ET92-66 (1992)
- (45)植野真臣、永岡慶三："多肢選択式項目における回答選択過程モデル"、電子情報通信学会論文誌(A)VolJ76-A. No. 5. pp. 752-758 (1993)
- (46) 植野真臣："適応型テスト"、植野真臣、永岡慶三編著、"e テスティング"(pp. 49-73), 培風館、東京 (2009)
- (47) Cronbach, L. J. and Gleser, G. C. : Psychological test and personnel decisions (2nd Ed), University of Illinois Press, Urbana: (1965)
- (48)Owen, R. J. ."A Bayesian approach to tailored testing", (Research Report 69-92) Princeton, NJ: Educational Testing Service (1969)
- (49) Lord, F. M. ."Some test theory for tailored testing". In W. H. Holtzman (Ed. ), Computer assisted instruction, testing, and guidance (pp. 139-183). Harper and Row, New York, (1970)
- (50) Larkin, K. C. and Weiss, D. J.:"An empirical comparison of two-stage and pyramidal adaptive ability testing", ( Research Report, 75-1). Minneapolis: Psychometrics Methods Program, Department of Psychology, University of Minnesota (1975)
- (51) Birnbaum, A. ."Some latent trait models and their use in inferring an examinee's ability", In F. M. Lord and M. R. Novick, Statistical theories of mental test scores, (pp. 397-479), Reading MA, Addison-Wesley (1968)
- (52) van der LINDEN, W. J. and Glas, C. A. W. : "Computer Adaptive Testing: Theory and Practice". The Netherlands: Kluwer Academic Press, Dordrecht, (2000)
- (53) 植野真臣、大西仁、繁樹算男、確率ネットワークを組み込んだテスト理論の提案、電子情報通信学会論文誌(A)、Vol. J76-A, No. 5, 752-758, (1994)
- (54)植野真臣、大西仁、繁樹算男、教師の領域構造知識を利用した適応型テストの開発、日本教育工学雑誌、18(1)、15-24, (1994)
- (55) Chang, H. H. and Ying, Z. ."A global information approach to computerized adaptive testing", *Applied Psychological Measurement*, 6, 431-444 (1996)
- (56) Veerkamp, W. J. J. and Berger, M. P. F. ."Item-selection criteria for adaptive testing", *Journal of Educational and Behavioral Statistics*, 22, 203-226, (1997)
- (57) van der Linden, W. J. ."Bayesian item-selection criteria for adaptive testing", *Psychometrika*, 62, 201-216, (1998)
- (58) 平嶋 宗："自動作問技術"、植野真臣、永岡慶三 編著、"e テスティング"(pp. 121-141), 培風館、東京(2009)
- (59) 國近 秀信, 花多山 知希, 平嶋 宗, 竹内 章:"英語長文読解学習のための質問自動生成機能", 電子情報通信学会論文誌 D-I, Vol. J83-D-I, No. 6, 702-709 (2000)
- (60) Martin, B. and Mitrovic, A. : "Automatic Problem Generation in Constraint-Based Tutors", Proc. the 6th International Conference on Intelligent Tutoring Systems: ITS2002, 388 – 398 (2002)
- (61) Le, N. T. and Menzel, W. ."Constraint-based problem generation for a self-assessment system", In ICCE2006 Workshop Proceedings of Problem-Authoring, -Generation and -Posing in a Computer-Based Learning Environment, the 14th International Conference on Computers in Education, 33 – 40 (2006)

- (62) 平嶋 宗, 中村 祐一, 池田 満, 溝口 理一郎, 豊田 順一: "ITSを指向した問題解決モデルMIPS", 人工知能学会学会誌, Vol. 7, No. 3, 475-486 (1992)
- (63) 平嶋 宗, 河野 隆宏, 柏原 昭博, 豊田 順一: "算数の文章題を対象とした問題演習支援機能の実現", 電子情報通信学会論文誌, Vol. J75-A, No. 2, 296-304 (1992)
- (64) 平嶋 宗, 東 正造, 柏原 昭博, 豊田 順一: "補助問題の定式化", 人工知能学会誌, Vol. 10, No. 3, 413-420 (1995)
- (65) 平嶋 宗, 梅田 多一, 志岐 隆弘, 竹内 章: "XMLを用いた算数の文章問題の作成・共有環境", 教育システム情報学会誌, Vol. 18, No. 3, 284-296 (2001)
- (66) 金西計英, 林賢太郎, 光原弘幸, 矢野米雄: "教材知識に基づき WBT 上で演習問題を生成する機能の実現", 教育システム情報学会誌, Vol. 20, No. 2, 71-81 (2003)
- (67) 隅田 英一郎, 菅谷 史昭, 山本 誠一: "英語能力測定のための空所補充問題の自動生成手法", 電子情報通信学会技術研究報告. TL, Vol. 104, No. 503, 17-22 (2004)
- (68) 松田憲幸, 小川修史, 平嶋宗, 瀧寛和: "選択問題の自動生成に関する構想", 教育システム情報学会研究報告, Vol. 20 No. 3, 37-38 (2005)
- (69) 舟生 日出男, 穂山 雅史, 平嶋 宗: "問題解決モデルを利用した選択問題の選択肢と解説の自動生成", 人工知能学会全国大会, 2E4-4 (2007)
- (70) 植野真臣: "知識社会における e ラーニング", 培風館, 東京 (2007)
- (71) 森本康彦: "e ポートフォリオの理論と実際", 教育システム情報学会誌, 25(2), 245-263 (2008)
- (72) G ポリア, 柿内賢信 (訳): "いかにして問題をとくか?", 丸善株式会社, 東京, (2001)
- (73) Brown, S. and Walter, M.: "Problem Posing: Reflections and Applications", Lawrence Erlbaum, Princeton, 1993
- (74) Silver, E. A. and Cai, J.: "An analysis of Arithmetic Problem Posing by Middle school students", *Journal of Research in Mathematics Education*, Vol. 27, No. 5, 521-539 (1996)
- (75) 中野洋二郎, 坪田耕三, 滝井章: "子どもが問題をつくる", 当洋館出版社, 東京 (1999)
- (76) 中野明, 平嶋宗, 竹内章: "問題を作ることによる学習の知的環境支援", 電子情報通信学会論文誌, Vol. J83-D-I, No. 6, 539-549, (2000)
- (77) 仲林清: "e テスティングと標準化", 植野真臣, 永岡慶三 編著, "e テスティング"(pp. 75-93), 培風館, 東京 (2009)
- (78) ISO/IEC: "Information technology Learning, education and training — Quality management, assurance and metrics — Part 1: General approach". ISO/IEC 19796-1 (2005)
- (79) IMS Global Learning Consortium, Inc.: "IMS Question and Test Interoperability" Version 2.1 Public Draft (revision 2), (2008)
- (80) ISO/IEC: "Information technology — A code of practice for the use of information technology (IT) in the delivery of assessments", ISO/IEC 23988, (2007)
- (81) 仲林 清, 中村明仁, 吉岡俊正, 相良貴子, 加賀田俊, 永岡慶三: "標準規格に準拠したオンラインテストシステム", 日本教育工学会論文誌, 29(3), 299-307 (2006)

# A NAVIGATION METHOD DEPENDING ON INDIVIDUAL E-LEARNING STRATEGIES

Makoto Takeya, Masanao Kobayashi, Tatsuya Nakauchi, Ikuo Kitagaki\* and Keizo Nagaoka\*\*

Takushoku University

Tokyo 193-0985 Japan

{takeya@cs, koba@ner,r38478@st }.takushoku-u.ac.jp

\*Hiroshima University

Hiroshima 739-8512

Kitagaki@hiroshima-u.ac.jp

\*\*Waseda University

Saitama 359-1192

k.nagaoka@waseda.jp

## ABSTRACT

One of the authors had been presented the Strategic Task Sequencing method, shortened to the STS method, by using a learning structural graph. The STS method presents a few optimum learning sequences according to learner's learning strategy. It is important for a navigation system to understand individual learners' learning characteristics and to coach their directions. This paper applies the STS method to the navigation method that is developed as a supporting tool for e-learning. Using the STS method, individual learners are able to decide their learning sequences according to their learning strategies. Next, this paper proposes a fuzzy estimation method for individual e-learning sequence strategies based on the STS method in order that the navigation system grasps their strategies. This estimation method applies Fuzzy decision-making to estimation for strategies based on the STS method, shortened to the FSTS method. This paper, firstly, shows the STS method and its algorithm. Secondary, this paper presents the FSTS method and its algorithm.

**KEYWORDS: E-LEARNING, NAVIGATION, LEARNING STRATEGY, LEARNING SEQUENCING**

## 1. INTRODUCTION

It is very difficult for a learner to set his /her target objective and to decide how to traverse e-learning objectives. One of the authors presented the Strategic Task Sequencing method, shortened to the STS method, by using a learning structural graph [1]. The STS method presents a few optimum learning sequences according to learner's learning strategy. This paper applies the STS method to a navigation method that is developed as a supporting tool for e-learning. Using the STS method, individual learners are able to decide their learning sequences according to their learning strategies. Next, this paper proposes a fuzzy estimation method for individual e-learning sequence strategies based on the STS method in order that individual learners grasp their strategies. It is important for the navigation system to understand individual learners' learning characteristics. This estimation method applies Fuzzy decision-making to estimation for strategies based on the STS method, shortened to the FSTS method [2]. This paper, firstly, shows a navigation method using the STS method and its algorithm. Secondary, this paper presents the FSTS method and its algorithm.

## 2. E-LEARNING STRUCTUREAL GRAPH

### 2.1 e-Learning spaces

In the e-learning, individual learners are setting target objective from complexity of a structural learning network. All of the learning objectives required toward the target are extracted according to individual prerequisites as shown in Fig.1. In order to support the e-learning, the navigation algorithm, called Strategic Task Sequencing (STS) method was developed, and the recommended sequence is presented to the individual learners based on the STS method. The extracted learning

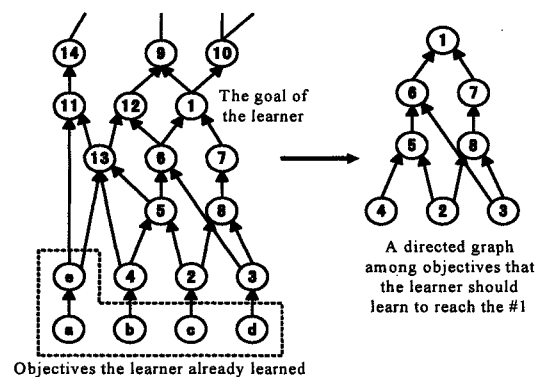


Fig.1 An extraction process model of learning structure based on orientation technology

objectives to learn can be represented as a digraph, where each node and arrow corresponds to each objective and ordering relation, respectively. Here, the digraph is a contraction of the words 'directed graph'. The learner sets node 1 as target of the extracted learning structure graph in Fig.1.

Objective 1 appears to be the terminal objective for the theme, which the learner is facing, and is superior to all other objectives. All other nodes are either directly or indirectly subordinate objectives of the element. On the other hand, objectives of 2, 3 and 4 are at the base level with no supporting objectives. One of these three objectives is the ideal one to start with for the learning sequence. Thus, one of these elements is a candidate to be the ideal starter for the learning sequence. No reverse ordering against arrows should be taken on the digraph. For example, the sequence:  $O_1 = \langle 4 \Rightarrow 2 \Rightarrow 3 \Rightarrow 5 \Rightarrow 8 \Rightarrow 6 \Rightarrow 7 \Rightarrow 1 \rangle$  is legal, but the sequence:  $O_2 = \langle 4 \Rightarrow 2 \Rightarrow 5 \Rightarrow 6 \Rightarrow 3 \Rightarrow 8 \Rightarrow 7 \Rightarrow 1 \rangle$  is illegal, because this sequence includes the reverse order of  $3 \rightarrow 6$ . Objective 3 should be completed before objective 6. Even in the model with only eight objectives in Fig.5.1, however, there are 52 reasonable sequences to satisfy the criterion of topologically sorted ordering. Generally speaking, many reasonable sequences exist that will satisfy the above condition.

## 2.2 Preparation of the STS method

The Strategic Task Sequencing (STS) method has been proposed to order learning tasks on the one dimensional scale, based on learning structure graph on two-dimensional learning space. The sequence is represented by  $O = \langle v_1 \Rightarrow v_2 \Rightarrow \dots \Rightarrow v_n \rangle$ . In related to the order between each objective,  $v_i$  is indispensable prerequisite objective to  $v_j$ . And it represented  $v_i \rightarrow v_j$  that  $v_i$  must learn previous to  $v_j$ .

- Superior objective and subordinate objective  
Superior objectives and subordinate objectives of a node  $v_i$  are defined by reachable nodes from the node  $v_i$  and nodes that are reachable to the node  $v_i$ , respectively.

- Superior objective set and subordinate objective set  
A superior objectives set and a subordinate objectives set of  $v_i$  is represented  $S_p(v_i)$  and  $S_o(v_i)$ , respectively. Common superior objectives set and common subordinate objectives set is called by coupling superior objectives set and coupling subordinate objectives set, respectively. Coupling superior objectives set and coupling subordinate objectives set is represented by  $S_p(v_i) \cap S_p(v_j)$  and  $S_o(v_i) \cap S_o(v_j)$ , respectively. When  $S$  is represented the set, number of elements in the set  $S$  is represented by  $N[S]$ .

- The prerequisite-directed and object-directed index

The prerequisite-directed index between a pair of objective  $v_i$  and  $v_j$  is defined as  $f_{ij}$ .

$$f_{ij} = N[S_p(v_i) \cap S_p(v_j)] / N \quad (i, j = 1, 2, \dots, n) \quad (1)$$

Similarly, the object-directed index between a pair of objective  $v_i$  and  $v_j$  is defined as  $g_{ij}$ .

$$g_{ij} = N[S_o(v_i) \cap S_o(v_j)] / N \quad (i, j = 1, 2, \dots, n) \quad (2)$$

- The sequencing index

The sequencing index represents that each of the prerequisite-directed index and the object-directed index is set and the learning strategy is set what-like putting weight. Specifically, the sequencing index between a pair of objective  $v_i$  and  $v_j$  is defined as  $\theta_{ij}$ .

$$\theta_{ij} = \theta(v_i, v_j) = \alpha \cdot f_{ij} + \beta \cdot g_{ij} \quad (i, j = 1, 2, \dots, n), \quad (3)$$

$$\text{where } |\alpha| + |\beta| = 1 \quad (4)$$

$\alpha$  and  $\beta$  are called by prerequisite-directed and object-directed weights, respectively. Also, constrained condition of  $\alpha$  and  $\beta$  adopts so that the sequencing index is normalized. Individual learning strategies are taken according to changes of  $\alpha$  and  $\beta$  values.

### 3. NAVIGATION BY USING THE STS METHOD

#### 3.1 Description of learning strategies

The values of  $\alpha$  and  $\beta$  change under a constraint condition of  $|\alpha|+|\beta|=1$ , as shown in Fig.2. There exist eight typical learning strategies as follows: fundamental-oriented, prerequisite-oriented, context-oriented, objective-oriented, application-oriented, non-prerequisite-oriented, non-context-oriented, and non-objective-oriented. The most typical four learning strategies are described in the following:

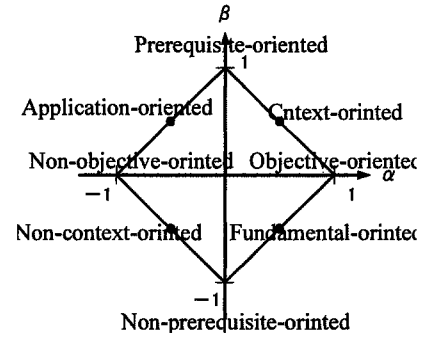


Fig.2 Relationship among typical sequence and  $\alpha$  and  $\beta$

- **Context-oriented strategy**  
This sequencing is made when the values of  $\alpha$  and  $\beta$  take  $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$  ( $\alpha = 1/2, \beta = 1/2$  as typical example). The nodes are selected focusing on continuousness according to both prerequisite and objective nodes. One of optimum sequence based on context-oriented strategy:  
 $O_3 = \langle 2 \Rightarrow 3 \Rightarrow 8 \Rightarrow 7 \Rightarrow 4 \Rightarrow 5 \Rightarrow 6 \Rightarrow 1 \rangle$ .
- **Fundamental-oriented strategy**  
This sequencing is made when the values of  $\alpha$  and  $\beta$  take  $-1 \leq \alpha \leq 0, 0 \leq \beta \leq 1$  ( $\alpha = 1/2, \beta = 1/2$  as typical example). One of optimum sequence based on fundamental-oriented strategy:  
 $O_4 = \langle 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 8 \Rightarrow 6 \Rightarrow 7 \Rightarrow 1 \rangle$ .
- **Application-oriented strategy**  
The sequencing is made when values of  $\alpha$  and  $\beta$  take  $0 \leq \alpha \leq 1, -1 \leq \beta \leq 0$  ( $\alpha = 1/2, \beta = -1/2$  as typical example). One of optimum sequence based on application-oriented:  
 $O_5 = \langle 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 8 \Rightarrow 6 \Rightarrow 7 \Rightarrow 1 \rangle$ .
- **Non-context-oriented strategy**  
This sequencing is made when the values of  $\alpha$  and  $\beta$  take  $-1 \leq \alpha \leq 0, -1 \leq \beta \leq 0$  ( $\alpha = -1/2, \beta = -1/2$  as typical example). This strategy should be never recommended.

The rest four strategies are omitted. The most recommended strategy is context-oriented strategy.

#### 3.2 Algorithm of the STS method

About algorithm of the STS method describes henceforth. Also, it illustrates Fig.3 as image of sequencing.

- 1) Hypothetical node  $v_0$  is added to connect to all the first selectable nodes ( $=V_1$ ), that is no descendent nodes. That is, arrows are drawn from  $v_0$  to any  $v_{1k} (\in V_1)$ .  $i=1$ , and go 2.
- 2) For any  $v_{ik} (\in V_i)$ ,  $v_i$  is extracted satisfying the following inequality:  
$$\theta(v_{i-1}, v_j) \leq \theta(v_{i-1}, v_i). \quad (5)$$

The selected node  $v_i$  is added to the end of the sequence.

- 3) And a set  $V_{i+1}$  of nodes including new selectable nodes are extracted. Here, the node  $v_i$  is deleted from  $V_{i+1}$ . This algorithm is terminated, when  $i = n$ . Otherwise,  $i = i + 1$ , and go to 2.

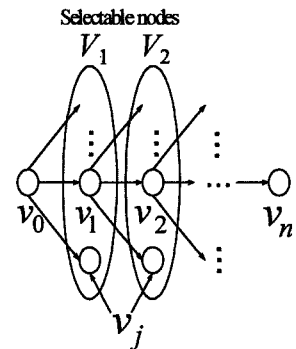


Fig.3 The image of sequencing process

## 4. A FUZZY ESTIMATION METHOD FOR LEARNING STRATEGY FROM E-LEARNING SEQUENCE RECORDS

Finally, this section describes the estimation method for individual strategies in accordance with the STS method. That is, when given e-learning structure graph and individual past sequence records, individual learning strategies are estimated. It is assumed that the sequence is given by the sequence  $O_4$ .

Next, it is assumed that all the nodes are selected as to satisfy Eq.(5). That is, each range of one-dimensional parameter  $\mu$ , concentrated the two-dimensional parameters  $\alpha, \beta$ , are extracted by  $\theta(v_0, v_1), \theta(v_1, v_2), \dots, \theta(v_{n-1}, v_n)$  as shown in Fig.4. Each stage is piled up on top of one another and as a result is described as shown in Fig5. This graph means a membership function for individual strategies on the given sequence. On the other hand, all of the membership functions of individual strategies are described as shown in Fig.6. A fuzzy decision-making is applied to the estimation method as shown in Fig.7. Fuzzy goals are defined by  $G_i (i = 1, 2, \dots, 8)$  and fuzzy constraint is defined by the given learning sequence  $O$ . Then, fuzzy decision of strategy of  $O$  is described by  $D$ . The goal  $D$  is divided into eight sub-goals  $D_i (i = 1, 2, \dots, 8)$ .

$$\text{That is, } D = \bigcup_i D_i \quad (6)$$

$$\text{Also, } M_{D_i}(\mu) = M_{G_i}(\mu) \cap M_O(\mu) \quad (7)$$

Therefore, maximum grade  $h(D)$  of fuzzy decision  $D$  is described as follows:

$$h(D) = \text{Max.}\{\bigcup_i (M_{G_i}(\mu) \cap M_O(\mu))\} \quad (8)$$

Let's represent a  $\alpha$ -level set of a fuzzy set  $A$  by  $A_\alpha$ . Then,

$$A_{h(D)} = \{\mu \mid M_D(\mu) \geq \alpha\} \quad (9)$$

In this example of  $O_4$ ,  $h(D) = 1.0$  and  $A_{h(D)} = \{\mu \mid \mu = 0.125\}$ . As a result, the sequence  $O_4$  is estimated as prerequisite-oriented strategy with the membership function value 1.0.

## 5. CONCLUSIONS

This paper presented a navigation algorithm depending on individual learning strategies. In future we are going to evaluate validity of the algorithm.

## REFERENCES

- Takeya, M., *Instructional Architecture, Design and Evaluation: Theory and practice of instructional architecture, design and evaluation*, Research Institute of Science and Engineering, Takushoku University, 1999
- Takeya, M., Funabashi, Y. and Nakauchi, T., An estimation method for strategy of a task sequence according to the Strategic Task Sequencing method, *Transactions of Institute of Electronics and Communication Engineers* (in Japanes) (in print)

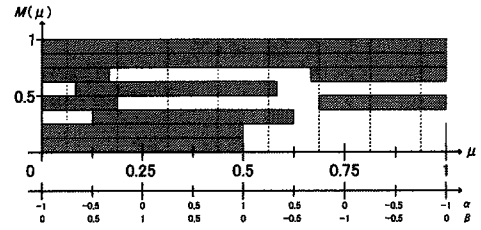


Fig.4 An illustration of construction process of a membership function of the sequence  $O_4$ .

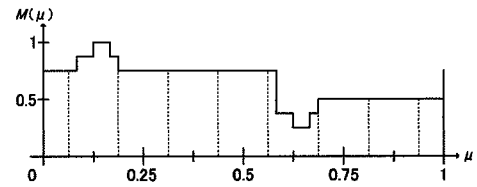


Fig.5 The membership function of of the sequence  $O_4$ .

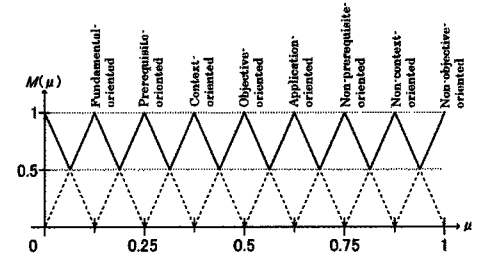


Fig.6 Composite membership functions of eight goals according to individual strategies.

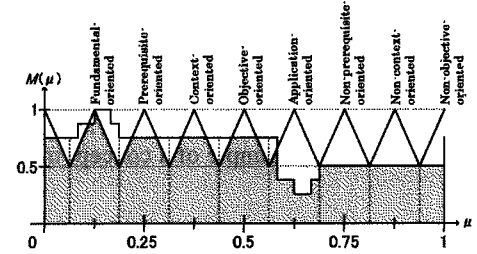


Fig.7 Fuzzy decision-making based on a fuzzy constraint  $O_4$  and fuzzy goals  $G_i (i = 1, 2, \dots, 8)$ .

# MEASUREMENT AND EVALUATION METHOD FOR A CONCEPT MAPPING TEST BY DRAWING ORDERING RELATIONS AMONG CONCEPTS

Makoto Takeya, Nobuaki Yasugi, Yoshio Funabashi & Keizo Nagaoka  
Takushoku University, Mitsubishi EIN, Fujitsu SSL & Waseda University  
Email: takeya@cs.takushoku-u.ac.jp

**Abstract.** The traditional paper test is able to measure mainly individual student's understanding level of individual bits of knowledge. However, it is difficult to measure the internal connection among bits of knowledge. Previously, Takeya, *et al.* (2004) had presented a new testing method, called a concept mapping test for a formative evaluation tool at the 1<sup>st</sup> CMC. This is a test by using concept maps based on the prerequisite relations among concepts. However, future subjects lie in an evaluation method for a concept mapping test based on ordering relations with transitivity law, such as casual relations, inclusion relations, and so on. This paper presents new measurement and evaluation method for these kinds of structural knowledge. Secondly, giving a few examples, this paper discusses scoring method based on qualitative degree of sequencing. Thirdly, these characteristics and validities of the measurement and evaluation method are discussed.

*Category/Categoria: Poster*

## 1 Introduction

Lectures and texts are arranged in linear or sequential order. Each learning unit is presented in order. That is, they naturally move from one idea to the next, and so forth, without ever systematically detailing the structural relationships among these ideas. The teacher is concerned with assessing and promoting the acquisition of knowledge by individual students. Attention has recently been drawn to what has become known as 'structural knowledge' or knowledge of interrelationships among ideas in their knowledge domain. The authors are becoming aware of the need to establish the internal connectedness of ideas and concepts to be learned. It is difficult to evaluate these internal relationships among ideas by using traditional paper tests, because these tests mainly measure the understanding level of individual bits of knowledge obtained by individual students.

## 2 Measurement and evaluation by a concept mapping test based on ordering relations

It is very important to check whether each relationship, *i.e.* each arrow in the student's map exists or not compared with the teacher's map. Especially, we have to investigate whether misunderstanding of relationships occur or not under the influence of existence or non-existence of individual arrows. Define a concept map by a digraph (directed graph)  $G = (V, E)$ , where  $V$  is a set of vertexes  $V = \{v_1, v_2, \dots, v_n\}$ , and  $E$  is a set of arrows  $E = \{e_1, e_2, \dots, e_m\}$ . The remarkable feature of the concept map  $G$  with respect of casual relations, inclusion relations, ordering relations, and so on is that the map is transitive. Here,  $G$  is transitive if and only if there exists an arrow  $(v_i, v_k)$  whenever both arrows  $(v_i, v_j)$  and  $(v_j, v_k)$  exists. That is, it is satisfied that whenever  $v_i \rightarrow v_j$ , and  $v_j \rightarrow v_k$ , then  $v_i \rightarrow v_k$ . Previously, Takeya (1999) and Takeya, *et al.* (2004) had presented the similarity index concentrating on relations of arrows. Contrary to this, a new similarity index should be discussed from a view point of ordering with transitivity law. Next, an example of a concept mapping test is shown before the new similarity index will be introduced in 2.2.

### 2.1 A concept mapping test on casual relationships

A concept map has been utilized in lectures on Environmental Science at Takushoku University in Japan. After lecturing and showing a video on the "Crisis of the Living Environment at the Foot of Mt. Fuji", a performance test is given by use of a concept map. Test contents are shown as follows:

Included are the following nineteen elements. Please draw a concept map, where an arrow " $a \rightarrow b$ " means a direct relation between cause  $a$  and effect  $b$ . Please show your map in the form of a hierarchical structure. Here, the element on the first level is only the element (1).



As shown in Figure 1, their contents include the following:

- (1) Crisis of the living environment at the foot of Mt. Fuji (the target)
- (2) Industrial use of underground water
- (3) Water pollution in rivers
- (4) Appearance of artificial valleys
- (5) Agricultural damage
- (6) Atmospheric contamination
- (7) Change of underground water to salt water
- (8) Deforestation
- (9) Sprinkling of agricultural chemicals
- (10) Factory construction
- (11) Increase of diseased trees in forest
- (12) Dumping of factory wastes into rivers
- (13) Gas emissions from automobiles
- (14) Decrease of underground water
- (15) Decrease of water retention power of the ground
- (16) Golf course development
- (17) Occurrence of avalanche accidents
- (18) Subaru-line road construction
- (19) Dumping of factory wastes into rivers

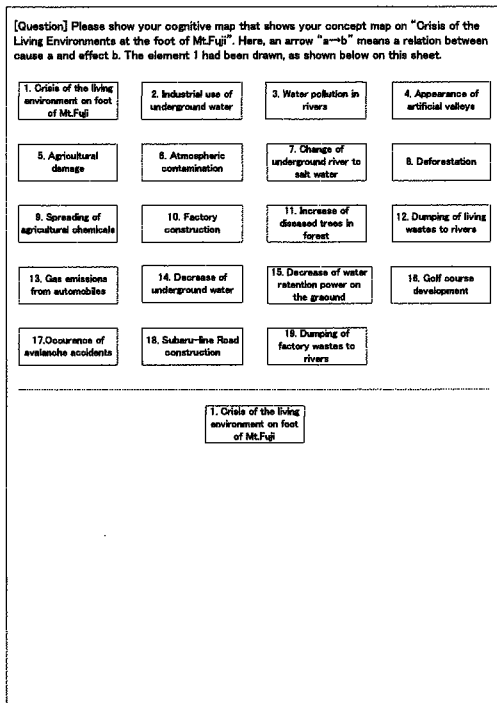


Figure 1. An example of the concept mapping test sheet.

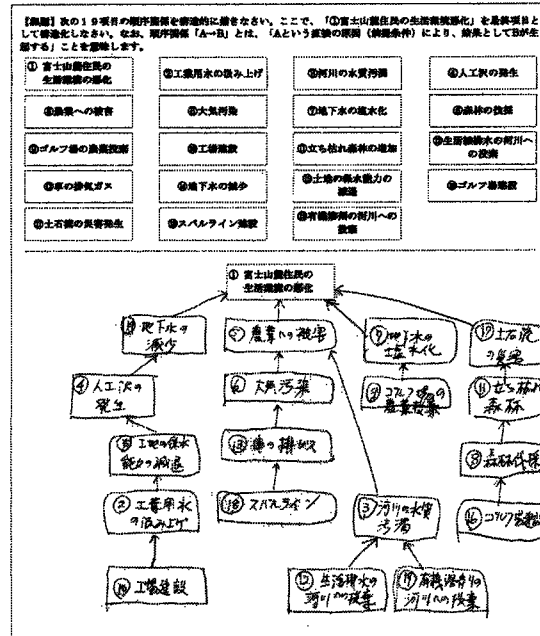


Figure 2 An original sheet of a concept mapping test drawn by one of students. (in Japanese)

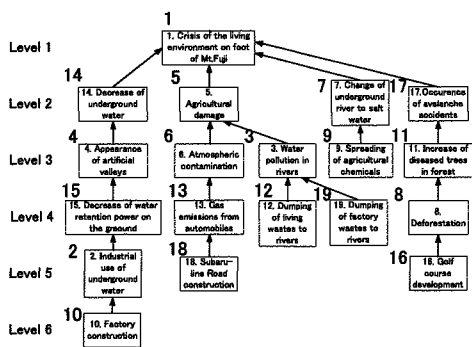


Figure 3 The rearranged English version of the concept map shown in Figure 2.

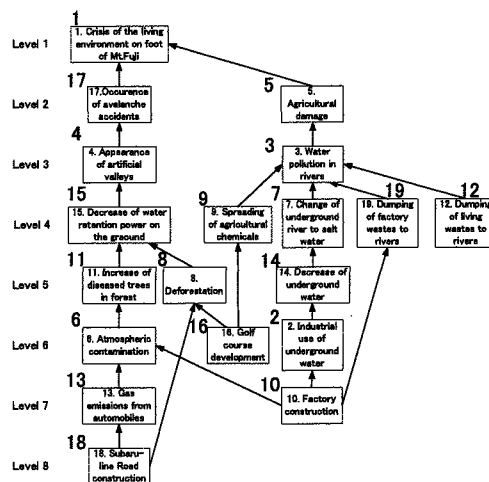


Figure 4 The concept map drawn by a professor.

Figure 1 shows a Japanese original of a concept mapping test drawn by one of students. Figure 2 represents rearranged English version of the concept map shown in Figure 1. Corresponding to Figure 3, Figure 4 shows the concept map drawn by the professor. According to the calculation of the performance score described in 2.2, the professor can understand the performance level for each student. In this case, the student receives a score of 42, described in 2.2.

## 2.2 Measurement and evaluation method

In the case where the digraph satisfies the transitive law, such as the map on casual relationships, and inclusion relationships, the teacher has to give attention to ordering of vertexes. Call a transitive digraph except a totally ordered digraph as a partially ordered digraph. So, represent both an antecedent set of the vertex  $v_j$  and a reachable set of  $v_j$  on  $G_X$ , by  $V\langle \rightarrow v_j \rangle^X$  and  $V\langle v_j \rightarrow \rangle^X$ , respectively. Then, similarity index  $S'(G_T, G_X)$  among the teacher's map  $G_T$  and the student's map  $G_X$  is defined as follows:

$$S'(G_T, G_X) \equiv \frac{\sum_{j=1}^n \left( \left( V\langle \rightarrow v_j \rangle^T \cap V\langle \rightarrow v_j \rangle^X \right) \cup \left( V\langle v_j \rightarrow \rangle^T \cap V\langle v_j \rightarrow \rangle^X \right) \right)}{\sum_{j=1}^n \left( \left( V\langle \rightarrow v_j \rangle^T \cup V\langle \rightarrow v_j \rangle^X \right) \cup \left( V\langle v_j \rightarrow \rangle^T \cup V\langle v_j \rightarrow \rangle^X \right) \right)}. \quad (1)$$

Naturally, the following inequality is obtained.

$$0 \leq S'(G_T, G_X) \leq 1 \quad (2)$$

Considering the maps in Figure 1 as partially ordered digraph, let's show examples of calculation process. For example, pay attention to the vertex 3.

$$V\langle \rightarrow 3 \rangle^T = \{2,7,9,10,12,14,16,19\}, \quad V\langle \rightarrow 3 \rangle^X = \{12,19\},$$

$$V\langle 3 \rightarrow \rangle^T = \{1,5\}, \quad V\langle 3 \rightarrow \rangle^X = \{1,5\}, \quad \text{and}$$

$$\left( \left( V\langle \rightarrow 3 \rangle^T \cap V\langle \rightarrow 3 \rangle^X \right) \cup \left( V\langle 3 \rightarrow \rangle^T \cap V\langle 3 \rightarrow \rangle^X \right) \right) = 2 + 2 = 4$$

Next, for the vertex 8,  $V\langle \rightarrow 8 \rangle^T = \{16,18\}$ ,  $V\langle \rightarrow 8 \rangle^X = \{16\}$ ,

$$V\langle 8 \rightarrow \rangle^T = \{1,4,15,17\}, \quad V\langle 8 \rightarrow \rangle^X = \{1,11,17\}, \quad \text{and}$$

$$\left( \left( V\langle \rightarrow 8 \rangle^T \cap V\langle \rightarrow 8 \rangle^X \right) \cup \left( V\langle 8 \rightarrow \rangle^T \cap V\langle 8 \rightarrow \rangle^X \right) \right) = 3.$$

In the same way as the vertex 3 and 8, these values of all the other vertexes involved in  $G_T$  and  $G_X$  are calculated.

## 2.3 Characteristics and validity of the similarity index

Here, to analyse similarity index,  $S'(G_T, G_X)$  is linearly transformed as follows:

$$\psi(G_T, G_X) \equiv 2S'(G_T, G_X) - 1. \quad (3)$$

That is,  $-1 \leq \psi(G_T, G_X) \leq 1$ . (4)

Let's refer to this index as the *Takeya's  $\psi$* . According to Eq.(3), it is considered that this index has very interesting characteristics. Ranking data can be interpreted as vertexes on a linear digraph such as  $G_L$  shown in Fig.5. Strictly speaking, note that the linear digraph is a totally ordered digraph which holds the transitive law. Whenever both  $G_T$  and  $G_X$  are linear graphs, then the *Takeya's  $\psi$*  coefficient is equivalent to both the *Goodman-Kruskal  $\gamma$*  coefficient, *Somers' d* coefficient and *Kendoll's  $\tau$*  coefficient well known as measures of association in ordinal data.

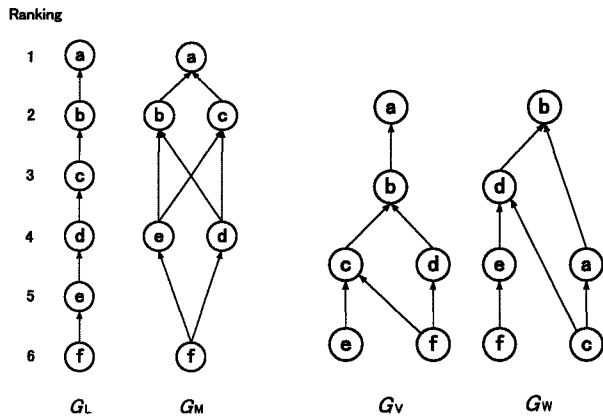


Figure 5 Graphical structure of totally ranking data .

Figure 6 Graphical structure of partially ranking data.

That is, whenever both  $G_T$  and  $G_X$  are linear graphs, then

$$\psi(G_T, G_X) = \gamma(G_T, G_X) = d(G_T, G_X) = \tau(G_T, G_X) . \quad (5)$$

Moreover, in the case where  $\gamma$  coefficient is defined in the expanded range of non-linear data structure as shown in Figure 6, the following relation is obtained.

$$\gamma(G_T, G_X) \geq \psi(G_T, G_X) . \quad (6)$$

The proof of the Eq.(5) and (6) are omitted by the space limitation. If examined in detail distributions of  $\psi$  and  $\gamma$  of several practical concept mapping tests, the values of *Takeya's*  $\psi$  coefficient cover the range from -0.5 to 1, but the most of values of *Goodman-Kruskal*  $\gamma$  coefficient converge on the range from 0.8 to 1. Also, in Figure 6,  $\gamma(G_V, G_W) = 1.00$ , but  $\psi(G_V, G_W) = .40$ . The ordered pairs  $(d, a), (e, a), (f, a), (e, c)$  and  $(f, c)$  exist on  $G_V$ , but never exist on  $G_W$ . On the other hand, the ordered pairs  $(c, d), (e, d)$  and  $(f, e)$  exist on  $G_W$ , but there never exist on  $G_V$ . The  $\gamma$  coefficient is taken no account of the above differences. On the other hand, those differences are built in to the  $\psi$  coefficient. Therefore, it is considered that the  $\psi$  coefficient is the rank correlation generalized to partially ordered data structure. Here, this theoretical description is omitted. As a result, the  $\psi$  coefficient is applicable not only to linear hierarchical structure, but also to non-linear one. That is, the *Takeya's*  $\psi$  coefficient is generalization of traditional several rank coefficients.

### 3 Summary

This paper presented measurement and evaluation for concept mapping tests based on casual relationships. Secondly the measurement method for structural knowledge was shown by using both the models of a concept map and the actual mapping tests. Thirdly, this paper discussed scoring methods according to qualitative degree of sequencing. Lastly, the validity was confirmed. Especially, it was derived that the scoring method for partially ordered map is based on the *Takeya's*  $\psi$  coefficient which is generalized from the traditional *Goodman-Kruskal*  $\gamma$ . One of the problems to be solved in near future is development on a reformation algorithm for individual students' misunderstanding.

### 4 References

- Takeya, M. (1999). Structure analysis methods for instruction: Theory and practice of instructional architecture, design and evaluation, Takushoku University Press, Seizansya.
- Takeya, M., Sasaki, H., Nagaoka, K., & Yonezawa, N. (2004). A performance scoring method based on quantitative comparison of concept maps by a teacher and students, Proceedings of the First International Conference on Concept Mapping, 2, 343-346.

# 機械設計授業におけるグループ学習の学習効果の検証

富田 聡<sup>†</sup> 神保 智広<sup>†</sup> 木原 幸一郎<sup>†</sup> 杉林 俊雄<sup>†</sup> 竹谷 誠<sup>†</sup>

<sup>†</sup> 拓殖大学 工学部 〒193-0985 東京都八王子市館町 815-1

E-mail: <sup>†</sup>{y7m316,r38455}@st.takushoku-u.ac.jp, <sup>††</sup>{kkihara,tsugi}@ms.takushoku-u.ac.jp

<sup>†††</sup>takeya@cs.takushoku-u.ac.jp

**あらまし** 論理的な問題を解決するプロセスを複数の要素と要素間の順序関係を有向枝で表現したロジカルフローグラフ（以下、LFG と呼ぶ）がある。この LFG を利用した学習者の理解構造の理解力を評価するロジカルフローテスト法（以下、LFT と呼ぶ）が提案されている。本稿では、LFG を用いたグループ学習を提案する。それに伴い個別学習とグループ学習それぞれにおける LFT の評価からグループ学習の学習効果の検証を行った。検証では、本学機械システム工学科の授業「システム設計 II」において LFT を適用し行った。LFT の結果よりグループ学習の学習効果に考察したのでそれについて報告する。

**キーワード** 構造的な理解, ロジカルフローグラフ, ロジカルフローテスト, グループ学習

## Validation of Group Learning Effects in the Mechanical Design Class

Satoshi TOMITA<sup>†</sup>, Tomohiro JIMBO<sup>†</sup>, Koichiro KIHARA<sup>†</sup>, Toshio SUGIBAYASHI<sup>†</sup>, and

Makoto TAKEYA<sup>†</sup>

<sup>†</sup>Faculty of Engineering, Takushoku University 1-2-3 Tatemachi, Hachioji, Tokyo, 193-0985 Japan

E-mail: <sup>†</sup>{y7m316,r38455}@st.takushoku-u.ac.jp, <sup>††</sup>{kkihara,tsugi}@ms.takushoku-u.ac.jp

<sup>†††</sup>takeya@cs.takushoku-u.ac.jp

**Abstract** The measurement method for students' structural understanding using a Logical Flow Graph (LFG) is called as Logical Flow Test (LFT) and the LFT is effective for measurement and evaluation of their logical thinking ability. This paper discusses effectiveness of group learning compared with individual learning by using their LFGs. That is, this paper describes a case study performed with "System Design II" class for the third grader in Department of Mechanical System Engineering of our university. As a result, it became clear that group learning by using LFGs is effective for their improvement of logical thinking ability.

**Keyword** structural knowledge, Logical Flow Graphs, Logical Flow Test method, Group Learning

### 1. はじめに

学校教育における教育測定評価は多義にわたるが、その中において、学習者の定量的評価は以前よりその重要性が指摘されている。学習者を評価する際、個々の知識の理解度や知識量の測定だけでなく、知識間の構造的な理解の測定の重要性も指摘されている。それを解決する方法として、これまでに定量的な評価の実現を目指して、認知構造を有向グラフで表現する手法の一つである LFG に着目し、教師と学習者が描画した LFG の類似性を定量的に評価する類似度に基づくロジカルフローテスト分析法（以下、LFT と呼ぶ）が提案されている [1]。一方、それとは異なる測定方法に則った差異度の提案がなされている [2]。これらの方法は、教師が学習課題の導出過程の関係を描画した LFG と学習者が学習課題の導出過程の関係を描画した LFG

の類似（または差異）の程度を定量化したものである。

この定量化された類似度（または差異度）により、学習者の導出過程の正しさの程度を測定評価することができる。また、グループ学習の教育的重要性も以前から指摘されている。これに関しては先に LFG を用いてグループで合意形成する方法論が提案されている [3]。

本研究では、LFG を用い個別学習とグループ学習で同じ課題に取り組ませ、個別学習による構造的な理解力の評価とグループ学習による構造的な理解力の評価を比較し、グループ学習の学習効果について検証を行うことを目的とする。本稿では、グループ学習の効果測定のために、本学の授業において実践検証したので報告する。具体的には、2. で LFG と LFT について説明

を行い、3.では、実験内容、予備実験また本実験について述べる。次に4.では、実験の結果と考察について述べる。最後に5.では、まとめと今後の予定について述べる。

## 1. LFG に関する基礎定義

### 2.1 LFG と LFT

基本定義に先立ち、LFT についての基本的な考え方を説明する[1]。授業において、学習者が単元や領域の知識を取得するとき、それぞれの知識の内容を取得することはもちろん大切であるが、複雑に関連する知識を関連付けて理解することも重要である。そこで、教師は、ばらばらな知識の羅列として指導するのではなく、LFG に基づき、知識をチャンクしやすいように指導内容と指導順序を決定する。つまり、複数の学習課題をひとつの認知ユニットとしてチャンク化し、いくつかのチャンク化された認知ユニットをまとめ、構造的な理解が取得しやすいように工夫するわけである。したがって、学習者が獲得した知識を体系的にどのように把握しているかを測定評価することが求められる。教師と学習者の LFG を比較し、観察しただけでどの程度類似しているかを定量化することは甚だ困難である。それを解決するために、LFG 間の類似度を提案するが、それに先立ち、諸定義を行う。

### 2.2 LFG の基礎定義

LFG  $G$  は、 $n$  個の要素を頂点  $v_i = (v_1, v_2, \dots, v_n)$  とした頂点集合  $V = \{v_1, v_2, \dots, v_n\}$ 、 $m$  個の順序関係を有向枝  $e_k = (k=1, 2, \dots, m)$  とした枝集合  $E = \{e_1, e_2, \dots, e_m\}$  を用いた、非循環有向グラフ  $G = (V, E)$  で表す。ここで、頂点  $v_i$  から頂点  $v_j$  への有向枝を、順序関係  $(v_i, v_j)$  で表す。また、教師の LFG と学習者の LFG をそれぞれ、 $G_T = (V, E_T)$ 、 $G_S = (V, E_S)$  で表す。前提として、教師の設定した要素集合と同一の要素集合を用いて、学習者が LFG を描画する。つまり、 $V_T = V_S$  とする。次に教師、学習者の枝  $(v_i, v_j)$  を通る経路の集合に含まれる始点と終点の集合をそれぞれ  $O(\rightarrow(v_i, v_j) \rightarrow)^T$ 、 $O(\rightarrow(v_i, v_j) \rightarrow)^S$  で表す。ここで、 $O(\rightarrow(v_i, v_j) \rightarrow)^Z (Z=T, S)$  は、枝  $(v_i, v_j)$  を通る経路の集合であることに注意されたい。つまり、 $O(\rightarrow(v_i, v_j) \rightarrow)^Z$  に含まれる経路はすべて、その始点は  $v_i$  の先行集合、その終点は  $v_j$  の可到達集合にそれぞれ含まれる。換言すると、 $O(\rightarrow(v_i, v_j) \rightarrow)^Z$  の始点集合は、 $v_i$  の先行集合、終点集合は  $v_j$  の可到達集合と一致する。

### 2.3 類似度の定義

$G_T$  と  $G_S$  の任意の枝  $(v_i, v_j)$  において  $(v_i, v_j)$  を通る経路の集合に含まれる始点と終点の集合の総和をその最大値で正規化した値として、 $G_T$  と  $G_S$  の類似度を定義する。ここで、 $O(\rightarrow(v_i, v_j) \rightarrow)^T$  を  $\alpha^T$  に、 $O(\rightarrow(v_i, v_j) \rightarrow)^S$  を  $\alpha^S$  に簡略化する。これより、 $G_T$  と  $G_S$  の類似度

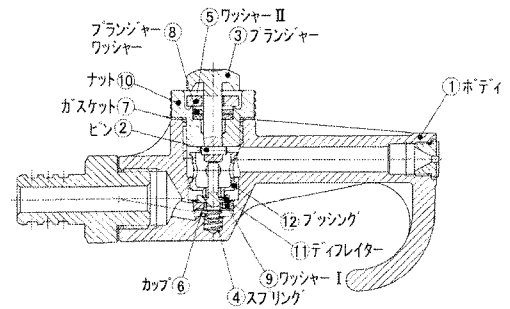


図1 スプレーブローガンの組立図の断面図

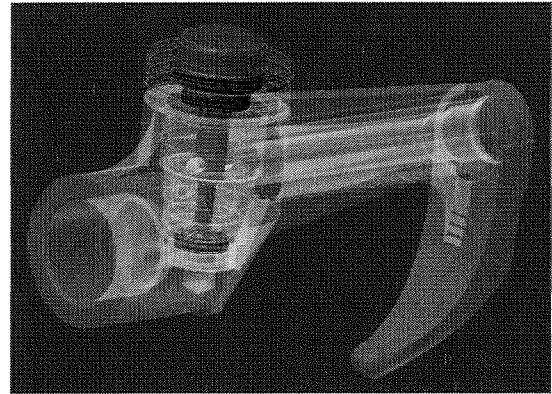


図2 スプレーブローガンの立体図

$\mathfrak{R}(G_T, G_S)$  を次のように定義する。

$$\mathfrak{R}(G_T, G_S) = \frac{\sum_{i,j=1}^n |\alpha^T \cap \alpha^S|}{\sum_{i,j=1}^n |\alpha^T \cup \alpha^S|} \quad (1)$$

当然ながら、 $0 < \mathfrak{R}(G_T, G_S) < 1$  である。この類似度  $\mathfrak{R}(G_T, G_S)$  により、学習者の知識間の構造的な理解を定量的に評価できる。これを用いて個別学習での LFT、グループ学習での LFT について評価を行っていく。

## 2. 実験内容

本学機械システム工学科の授業「システム設計 II」で昨年度行った事例を用い、個別学習での LFT を実施した。LFT の内容は、スプレーブローガンという塗装用の機械の組立プロセスを LFG として描画するものである。

### 3.1 授業内容

本学機械システム工学科の授業「システム設計 II」において、受講している学習者に対し問題形式でスプレーブローガンにおける部品図の組立プロセスを LFG として描画してもらった。この授業は、3 年生を対象とした授業であり、1 年次科目の「図学」・「電算機言語 I」、2 年次科目の「電算機言語 II」・「設計製図 I」・「設計製図 II」、3 年次科目の「設計製図 III」などの設計に関する授業の集大成の科目である。

ここでスプレーブローガンの動作について簡単に説明する。スプレーブローガンの組立図の断面図と立

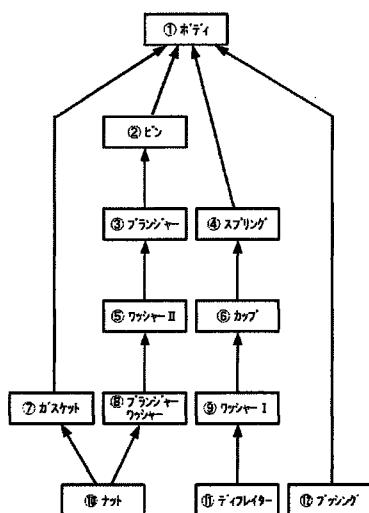


図3 教師が描画したLFG

体図をそれぞれ図1, 2に示す。まず, ③を下に押し込むことにより, 各部品が運動して動作し, ①と⑨の部分に隙間が出来る。その隙間から左端部の圧縮空気が開放され右端部から噴出するしくみになっている。組立図からも分かるように, さまざまな部品が関連しあって1つの役割をするため, ある1つの部品図を作成するとしても, その部品に関連している部品を考慮し製図する必要がある。よって, 部品図を作成する場合には, そのシステム全体を把握し部品間の関連の組立図から読み取り作成していくことになる。

### 3.2 問題形式

学習者には問題用紙とスプレーローガンの組立図の断面図と各部品についての部品図を配布し, 問題用紙にLFGを描画してもらう。このとき問題用紙には, 初期前提課題として⑩と⑪と⑫, 最終目標課題として①をあらかじめ提示している。

### 3.3 スプレーローガンのLFG

スプレーローガンを製図する場合, 互いに独立したのではなく関連をもつため, 部品と部品の関連を考慮する必要がある。ここでスプレーローガンの部品図作成を問題として捉えたとき各部品図の作成は小問題となり, 各部品図を項目・有向枝をその解決順序としてスプレーローガンにおける部品図の組立プロセスをLFGとして描画させる。教師が描画したLFGを図3に示す。このLFGは, ⑩と⑪と⑫の情報を既知であるとし, 最終的に①に組み付けるとき, その間の部品図の組立プロセスを示している。例えば, ⑦は⑩と①の隙間を埋める役割をしている。そのため図3の教師が描画したLFGでは, ⑩→⑦→①の関係が表現されている。

### 3.4 予備実験における考察

本実験の前段階として予備実験を行った[5]。予備実

表1 予備実験によるグループ学習による効果

グループ名	対象	Grとの類似度	Geとの類似度
1	グループ	1.00	
	学習者A	0.45	0.45
	学習者B	0.43	0.43
	学習者C	0.55	0.55
2	グループ	0.94	
	学習者D	0.63	0.60
	学習者E	0.51	0.51
	学習者F		
3	グループ	0.60	
	学習者G	0.44	0.66
	学習者H	0.65	0.42
	学習者I		

験では, この個別学習でのLFTと同様の内容でグループ学習によるLFTを行った。グループは各グループ3人の計3グループになり実施した。この9人に関しては, 昨年度の実験を受けている学習者が9人中7人, 受けていない学習者が2人となっている。表1に個別学習でのLFTで得た類似度の結果とグループ学習のLFTで得た類似度の結果が表されている。また, データが存在しないものに関しては斜線を引いてある。ここで,  $G_G$ をグループでのLFGとする。測定データのある学習者7人の内, 6人は単独でLFTをした場合に比べてグループの類似度が上昇しており, グループ学習の効果が期待できる結果となった。効果の下がった学習者のいるグループ3に着目すると, 効果の下がった学習者Hの $G_S$ では, 誤理解や理解不足が計3箇所の比較的正確に近い理解構造が見られた。 $G_G$ と $G_S$ を比較すると $G_G$ が学習者Gの理解構造に近い要素間の関連付けが見られた。これは, 組み立てた部品同士を組み付けて最後に本体に組み付けるという学習者Gの意見が強く反映, その方法で合意形成され, その部品間の関連付けを多く間違え, 正解に近かった個別より類似度が下がったと考えられる。

### 3.5 個別学習によるLFT

本学機械システム工学科の授業「システム設計II」において, 個別学習によるLFTを行った。受講する学生に対してLFTを実施した。まずはじめに機械システム工学科の学生はLFGについての知識が無いため, LFGについての説明を行った。説明の後, クランプの組立プロセスをLFGで描画する練習問題を行ってもらった。クランプは, スプレーローガンより部品が少なく構成されており比較的簡単な問題である。練習問題を回答させ, 答え合わせを行った。このとき, 組立プロセスを図で示しながら説明を行った。最後にスプレーローガンの組立プロセスの問題を回答してもらった。回答用紙のほかに, スプレーローガンの組立図の断面図と各部品図を配布した。

表2 個別・グループのLFTでの評価

グループ名	対象	Grとの類似度	Ggとの類似度
1	グループ1	0.94	
	学習者1	0.28	0.26
	学習者2	0.95	0.93
	学習者3	0.63	0.58
2	グループ2	0.79	
	学習者4	0.26	0.21
	学習者5	0.45	0.60
	学習者6	0.27	0.40
3	グループ3	0.41	
	学習者7	0.66	0.53
	学習者8	0.22	0.16
	学習者9	0.41	1.00
4	グループ4	1.00	
	学習者10	1.00	1.00
	学習者11	1.00	1.00
	学習者12	0.38	0.38
5	グループ5	0.50	
	学習者13	0.60	0.48
	学習者14	0.35	0.20
	学習者15	0.50	1.00
6	グループ	0.44	
	学習者16	0.75	0.50
	学習者17	0.75	0.37
	学習者18	0.32	0.41
7	グループ	0.63	
	学習者19	0.63	1.00
	学習者20	0.54	0.37
	学習者21	0.43	0.44
8	グループ	0.58	
	学習者22	0.63	0.92
	学習者23	0.21	0.32
	学習者24	0.21	0.32
9	グループ	0.63	
	学習者25	0.63	1.00
	学習者26	0.33	0.42
	学習者27	0.16	0.24

### 3.6 グループ学習によるLFT

グループによるLFTは、3人グループで9グループに対して実施した。グループ学習の利点としては、グループでディスカッションしながら学習を進めることによる知識の獲得の促進やコミュニケーションスキルの向上などが挙げられる。しかし、利点だけではなく問題点も存在する。一方的なメンバーがいた場合、意見を妥協せざるを得なくなったり、他力本願でまったく参加しないメンバーもでてくる。こうしたマイナス面を防ぐために、まず課題を個人で解決し、次にグループで意見を述べ、全員で話し合って合意した考えをまとめることが必要である。よってグループによるLFTはまず、3人それぞれどのようにしてLFGを描画したのかについて発表してもらい、次に互いの意見を出し合いディスカッションした上で合意し新たにグループでのLFGを描画してもらう方法で行った。このとき、ビデオカメラでグループを撮影しながら実施した。ビデオ撮影を行ったグループは、9グループ中5グループに対し行った。

#### 1. 実験の結果と考察

個別でのLFTとグループでのLFTの評価を表2に示す。表の見方は上記でも述べたが、まず最初に $G_T$ と各 $G_G$ との類似度、また $G_T$ と各 $G_S$ との類似度を示している。次に $G_G$ と各グループ中の $G_S$ との類似度を表している。これにより、学習者それぞれが $G_G$ に与えた影響がどのくらいあるかが数値として表される。表2より、 $\mathfrak{R}(G_T, G_S) > \mathfrak{R}(G_T, G_G)$ を満たす学習者は、27人中6人(全体の22%)となった。グループで考えると、 $\mathfrak{R}(G_T, G_S) > \mathfrak{R}(G_T, G_G)$ を満たす学習者が2人いるグループが1グループ、1人いるグループが4グループ、1

人もいないグループが4グループとなった。換言すると、27人中21人はグループ学習によって学習効果が得られたといえる。グループで考えると、9グループ中8グループ( $\mathfrak{R}(G_T, G_S) > \mathfrak{R}(G_T, G_G)$ )を満たす学習者が1人いるグループも含む)と8割以上はグループ学習により学習効果が得られたと考えられる。

次に、グループ学習によって評価が上がったグループ、評価が下がったグループについて考察を行った。また、ビデオ撮影を行ったグループに関しては、プロトコル解析を行い考察した。

#### 4.1 評価が上がったグループに対する考察

ここでは、評価が上がったグループについて考察を行う。表2よりグループ1, 2について考察をする。なおこの2つのグループはビデオ撮影を行っている。

##### ● グループ1

図4に各学習者のLFGとグループでのLFGを示す。まず各学習者のLFGを見てみると、学習者1では、それぞれ初期前提課題から順々に関連付けを行っているのだが、②の関連付けが図面が読み取れず機械下部と関連があると考え誤理解をしている。次に学習者2は、教授者のLFGとほぼ一致しているのだが、⑦からの関連付けをどこにすればいいのか分からなかったと考えられ、⑦から関連付けがない状態となっている。最後に学習者3は、初期前提課題から順々に関連付けして①に関連付けしている。

次にグループのLFGを見てみると、⑩からの経路は教授者と一致している。また、⑪からの経路も一致している。しかし、⑩からの関連付けを誤理解している。このグループではまず、互いの考えを述べながらディスカッションを進めた。まず、学習者1のLFGにおけ

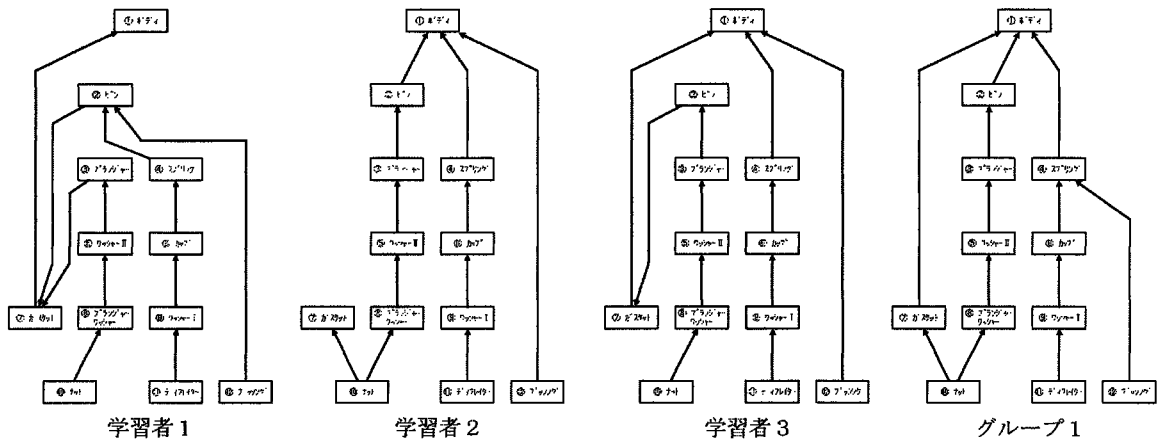


図4 グループ1の各学習者のLFGとグループでのLFG

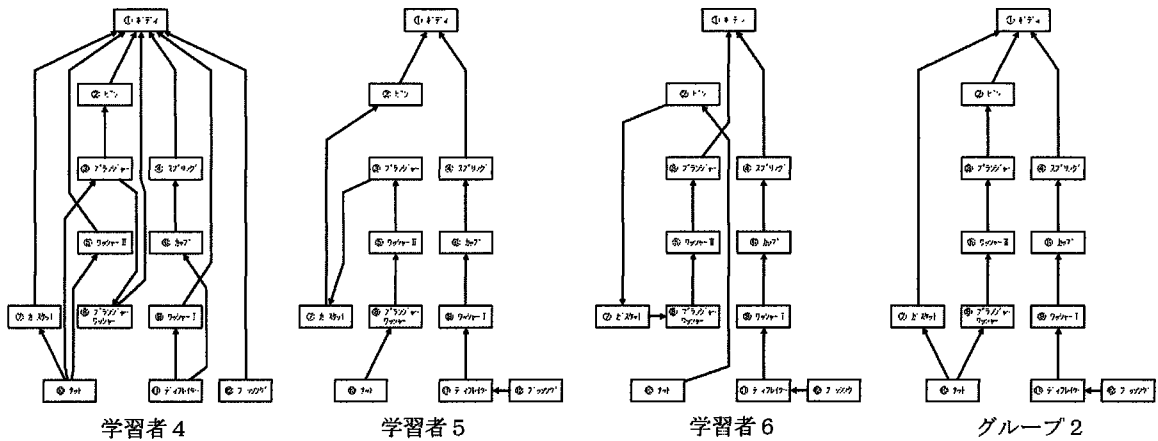


図5 グループ2の各学習者のLFGとグループでのLFG

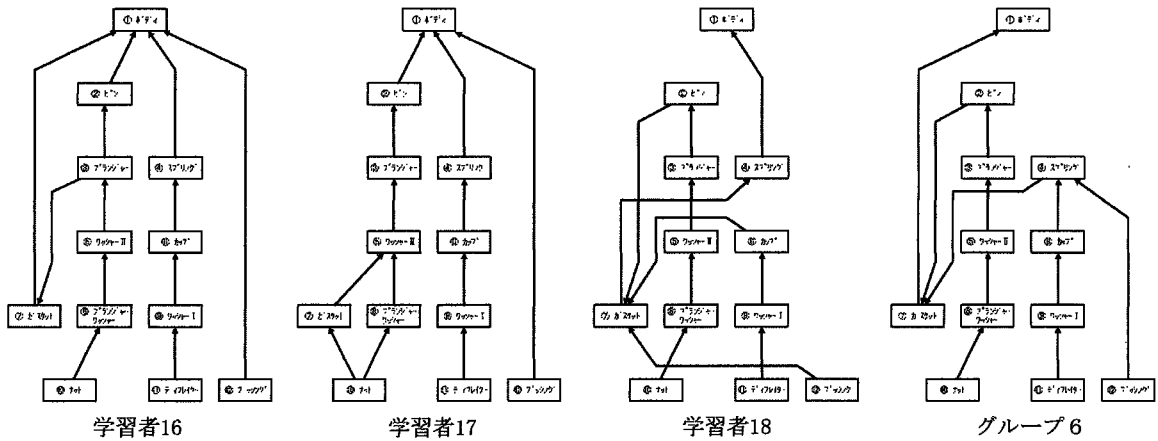


図6 グループ6の各学習者のLFGとグループでのLFG

る②→⑦の関連付けに対して、他の二人から「②と⑦は関連がない」と指摘を受けた。また、同様に「②は⑩と④にも関係ない」と指摘を受けた。学習者1はこの指摘に対し納得し、間違った考えをしていたと気づいた。グループのLFG作成過程として、主に⑦の関連付けと⑫の関連付けについて議論していた。⑦に関し

ては、まず⑩に関連があるのかわからないので意見が分かれた。ディスカッションを進める中で、学習者2が「⑦は、⑩と④に関連付けられるのではないかと意見があり、これにより⑦は⑩と関連があり、また、⑦を独立して考え、⑩→⑦→④と関連付けた。また、⑫に関しては、⑪と関連があると考えたが、LFGの記述方式



上の注意として初期前提課題に有向枝を繋げてはいけないので、⑩→⑨→⑥→④と組み立てた機械下部の部品に⑫を組み付けるという考えが合意した結果、⑩→④という関連付けを行った。このグループでのLFGはこの⑫からの関連付けを誤理解しただけの正解に最も近似したLFGとなった。

#### ● グループ2

グループ2では、3者とも個別のLFGに比べグループのLFGの方が評価が高くなっていることがわかる。図5に各学習者のLFGとグループでのLFGを示す。各学習者のLFGを見てみると、学習者4は、⑩から3つに分かれて有向枝を伸ばしている。⑩から⑦については理解しているのだが、⑩→⑤、⑩→③は、構造上有り得ない関連付けをしており誤理解をしている。また、⑪→⑨、⑪→⑥も同様のことが言える。これは、LFGの描画方法をしっかりと理解していないと言える。次に学習者5は、機械の上部と下部の2つに分けて考え各部品を組みつけている。しかし、⑫→⑩の様に初期前提課題から初期前提課題に関連付けを行っている。これは、LFGの描画方法の誤理解である。これは、学習者6にも言えることである。学習者6は、⑩→②という誤理解をしている。この部分に関して学習者6はスプレーローガンの構造を把握していないことがわかる。

次にグループのLFGについて考察する。まず、⑩→⑪→⑨→⑥→④→①の部分は、まず最初に学習者4が「⑫は①に直接関連付けするのではないかと意見をしたが、他の2人が同じ考えだったため2人の意見に合意し、⑫→⑩の間違った関連付けをしてしまった。次に、⑩の関連付けに関しては、学習者4が「同時に組み付けが出来る部品は組み付けたほうがよいのではないかと2人に意見し、そこから再度⑩の関連付けを考えた結果、⑩から⑧、⑩から⑦と二手分かれ⑧→⑤→③→②と関連付け①に関連付け、⑦はそのまま①に関連付けるということで合意し形成された。これによりグループでのLFGの類似度が上昇したと考えられる。

#### 4.2 評価が下がったグループに対する考察

ここでは、評価が下がったグループについて考察を行う。表2のグループ6について考察を行う。グループ6では、学習者18に関しては類似度は上昇しているが、学習者16、17に関しては下降していることがわかる。図6に各学習者のLFGとグループでのLFGを示す。まず各学習者のLFGについて見てみると、学習者16は各初期前提課題から順々に部品に関連付けている。⑩の部分を見てみると③から二手に分かれ①に関連付けている。学習者17も同様に初期前提課題から順々に関連付けを行っている。学習者17は学習者16

とは違い、⑩から二手に分かれ関連付けを行っている。最後に学習者18は、初期前提課題から順々に関連付けをし、各部品を一塊にし①に関連付けている。また、この学習者は、設問にある「組立プロセスを考える際は、④と①の隙間が小さいため①に④を直接入れることは困難であることに注意しなさい。」という説明から最後に④スプリングに関連付けをした後に①ボディに関連付けしたと推測できる。

次にグループのLFGについて考察を行う。グループでのLFGでは、各部品を一塊にして①に関連付けている。類似度の高い2人の学習者16、17に関しては、⑩からの経路、⑫からの経路は教授者のLFGと一致している。にも関わらず、それがグループのLFGに反映されなかった原因として、学習者16、17が学習者18の意見（各部品を一塊にして①に関連付ける考え方）に合意してグループのLFGが描画されたからだと考える。これにより、評価が下がったものと考えられる。

以上、評価が上がったグループと下がったグループそれぞれについて考察を行った。これより、LFGを用いたグループ学習の学習効果が期待できる。また上記の結果から今後本手法を実施する場合、各グループに対して、教授者がアドバイスしたり、ヒントを提示しながら適用することでより学習効果が上がると期待できる。

#### 1. おわりに

本稿では、LFGを用いたグループ学習を提案した。またそれについて本学機械システム工学科の授業「システム設計Ⅱ」で実施したスプレーローガンの組立プロセスにおけるLFTを実施し、学習効果の検証を行ったのでそれについて述べた。検証からLFGを用いたグループ学習において学習効果が期待できると結論付けた。今後は、上記に示したことを踏まえた上で、他の事例においてもLFTを適用し、グループ学習の学習効果を検証していく予定である。

#### 文 献

- [1] 竹谷誠,佐々木整:"学習者描画の認知マップによる理解度評価",信学論 D-II,vol.j80-II NO.1 pp.336-347(1997).
- [2] 米澤宣義,新藤康正,平井和人,竹谷誠:"学習課題の関連構造を評価するための差異度",信学論 D,vol.J89 NO.5-D pp.967-978(2006).
- [3] 千葉樹里,小林政尚,佐々木整,竹谷誠:"認知マップを利用した合意形成型グループウェアシステム(Ⅱ)",電子情報通信学会 1996 年情報・システムソサイエティ大会講演論文集,D-488,pp.491(1996).
- [4] 富田聡,植田光,中川麻衣子,小倉智幸,木原幸一郎,磯野宏秋,杉林俊雄,近藤秀文,竹谷誠:"機械設計授業におけるバグ修正アルゴリズムの有効性の検証",電子情報通信学会 2008 年総合大会講演論文集,D-15-21 pp.215(2008).
- [5] 神保智広,富田聡,近藤秀文,竹谷誠:"ロジカルフローグラフを用いたグループ学習の学習効果測定と評価",日本教育工学会第24回大会講演論文集,2a-A206-07,pp.509-510(2008).

# ロジカルフローグラフを用いたグループ学習の学習効果と評価

Measurement and Evaluation of Group Learning Effects by Using Logical Flow Graphs

神保 智広 富田 聡 近藤 秀文 竹谷 誠

Tomohiro JIMBO, Satoshi TOMITA, Hidehumi KONDO and Makoto TAKEYA

拓殖大学 工学部

Faculty of Engineering, Takushoku University

## 1 まえがき

筆者らは、課題に対して、問題解決のプロセスを有向グラフによって表現するロジカルフローグラフ(以降、LFGと呼ぶ)を提案し、LFGを用いて構造的理解度を測定するロジカルフローテスト(以降、LFTと呼ぶ)法を提案した[1]。また10年前より、機械システムの組立図と部品図から、システムの組立手順を構造的に表現する方法にLFGを適用し、LFTによる構造的理解度評価法の教育効果を報告した[2]。

本稿では、機械システム設計の授業で、LFTをグループ学習に適用し、学習効果の測定評価を行ったので報告する。

## 2 グループ学習の実施と測定

まず、本グループ学習の方法について述べる。対象の学生は、本学機械システム工学科3年生で、実践授業は「システム設計II」である。LFTの課題は、スプレーブローガン(塗装用スプレーガン)の組立順序を組立図と部品図から読み取り、LFGで記述する課題である。問題用紙と回答例を図1に示す。はじめに学生個人で課題に取り組み、その後学生をグループに分け、グループ学習を行う。グループ学習では、まず各学生がどのような考えでLFGを描画したかを発表し、その後議論を行い、最終的に全員が合意したLFGを描画する。またグループ学習の際、議論のプロセスを記録することは重要であるため、ビデオ撮影を行い、グループ内の各学生の発話のプロトコル解析を行い、LFG描画プロセスを考察した。

## 3 グループ学習の学習効果の評価と考察

全学生27人を9グループ(各グループ3人)に分け、ビデオ撮影は5グループに対して実施した。また構造的理解度は、LFGの類似度[3]により評価した。学生個人のLFGの評価と比較して、グループのLFGの評価が下降している学生は27人中6人(22%)であった。4人に1人はグループ学習効果が得られなかったと考えられる。この結果より、LFGを用いたグループ学習の学習効果が期待できるが、学習効果が得られなかった学生への対策を検討しなければならない。対策としては、グループ編成に工夫をこらしたり、各グループに対して、教授者がアドバイ

スやヒントを提示したりする方法が考えられ、これらは今後の課題と考える。なお、プロトコル解析の詳細は紙面の都合で割愛した。

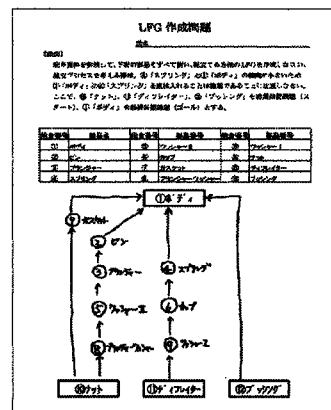


図1: スプレーブローガンの組立順序の問題回答例

## 4 むすび

本稿では、LFGを用いたグループ学習の方法と学習効果の測定評価法を提案した。一例ではあるが、LFGを用いたグループ学習の学習効果が期待できるとの結果を得た。今後は他の事例においてもLFTを適用し、グループ学習の学習効果を検証していく予定である。

末筆ながら、授業をご担当頂いた本学工学部機械システム工学科の杉林俊雄・木原幸一郎両先生には、テスト設計から評価までご議論頂きました。ここに謝意を表します。

## 参考文献

- [1]竹谷誠,佐々木整:"学習者描画の認知マップによる理解度評価",信学論D-II,vol.j80-II NO.1 pp.336-347(1997).
- [2]佐々木整,木原幸一郎,磯野宏秋,杉林俊雄,竹谷誠:"ロジカルフローグラフに基づく授業設計と改善-4年間にわたる継続的実践研究と有効性の検証"日本教育工学会論文誌,27,(4),pp.387-395(2003).
- [3]富田聡,神保智広,木原幸一郎,杉林俊雄,竹谷誠:"機械設計授業におけるグループ学習の学習効果の検証",電子情報通信学会信学技法,EIT2008-62, pp.23-28(2008).

# Development of An Algorithm for Groupware Modeling for A Collaborative Learning

Ikuo Kitagaki, Atsushi Hikita, Makoto Takeya, Yasuhiro Fujihara

**Abstract:** This paper reports an algorithm for forming groups of students with regard to a computer system for collaborative learning designed to give a cue for debate utilizing mobile terminals. With this system, questionnaires which should be used as the seeds for debates are prepared in advance on the Web and all students attending the class answer to the questionnaires on the Web through their mobile terminals. Following this, the computer assigns students to appropriate groups based on the results of answers, and transmits each of answers and information of the group member to terminals of the students. Based on the information, students form groups and each group starts debate. In this study, system composition is dealt with first and algorithm for forming groups using answers by the students is discussed.

**Keywords:** Group forming, algorithm, collaborative learning, mobile terminal, debate system, groupware

## 1 Introduction

This study relates to university class in Japan. In many cases, lectures are given in one-sided manner, and it is pointed out that some schemes are necessary to encourage students to express their opinions or to induce discussions by students themselves[5][8]. Considering the recent trend that almost all students carry their own mobile terminals[6][7], this study relates to a learning system that utilizes mobile terminals as an ancillary tools for debate. Students are divided into groups making use of the information of answers by the students against questionnaires presented beforehand. On this occasion, information necessary for grouping and results of grouping are transmitted together with contents of each of answers to mobile terminals of the students. Students are then requested to form groups based on these information and to initiate debates in each group. In other words, this system is one sort of blended learning in which face-to-face debate is carried out utilizing electronic educational information communication. In this study, algorithm for grouping based on information of answers by the students is dealt with. With collaborative learning as mentioned, two points can be expected as mentioned below.

1. Groups are being formed considering education effects. Therefore, every student can know why he/she is assigned to the current group.
2. Students are instructed to start debate using the information on their answers as the seeds. Therefore, they can easily grasp a cue for debates.

According to the conventional method of grouping of students, groups are formed simply mechanically in the order of student ID numbers or those seating nearby are assigned to the same group. Meanwhile, in this study, it is attempted to form groups appropriately based on answers against some sort of assignments including tests. Teaching by groups which were formed based on answers of the students was attempted by the prior study[2][3]. At the time of this study, since possession of a mobile terminal by every student was not practical, test results were collected and marked by the teacher, groups were formed based on the results. Test papers were returned to the students, groups were formed, and students in each of the groups are requested to discuss primarily incorrect answers. For this reason, it took enormous time from answering to the questionnaires by the students to group discussion. This time, fully making use of information technology available today, we configured a new collaborative learning

system. Thanks to assistance by mobile terminals, it is expected that from answering by the students till execution of discussions can be made promptly and smoothly. This system is currently under development. In Section 2, outline of system composition will be described, and an algorithm for forming groups will be introduced in Section 3 onward. Although the algorithm is one portion of the system, it is significant to make that in order to put the system into practical use, which may promote the practical e-learning in universities[1].

## 2 System Flow

Total four files are prepared and used in this system, student file, questionnaire file, aggregation file and groupware file. These files are used in this order.

1. Student file  
Basic information such as mail address, age, gender or the like of students are summarized in this file.
2. Questionnaire file  
Questionnaire or the like concerning discussion theme are summarized in this file. This file can be consisted of any of three types, multiple-choice method that allows only one selection, multiple-choice method that allows a plurality of selections, and free composition. (Multiple-choice method alone is subjected to processing of item 4. onward.)
3. Aggregation file  
Answers by the students for questionnaires in 2. are aggregated in this file.
4. Groupware file  
The computer forms student groups appropriately based on results of answers by the students in 3. This file also includes information on, in addition to answers by the students, which students form one group.

Execution procedures for preparing and utilizing above-mentioned files are shown in Fig.1.

- (1) A student accesses the predetermined URL, enters basic information such as gender, ID number or the like and transmits it. His/her file (EXCEL) is then completed while the teacher may execute input and deletion of the information by operating EXCEL directly. Basic information of a certain class is stored here and this file can be utilized at every round-table discussion held by the class.
- (2) With regard to a discussion planned by a certain teaching, a group of questionnaires which may benefit grouping is prepared and is registered to the system as a questionnaire file.
- (3) Several groups of questionnaires are being registered in the questionnaire file. The teacher picks up specific groups of questionnaires and transmits URL for browsing questionnaire items to mail addresses of all students.
- (4) Students answer the questionnaires and transmit them to create an aggregation file.
- (5) Based on the information in the aggregation file, students are assigned appropriately to a certain group.
- (6) Information on grouping is transmitted to mobile terminals of the students.

Upon completion of above-mentioned procedures, the students are grouped in the class and start discussions using results of answers as the cue.

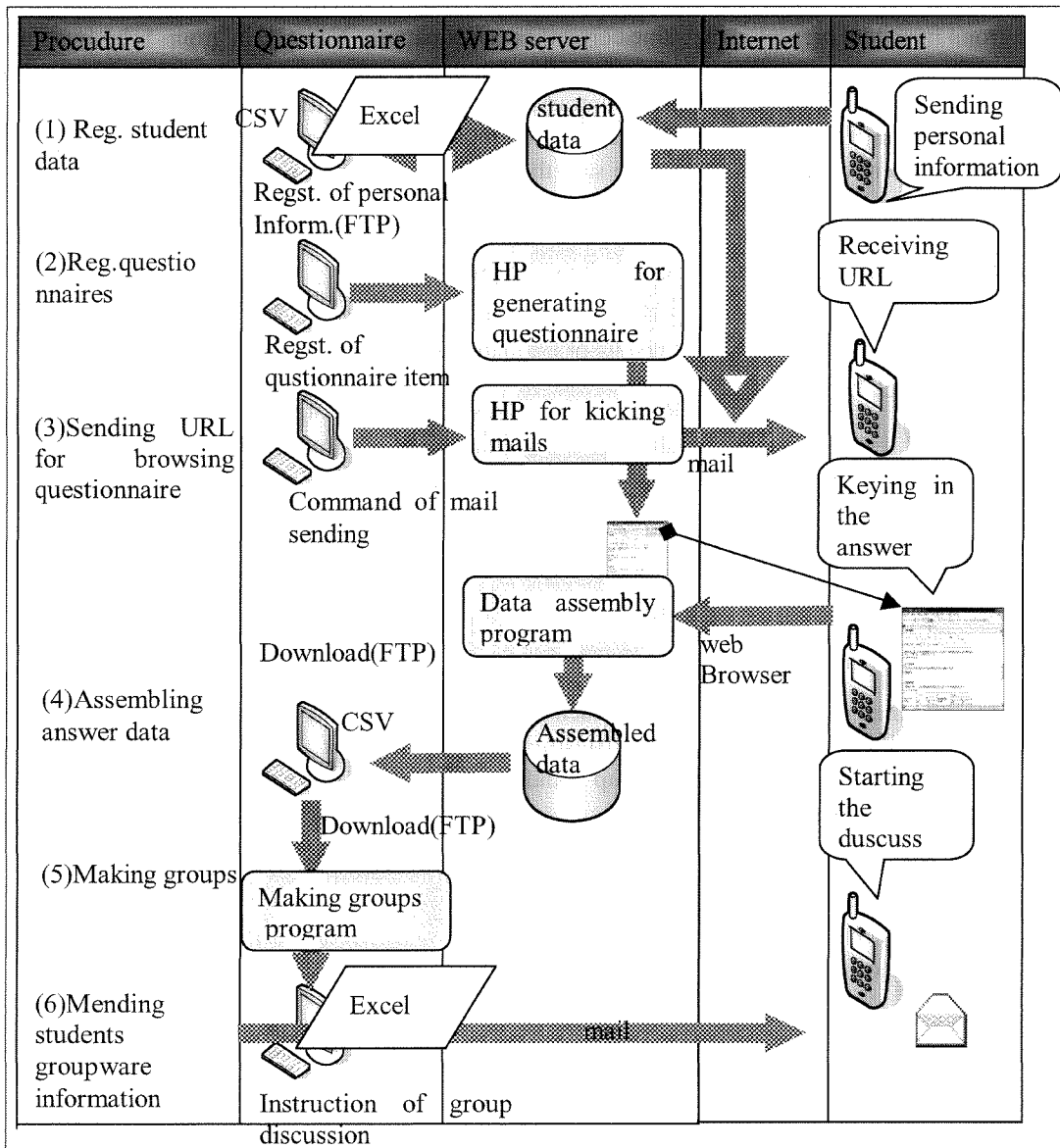


Figure 1: System flow

### 3 Construction of students groups

Suppose that there are four students A, B, C and D, and three questions a, b and c are given. Results of evaluation of answers to these questions are shown by O and X as shown in Table 1. One group consists of two students and discussion is made within the group primarily for questions which resulted in incorrect answers. In this case, two types of grouping shown in the table are compared. With grouping p, students A and B and students C and D constitute one group, respectively. However, in both groups, each of group members presented the same correct/incorrect pattern. Namely, in group 1, both members gave incorrect answer for question b and a possibility that they can teach each other for problem solving is low. The same also applies to group 2. Contrary, with grouping q, students A and C and students B and D constitute one group, respectively, and the said problem does not occur in this case.

From above-mentioned cases, it may be said that grouping q is superior to grouping p. When we focus on problem solving of correct answers and incorrect answers of the test, possibility of problem solving is considered to be higher for such a case where there are different patterns of answers rather than a case where there is only one pattern in the same group.

Table 1: Example of results of evaluation of answers and grouping

Student	Result of evaluation of question a	Result of evaluation of question b	Result of evaluation of question c	Grouping p (Group ID)	Grouping q (Group ID)
A	○	×	○	1	1
B	○	×	○	1	2
C	×	○	×	2	1
D	×	○	×	2	2

Suppose that the number of members in each of groups is the same. Then excellence of grouping  $u_g$  in a certain group  $g$  ( $g \in G$ ) is defined. Question  $j$  and question set are expressed by  $m_j$  and  $M$  ( $m \in M$ ), respectively, and student  $i$  in one group  $g$  and the student sets is expressed by  $s_i$  and  $S_g$  ( $s_i \in g$ ), respectively. Further, result  $e$  of evaluation of answer for question  $m_i$  and for student  $s_k$  are expressed by  $e(m_i, s_k)$ .

- Answer is correct  $\rightarrow e(m_i, s_k) = 1$
- Answer is incorrect  $\rightarrow e(m_i, s_k) = 0$

$$u_g = \frac{\sum_{s_k \in g} \sum_{m_i \in M} (1 - e(m_i, s_k)) \cup \sum_{s_j \in g} e(m_i, s_j)}{|M| |S_g|} \quad (1)$$

where,  $|M|$  denotes the number of elements of sets  $M$ . Grouping  $u$  of entire class is expressed by Equation (2)

$$u = \frac{\sum_{g \in G} u_g}{|G|} \quad (2)$$

Although two  $u_g$ 's obtained by equation (1) are comparable to each other only for the case that both groups consist of the same number of the students, the proposed idea in this paper will be able to be

extended to the different number of groups in a classroom by equation (3) where both equations (1) and (2) are combined:

$$u_g = \frac{\sum_{g \in G} \sum_{s_k \in g} \sum_{m_i \in M} (1 - e(m_i, s_k)) \cup_{s_j \in g} e(m_i, s_j)}{|M| |S_g|} \quad (3)$$

where N means the number of the students in a classroom. Actually there could be that several groups are different in the number of group members because the total number of the classroom cannot be well divided into an integer by appropriate integers.

Whether the number of the students in a group is equal or not, it is necessary to calculate equation (2) or (3) for all the combination of student group in order to obtained the maximum solution. Although it is possible to do that theoretically, it is difficult to obtain the strict solution because of the enormous number of calculations. In reality, if the number of group members is supposed to be 4, the computer can execute computations of a class consisting of 50 peoples at the most. Then more simplified method should be used. In other words, with the relevant algorithm, suppose that students of one class are re-expressed as  $s_1, \dots, s_N$ , the following replacement is made which results in calculations of N (N-1) times in total, and a group in which u becomes the maximum is judged to be the optimum grouping.

$$\begin{aligned} & s_1 \text{ and } s_2, s_1 \text{ and } s_3, \dots, s_1 \text{ and } s_N \\ & s_2 \text{ and } s_3, s_2 \text{ and } s_4, \dots, s_2 \text{ and } s_N \\ & \dots\dots\dots \\ & \dots\dots\dots s_{N-1} \text{ and } s_N \end{aligned} \quad (4)$$

It is already known that when one class consists of 20 and several members and one group consists of four members, excellence of grouping by the simplified algorithm as used in the current proposal is, compared with a case where the same is obtained for all combinations without simplification, approximately 60 ~ 100% optimization rate depending upon the answer pattern[4].

#### 4 Discussion in a classroom

Two kinds of discussion were administrated concerning the proposed method.

The first kind is the one for discussing test answers, where the evaluation either of correct or wrong was determined. Digital mathematics served as the subject. Several test problems were given to the students. All the test problems had multiple choice type of answers thus the students selected an answer among them. Because the evaluation was already done, the instructor, after group making, told them to start by telling the evaluated result then discuss the answer which was evaluated wrong.

The second kind is the one for discussing opinions, where the evaluation either of correct or wrong cannot be determined. Training program for job hunting served as the subject. The example of a question done by an interview and the answer by the interviewee was presented followed by several opinions for the answer then the students selected an opinion which they thought the closest to their impression. The instructor, after group making, told them to start by telling their own choice, discuss a better answer then make the report as group co-working.

Through several administrations, the method was thought to give the students learning motivation because, every time when the proposed discussions were done, group member changed for clear reason leading to their refreshment. In addition to that, we thought that, if the more detail direction in what point the discussion ought to be done was given to the students, they may think it easier to start.

## 5 Consideration

According to the algorithm for forming groups used in the current proposal, grouping is made so that results of evaluation of answers of the test by students may become different as much as possible. For the sake of establishment of more generalized algorithm, overall investigations are necessary which includes what sort of group discussions are actually available, what sort of information are effective for forming groups, what judgment criteria should be used to determine good or bad of grouping based on the said information. Although in the present study, results of evaluation of answers are digitized to either 0 or 1, the authors intend to investigate another algorithm capable of coping with more diversified evaluation results. At the same time, utility of this group discussion system and points to be improved will be checked from both software and hardware aspects through practicing.

## 6 Acknowledgements

The present study has been promoted under the grant partly from Grant-in-Aid for Scientific Research, germination study No.17650260, fundamental study (B) No.18300288, and germination study No.17300275, each sponsored by the Ministry of Education, Culture, Sports, Science and Technology. We are sincerely thankful to all the persons and bodies helping our works.

## References

- [1] Styliadis D. A., Karamitsos D.I. and Zachariou I.D: Personalized e-learning implementation-The GIS case *International Journal of computers, Commjunicaitons & Control*, Vol. 1,1, pp. 59-66, 2006.
- [2] Kitagaki I., Shimizu Y. and Suetake K., An instructional method which permits the students to critically discuss their own test answers, *Jour.of Educ. Technol.*, Vol.5, No.1, pp.23-33, 1980.
- [3] Kitagaki I., Shimizu Y., Consideration on an educational system which permits the students to discuss over their own test answers, *Res. Sci.Educ.*, Vol.5, No.1, pp.22-28,1981.
- [4] Kitagaki I, A Consideration on an Educational groupware Algorithm Using Fuzzy Integral, *IEICE, Tech.Rep.*,Vol.ET94-101, pp.61-66,1994.
- [5] Nagai M., Kitazawa T., Koshikawa H., Kato H., Akahori K, Eevelopment and Verification of the Formative Evaluation System with Utilizing Mobile Phone for Web-Based Collaborative Learning, *Japan Journal of Educational Technology* , Vol.28, No.4, pp.333-342, 2005.
- [6] Nagamori M., Ueno M., Ando M., Pokpong S., Endo K., Nagaoka K., Response Analyzer System Using Mobile Phones for Distance Education, *Japan Journal of Educational Technology* , Vol.29, Suppl., pp.57-60, 2005.
- [7] Otuska K., Yahiro T., The Input Interface of the Value of Evaluation in an Evaluation System of Instruction Using Mobile Phones, *Japan Journal of Educational Technology*, Vol.30, No.2, pp.125-134, 2006.
- [8] Ozawa S., Mochizuki T., Egi H., Kunifuji S, Facilitating Reflection in Collaborative Learning Using Formative Peer Evaluation Among Groups, *Japan Journal of Educational Technology* Vol.28, No.4, pp.281-294, 2005.



Ikuo Kitagaki  
Hiroshima University, Japan  
Graduate School of Education  
The research Institute for Higher Education  
1-2-2 Kagamiyama, Higashi-hiroshima, 739-8512, Japan  
E-mail: kitagaki@hiroshima-u.ac.jp

Atsushi Hikita  
Hiroshima University, Japan  
Community Cooperation Center  
E-mail: hikita@hiroshima-u.ac.jp

Makoto Takeya  
Takushoku University, Japan  
Faculty of Engineering  
Department of Computer Science  
815-1 Tate-machi, Hachioji, Tokyo, 193-0985 Japan  
E-mail: takeya@cs.takushoku-u.ac.jp

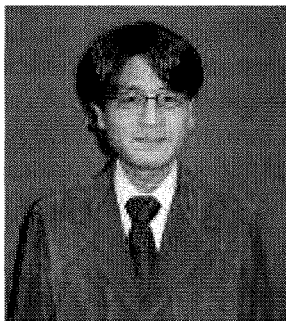
Yasuhiro Fujihara  
Iwate Prefecutural University  
Faculty of Software and Information Science  
fuji@iwate-pu.ac.jp

Received: November 7, 2006

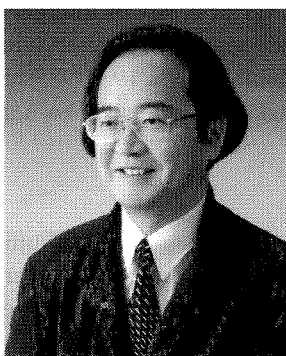
**Editor's note about the authors:**



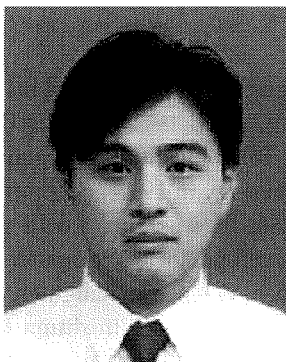
Ikuo Kitagaki, born in Aichi on August 9, 1947, received his BE and ME degrees in electronics in 1970 and 1972, and his Doctor Engineer degree in 1981 from Tokyo Institute of Technology. Joining Tokyo Institute of Technology in 1973 and Employment Promotion Corporation in 1986, he was engaged in various research areas including the development of computer applications in education, fuzzy science, science of laughter, and so on. Since 2000, he has been belonging to Hiroshima University, Hiroshima, Japan. Presently he is Professor of Graduate School of Education/The research Institute for Higher Education, Hiroshima University. One of his most important recent works is "University Authority (in Japanese)". His career is shown also in "Who's Who in Science and Engineering, MARQUIS 9th ed.(2006-2007), USA".



Atsushi Hikita (born on January 30, 1968) graduated the faculty of Physics of Sophia university, Japan in 1991, and got his master degree at the same faculty in 1993. He worked in Mitsubishi Research institute as a media and information planner and researcher (1993-2000). His main research fields are information design (information and its places, especially map and pictograms as a non-verbal communications), and media design (media communications for society, museum, library and classroom). He has co-authored and co-edited, several books and papers for these fields. He is a chair of APNG (Asia Pacific Networking Group) Education and Live E! WG, committee member of Chugoku-Shikoku Internet council, Hiroshima region IPv6 deploy committee.



Makoto Takeya, born in Tokyo on November 2, 1941, received his BE and ME degrees in applied physics in 1966 and 1968, and his Doctor of Science degree in 1981 from Waseda University. Joining NEC Corporation in 1968, he was engaged in various research areas including the development of computer applications in education. Since 1986 he has been belonging to Takushoku University, Tokyo, Japan. During 1992-1993 he was a visiting scholar of Department of Educational Psychology, University of Illinois. Presently he is Professor of Department of Computer Science, Takushoku University. Among his most important books are "A New Test Theory: Structure Analysis Methods for Educational Information" (in Japanese) and "Structure Analysis Methods for Instruction: Theory and Practice of Instructional Architecture, Design and Evaluation". He is the recipient of the 1976 Yonezawa Memorial Award from the Institute of Electronics and Communications Engineers of Japan, the 1989 Excellent Research Award from Behaviormetrics Society of Japan, the 1996 Winning Paper Award from Japan Society of Educational Technology, and the 1999 Engineering Education Award from Japanese Society for Engineering Education.



Yasuhiro FUJIHARA (born on January 8, 1971) graduated the Faculty of Education of Kobe University in 1993. Presently he is an assistant professor at Faculty of Software and Information Science, Iwate Prefectural University, Japan. His main research fields are educational technology (educational evaluation, e-learning) and ICT education. He is a member of Japan society for educational technology, and the institute of electronics, information and communication engineers.

# 理解構造と差異度に基づく学習者のグループ化アルゴリズム

米澤 宣義<sup>\*1</sup>

野本 竜哉<sup>\*1</sup>

平井 和人<sup>\*2</sup>

<sup>\*1</sup>工学院大学 情報通信工学科

<sup>\*2</sup>アイネス株式会社

**あらまし** 同一の教材と学習環境で学習しても各学習者の理解状態は異なる。実際に理解構造を描かせると実に様々な形の理解構造を描く。本論は、グループ学習を用いて教師の指導なしで理解構造を正しい構造に接近させる一つの方法を試みる。この方式においては、学習者のグループ構成法が重要となる。ここでは、グループ化は以前に提案した差異度を応用し、さらにグループの学習は1度だけではなく、繰り返しグループを再編してグループ学習する再グループ化を提案する。

**キーワード** グループ学習, グループ化, 差異度, 理解構造

## An Algorithm for Configuration of Learners' Groups Based on Structure of Learner's Understanding and Dissimilarity

Nobuyoshi YONEZAWA<sup>\*1</sup>

Tatuya NOMOTO<sup>\*1</sup>

Kazuhiro HIRAI<sup>\*2</sup>

<sup>\*1</sup>Department of Computer Science and Communication Engineering, Kogakuin University

<sup>\*2</sup>INES Corporation

**Abstract** The students draw actually various understanding structure despite the same lesson in the same learning environment, because each student in class has his own level of understanding. This paper proposes a method for guiding their understanding structure to the correct structure by group learning without teacher. In this method, it is very important to divide students into groups based on their understanding. Here the students are classified into groups by structural dissimilarity which we proposed before, and the students discuss the problem in each group. The learning by the grouping is not the only once, but repeated several times.

**Key words** group learning, grouping, dissimilarity, understanding structure

### 1. はじめに

同一の学習環境で同じ教材を用いて学習しても学習者一人一人の理解状態は異なる。実際、学習後に理解構造を描かせると、学習者は実に様々な形の理解構造を描く。本論は、各学習者の理解構造を教師の指導なしで正しい構造（教師が作成した教材構造）に接近させるグループ学習とグループ構成法（以後グループ化）を提案する。

協調学習やグループ学習における学習者のグループ化は、システムの評価あるいは学習の成果を大きく左右する重要な基礎要素技術であるにもかかわらず、意外とこのグループ化に関する研究は少ない<sup>(1)(2)</sup>。

文献1は、グループ学力の評価指標を定義することによりグループの学力を測定し、さらにはどの程度協調効果あるいはグループ効果があったかを調べる試みである。この研究はグル

ープ化に利用できる可能性もあるが未知である。また文献2は、学習者グループのあり方に焦点をあてて、電子掲示板のようにオープンな場で質疑応答するコミュニケーション形式の協調学習において学習者集団を動的にグループ化する手法である。しかしこれは数学や物理系の学習には適用できない。

本論は数学や物理系の問題解決を行うグループ学習のグループ構成法に関する研究である。グループ内の議論の内容は必然的に導出過程や因果関係が正しいかどうかの判断ならびに計算が正しいかどうかのチェックが中心になると思われる。したがってグループ学習では、学習者がお互いに導出過程を描いた理解構造を見せ合い、さらに統一見解をディスプレイに描きながらビジュアル的に議論を行うほうが分かりやすく、また議論も活性化すると考えられる。このように議論は理解構造をベースに行

うので、グループ構成は理解構造の差異度<sup>(3)</sup>を応用することにする。

教員はグループ化を行うが学習指導には一切関わらない方式を用いるので、全てのグループの結論が出次第、グループの再編を繰り返すことによって沢山の他学習者と議論して理解を修正する機会を与える再グループ化手法を提案する。

### 1. 諸定義

教材構造は、学習課題とその順序関係をそれぞれ頂点と有向辺に対応させた閉路のない非循環な有向グラフで表現される。各学習課題は解説と正答を含めた問題で構成され、前提学習課題で定義される式や用語から導出されるように記述してある。つまり学習課題の順序関係は導出過程の関係であり、教材構造は全課題の導出関係を表わしたものである。

学習者に教材構造を見せずに学習課題系列順で学習課題を提示して学習させる。全学習課題を提示し終わった後、正答を見ずに同一の問題に解答させる。

その際に学習課題と前提学習課題を線で結ばせ、最終的に学習課題全体の導出過程の関係図を描画させる形成的テストを行う。この形成的テストで得られる有向グラフを理解構造という。

各学習課題の問題は前提学習課題からのみ導出されるように記述してあり、学習者は必要な前提学習課題の個数等の情報を問題から読み取ることができる。したがって、学習課題を理解していれば形成テストで前提学習課題を正しく関連付けることができ正しい理解構造を描くことができる。しかし、教材構造に記述してある知識以外の知識による先入観や暗記等が学習中や学習後に混ざりこみ完全に学習課題が理解されていない場合がある。このような場合は、前提学習課題に正しく関連付けがなされないために理解構造を正確に描けないことが多々生じる。

本論は、この不完全な理解構造をグループ学習で教材構造に導く方法を提案する。

最初に有向グラフの用語を定義しておく。非循環な有向グラフを  $G = [V, E]$  とする。ここに  $V$  は  $G$  の頂点集合  $V = \{v_1, v_2, \dots, v_n\}$ 、 $E$  は有向辺集合である。集合  $A$  の要素数を  $N(A)$  で表せば  $n = N(V)$  である。ある頂点  $v_a$  と  $v_b$  に接続し、かつ  $v_a$  から  $v_b$  へ向かう有向辺があるとき  $e_{ab} = 1$ 、有向辺がない場合  $e_{ab} = 0$  で表わす。

ここで、教材構造と理解構造を次のように定義する。

**【定義 1】**教材構造を  $G_T$ 、学生  $S_i$  の理解構造を  $G_{S_i}$  とする。 $G_T$  と  $G_{S_i}$  は同じ頂点集合  $V_T$  を持ち有向辺集合だけが異なる非循環な有向グラフで、

$$G_T = [V_T, E_T], \quad G_{S_i} = [V_T, E_{S_i}]$$

$$V_T = \{v_1, v_2, \dots, v_n\}, \quad n = N(V_T)$$

とする。ここに、 $E_T$  と  $E_{S_i}$  はそれぞれ  $G_T$  と  $G_{S_i}$  の有向辺集合である。

**【定義 2】**学生の集合  $S = \{S_1, S_2, \dots, S_m\}$  に対する、理解構造の集合を

$$G_S = \{G_{S_1}, G_{S_2}, \dots, G_{S_m}\} \text{ とする。}$$

ここで、 $G_S$  の部分集合  $R \subset G_S$  をグループと呼ぶ。本論では  $N(R) = 3$  である。

いま、 $R = \{G_{S_j}, G_{S_k}, G_{S_l}\}$  のとき、多数決グラフを  $R^* = [V_T, E_{R^*}]$  と定義する。ここに  $E_{R^*}$  は

$R^*$  の有向辺集合で、 $e_{ab} \in E_{R^*}$  は、 $e_{ab}^i \in E_{S_j}$ 、

$$e_{ab}^k \in E_{S_k}, \quad e_{ab}^l \in E_{S_l}$$

に対する次の多数決演算の値

$$e_{ab} = e_{ab}^j e_{ab}^k \vee e_{ab}^j e_{ab}^l \vee e_{ab}^k e_{ab}^l$$

である。

図 1 を教材構造、図 2 の A, B, C を理解構造としたときの多数決グラフを図 3 に示す。

**【定義 3】**多数決により得た  $e_{ab} \in E_{R^*}$  が  $e_{ab}^j \in E_T$  に等しくなり有向辺が正しく修正された場合は、多数決による改善と呼ぶ。

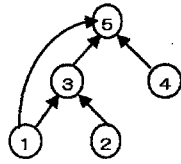


図1 教材構造

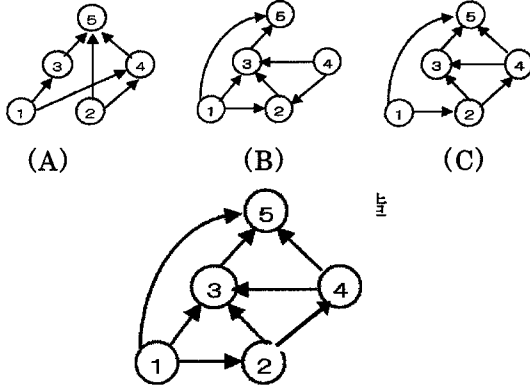


図3 図2から得られる多数決グラフ

また、 $e_{ab} \in E_R^*$  が  $e_{ab}^T \in E_T$  と等しくならず多数決により正しい結果が得られない場合は、多数決による改悪と呼ぶ。ただし、 $e_{ab}^i$ ,  $e_{ab}^k$ ,  $e_{ab}^l$  が全て1あるいは0の場合は、議論は起きないと考えられるので、その有向辺は多数決の対象から外す。

図3において、多数決による改善は  $e_{13}$ ,  $e_{14}$ ,  $e_{15}$ ,  $e_{23}$ ,  $e_{25}$ ,  $e_{45}$ , 改悪は  $e_{12}$ ,  $e_{24}$ ,  $e_{43}$  で、残りの有向辺は対象外である。

$S_j$ ,  $S_k$ ,  $S_l$  が  $e_{ab}^i$ ,  $e_{ab}^k$ ,  $e_{ab}^l$  について議論

した結果、 $e_{ab}^T$  に等しくなったとき、議論による改善と呼ぶ。等しくならなかったときは、議論による改悪と呼ぶ。

**【定義4】**文献3で教材構造  $G_T$  と理解構造  $G_{S_i}$  の可到達性に関する差異度ならびに先行性

に関する差異度をそれぞれ  $D_{TS_i}^+$  と  $D_{TS_i}^-$  として  $G_T$  と  $G_{S_i}$  の差異度を次のように定義した。

$$D_{TS_i} = \frac{1}{2}(D_{TS_i}^+ + D_{TS_i}^-) ; 0 \leq D_{TS_i} < 1 \quad (1)$$

ここに、 $D_{TS_i}$  は  $G_T$  と  $G_{S_i}$  の隔たりを表し、 $D_{TS_i}$  が小さい程  $G_{S_i}$  は  $G_T$  に接近する。

ここで、学習者  $i$  と  $j$  の理解構造  $G_{S_i}$  と  $G_{S_j}$  の差異度を式1の  $T$  を  $S$  に置き換えて

$$D_{S_i S_j} = \frac{1}{2}(D_{S_i S_j}^+ + D_{S_i S_j}^-) ; 0 \leq D_{S_i S_j} < 1$$

と定義する。

図4は図1の教材構造を中心とする同心円上に理解構造  $G_{S_i}$  を  $D_{TS_i}$  の大きさの順に配置したものである。中心に近いほど  $D_{TS_i}$  は小さい。理解構造間の矢印は、有向辺を一つ修正したときの理解構造の推移を示す。この矢印が示す理解構造の連続を修正系列<sup>(3)</sup>という。

$D_{S_i S_j}$  は、 $G_{S_i}$  と  $G_{S_j}$  の隔たりを表す。同じ修正系列上の  $G_{S_i}$  と  $G_{S_j}$  は接近するほど構造が類似し  $D_{S_i S_j}$  が小さくなる。また、 $D_{S_i S_j}$  が小さいと  $G_{S_i}$  と  $G_{S_j}$  は同一修正系列上にあるか、あるいは  $G_{S_i}$  と  $G_{S_j}$  の修正系列が接近している。

### 1. グループ構成アルゴリズム

グループは、グループ内の学習者に議論が生じ、その結果、グループの3人の理解構造の平均差異度が減少する学習者で構成することを原則的ルールとする。

このルールを満足するためには、同程度の出来具合で、しかも異なる考え方を持つ学習者で構成すればよいと考えられる。

これを差異度を用いて具体的に表現すると次のようになる。 $D_{S_i S_j}$  が小さい  $G_{S_i}$  と  $G_{S_j}$  を同一グループに入れると構造が類似しているため議論が活性化しない可能性がある。また、 $D_{TS_i}$  と  $D_{TS_j}$  が大きく異なる  $G_{S_i}$  と  $G_{S_j}$  を同一

グループに入れると大きな能力の違いから議論が活発化しない.そこで,最初一人目の  $G_{S_i}$  を決めて,この  $D_{TS_i}$  と同程度の値をもつ理解構造の中から  $D_{S_i S_j}$  が  $\alpha$  以上の  $G_{S_j}$  を選びグループを構成する. 3人目は,この3人の理解構造の多数決による改善の個数が改悪の個数より大きく,さらに多数決グラフの差異度が3人の理解構造の平均差異度を下回り,かつ  $D_{S_i S_k}$  と  $D_{S_j S_k}$  が  $\alpha$  以上となるような  $G_{S_k}$  を選ぶ. ここに  $\alpha$  は,全理解構造の  $D_{TS_i}$  の平均とする.

この操作を繰り返して3人ずつグループ化する.

図4でイメージ的に説明すれば,外の円上にある理解構造から中心に向かって,異なる理解構造を持ち,さらに多数決グラフの  $D_{TR^*}$  が小さくなるような学習者を3人ずつグループ化することを意味する.

文献4で図4の理解構造を上記ルールでグループ化し,各グラフの多数決グラフを求め,この多数決グラフがグループの各学習者の理解構造として,さらにグループ化を繰り返すシミュレーションにより学習者全体の平均差異度を段階的に減少させることができることを

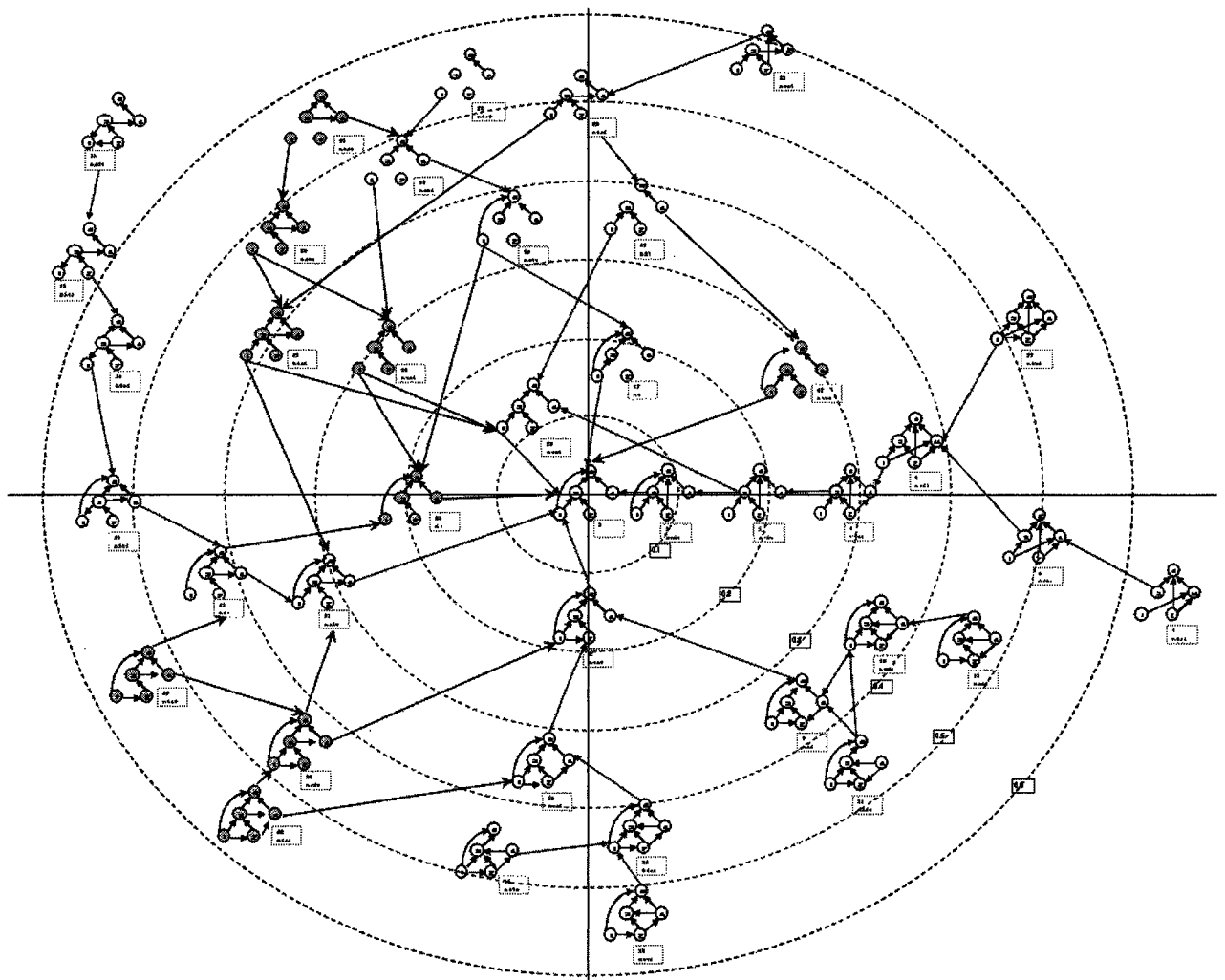


図4 教材構造を中心とした同心円状修正系列

示した。

実際にグループで議論し得られる結果は、多数決ではなく、少数意見が勝つ改善もあれば改悪もある。これは学習者の説得力やリーダーシップ等の総合的能力ならびに性格に左右されると考えられる。しかし、実際の学習においてもこの繰り返しグループ化の平均差異度の減少特性は、多数決法と同様に減少すると考えられるので実験による検証をおこなった。

### 1. 実験

被験者9人(学部2年生)に、直流回路に関する全頂点数17の教材を2章で示した方法で学習をさせた。その後、17頂点の内10頂点間の正しい有向辺を書いた構造を完成させる形式で各学習者に理解構造を描かせた。最初に提示した構造の差異度は0.497であるので、これを得点0, 差異度0を得点100に規格化した。

時間配分は、最初の教材による学習55分、問題の解答20分、理解構造作成15分、休憩15分(この間グループ化処理)、1回目グループ学習30分、休憩5分(2回目グループ化処理)、2回目グループ学習30分とした。1回目のグループ学習は個人が解答した問題と理解構造を参照して行う。2回目グループ学習はこれに加えて1回目の学習の結果を参照して行う。学習者には実験の意図と問題の正解や得点等の情報は一切教えない。

各学習者が描いた最初の理解構造の差異度とその得点(初期差異度、初期得点)ならびに1回目と2回目のグループ化を表1に示す。ただし、表は初期得点を昇順にソートし左端に最低得点から順番に学習者の識別番号を付けてある。また  $R_j^i$  は  $i$  回目の  $j$  グループを意味する。

表2は、1回目の各グループの3人の理解構造の学習前平均得点、グループの多数決グラフの得点(期待得点)、議論して得た実際のグループの理解構造の得点を示してある。表3は各

グループの多数決による改善の個数(期待改善個数)に対して、議論により期待通りに改善された個数と期待に反して少数意見が勝ち改悪された個数、ならびに多数決による改悪の個数

	初期差異度	初期得点	1回目グループ	2回目グループ
1	0.47	5.1	$R_1^1$	$R_1^2$
2	0.36	28.0	$R_1^1$	$R_2^2$
3	0.33	33.6	$R_1^1$	$R_3^2$
4	0.32	36.5	$R_2^1$	$R_2^2$
5	0.25	50.1	$R_2^1$	$R_1^2$
6	0.22	55.3	$R_2^1$	$R_3^2$
7	0.22	55.3	$R_3^1$	$R_3^2$
8	0.20	59.3	$R_3^1$	$R_2^2$
9	0.16	68.7	$R_3^1$	$R_1^2$

表1 初期得点とグループ化

	$R_1^1$	$R_2^1$	$R_3^1$	全体平均
学習前平均得点	22.2	47.3	61.2	43.56
多数決期待得点	43.5	56.5	77.2	59.1
実際の得点	58.9	66.8	75.1	66.9

表2 1回目グループ学習の学習効果

		$R_1^1$	$R_2^1$	$R_3^1$
多数決	期待改善個数	16	12	9
議論	期待通りに改善された個数	15	11	7
	期待に反して少数意見が勝ち改悪された個数	1	1	2
多数決	予測改悪箇所	4	5	1
議論	予測通り改悪された個数	2	2	0
	予測に反して少数意見が勝ち改善された個数	4	3	1

表3 1回目グループ学習における改善と改悪個数

	$R_1^2$	$R_2^2$	$R_3^2$	全体平均
学習前平均得点	66.9	66.9	66.9	66.9
多数決期待得点	78.3	78.3	78.3	78.3
実際の得点	68.7	69.1	64.1	67.3

表4 2回目グループ学習の学習効果

		$R_1^2$	$R_2^2$	$R_3^2$
多数決	期待改善個数	8	8	8
議論	期待通りに改善された個数	4	7	3
	期待に反して少数意見が勝ち改悪された個数	4	1	5
多数決	予測改悪個数	1	1	1
議論	予測通り改悪された個数	0	1	0
	予測に反して少数意見が勝ち改善された個数	1	0	1

表5 2回目グループ学習における改善と改悪個数

(予測改悪個数)に対して予測通り改悪された個数と予測に反して少数意見が勝ち改善された個数を示す。

$R_1^2$ と $R_2^2$ は表3より議論により期待通りに改善された個数と予測に反して少数意見が勝ち改善された箇所が多いため、多数決期待得点より実際の得点が大きくなり、実際の学習者全体の平均得点も増加し、グループ学習の効果があることがわかる。

2回目のグループ学習においては、表4が示すように学習前の平均得点より実際の得点の全体平均はわずかに増加しているが、全グループの実際の得点は多数決期待得点を超えな

った。 $R_1^2$ と $R_3^2$ は表5より議論により期待通りに改善された個数と期待に反して少数意見が勝ち改悪された箇所が同数に近く多数決期待得点より実際の得点が小さい。この原因としては2回目のグループ学習においては、初期の個人の学習結果と1回目のグループの結果の両方を参考にして議論するので混乱が生じるためであると考えられる。 $R_2^2$ に関してはよく改善されているが、重みの大きい有向辺1本が期待に反して少数意見が勝ち改悪され多数決予測得点に届かなかった。

### 1. むすび

サンプルが少なく断定はできないが、提案する繰り返しグループ化による学習は、学習者に議論をさせ、さらに全理解構造を学習の指導なしで教材構造に接近させる一つの有効な方法であると考えられる。

### 参考文献

- [1] 永岡慶三 植野真臣 “グループ学力の測定・評価—協調効果の計量—”, 日本教育工学会, 第22回全国大会講演論文集, pp469-470, (2006)
- [2] 井上智雄 重野寛 岡田謙一, “協調学習における掲示版ユーザの動的グループ化による情報アクセス性の向上”, 情報処理学会論文誌, Vo44, No.10, (2003), pp.2490-2494
- [3] 米澤宣義 新藤康正 平井和人 竹谷誠 “学習課題の関連構造を評価するための差異度”, 電子情報通信学会論文誌, Vo J89-D, No.5, (2006), pp967-978
- [4] 米澤宣義 野本竜哉 平井和人 “差異度に基づくグループ化アルゴリズム”, 教育システム情報学会, 第33回全国大会講演論文集, pp448-449, (2008)



# 自己組織化マップと完全情報最尤法から見た

## 平成 17 年度センター試験の教科学力構造

荘島 宏二郎, 大津 起夫, 田栗 正章

独立行政法人 大学入試センター

本研究では、2005 年の大学入試センター試験データにおける教科学力構造を自己組織化マップ (SOM) を用いて探索的に分析した。その後、完全情報最尤法 (FIML) を用いて、主要 16 科目の教科学力の平均・共分散・相関構造を計算した。分析の対象は、現役生・健常者・本試験受験者であった。SOM によって 9 つの学力層を取り出し、各層ごとに科目選択行動および得意科目が異なることが明らかになった。また、FIML による分析からは、教科学力構造の 2 因子性が示唆された。

**キーワード:** 教科学力, センター試験, 自己組織化マップ, 完全情報最尤法

Scholastic Achievement Structure of The National Center Test 2005

using Self-Organizing Map and Full Information Maximum Likelihood Method

Kojiro Shojima, Tatsuo Otsu, Masaaki Taguri

The National Center for University Entrance Examinations

The scholastic achievement structure in the data of the National Center Test 2005 was exploratorily analyzed using a self-organizing map (SOM). Next, the mean, the variance-covariance, and the correlation structure for the sixteen principal subjects were computed using the full-information maximum likelihood (FIML) method. The target population of the analysis was able-bodied high school graduands. Extraction of nine scholastic achievement layers using SOM, clarified that the subject-selection patterns and favorite subjects were different in each layer. FIML analysis also suggested that two factors which were abilities for liberal arts and sciences were assumed in the scholastic achievement structure.

**Keywords :** academic achievement, National Center Test, self-organizing map, full-information ML

### 1. 本研究で扱う学力

本研究では、2005年1月に施行された大学入試センター試験 (The National Center Test, NCT) において、受験者集団内に潜在する学力層を探索し、学力の相関構造を明らかにすることを目的とする。しかし、その前に、まず本研究で言うところの「学力」の定義を以下に説明しておく。

高浦 (2005) によると、我が国における「学力」の定義の仕方、学力とはどういうものかについての考え方、すなわち「学力観」には、戦後からのおおよそ 60 年の間で変遷が見られるという。その遷移を概観すると、1950 年代前半には「要素的学力」、後半には「概括的学力」が求められていた。1960 年代前半では、科学的な知識結果の体系的な獲得を目指す科学的な学力観が重視さ

---

Correspondence concerning this article should be sent to: Kojiro Shojima, The National Center for University Entrance Examinations, Komaba 2-19-23, Meguro-ku, Tokyo, Japan 153-8501. E-mail: shojima@rd.dnc.ac.jp

This research was supported by a Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science (C) No. 16650220.

れるようになり、60年代後半では「見かた・考えかた・感じかた」を含む科学等の基本概念や探究の仕方が重視された。さらに1970年代になると、より人間的な学力が求められ、知識過程も含む知識結果としての知識とともに「より内部に位置する能力」が重視されるようになった。さらに、平成10年の学習指導要領および平成12年の指導要録においては、新しい学力観として「生きる力」が取り上げられている。これは、平成8年7月の中央教育審議会の答申において、子どもたちが自ら学び自ら考える力などが、21世紀を展望した我が国の教育の在り方として重視されたからである。

このように、さまざまな学力観が提示されてきたが、時間を経るにつれて、意欲・態度などの情意面が重視されてきたように思われる。一方で、勝田(1962)の言うような「成果が計測可能なように組織化された教育内容を、学習して到達した能力」という学力観もあった。この勝田(1962)の学力観は、その後10年経過した頃から学力論争の火種となった(高浦, 2005)が、今でも市井一般に受容されている学力観の1つであろう。このような学力観は、「知能の定義」を巡って、F. N. FreemanやE. G. Boringが、「知能はテストによって測定されるもの」と定義した(e.g., 松原, 2005)ことに似ている。FreemanやBoringは操作的定義として、そのように知能を扱うことによって、測定することの根拠を確保したと捉えることができる。今日ではこのような考え方は、研究者の間では一般的に受容されていないが、巷間では、知能とは、知能テストで測られるものとする人々も依然として少なくないだろう。

つまり、勝田, Freeman, Boringの考え方は、学者や研究者の間では受け入れられていないが、一部の一般人に「信じられている」といったところであろう。少なくとも「知能や学力とは、テストで測定された知能や学力である」と胸を張って言えるように、テスト関係者が日々の努力を怠るべきでないといった教訓が彼らの学力観・知能観から得られるだろう。

さて、本研究では、2005年度のNCTにおける数十万人規模の教科得点の配列にはどのような特徴があるかを同定することが目的である。我々は、基本的には学習指導要領が定義するところの「生きる力を学力とする」という国家の教育政策と同様な立場をとり、勝田(1962)のように「センター試験で測定される学力が学力である」との立場はとらない。しかしながら、現行の学習指導要領の学力観は、幅広い概念であるので、センター試験で測定される教科別達成度得点も学力を構成する要素の1つであろう。そこで、本研究において

我々が報告する学力をとりあえず「教科学力」と呼んでおく。「教科学力」は、学習指導要領の「新学力」のうち、客観的測定を行いやすい能力の一群と見なすことができる。

## 2. ベールに包まれた大学入試センター試験

大学入試センター試験(NCT)は、全国の18歳人口のおおよそ3分の1が受験する大規模試験である。大学入試センター(2005a)によると、2005年は83の国立大学、76の公立大学が利用している。私立大学では408大学1,078学部が利用している。また、公立短期大学では15短期大学39学科、私立短期大学では104短期大学241学科が利用している(大学入試センター, 2005b)。

科目ごとの受験者数・平均点・最高点・最低点・標準偏差は大学入試センターのウェブサイトで公開されている。また、一部については、中畝・吉村・荘島(2004)および荘島・吉村・中畝(2004)のように、センター試験の分析結果が報告されている。また、吉村・荘島・杉野・野澤・清水・齋藤・根岸・岡部・フレイザー(2005)のように、NCT1990-2004の『英語』の経年比較を行っている研究もある。しかし、大規模試験であるNCTのある程度詳細な分析結果が公表されることは非常にまれである。特に、科目間の相関構造については報告されていない。TIMSS-RやPISAによる学力調査の結果をもとに学力低下論争が行われている昨今、18歳年齢の教科学力構造がどのようになっているかは、国家的な関心事であるべきである。

なぜ、そのような研究・分析がされなかったのかについては、情報保護の観点からの制約もあるが、それ以外に2つの技術的な困難があったと思われる。以下にそれらについて説明する。

### 2.1. 計算機に関わる問題

1つには計算機の問題点がある。センター試験外国語教科の『英語』は最も受験者数が多く、2005年の本試験では521,059人であった(大学入試センター, 2005c)。50万件以上のデータを手軽に扱えるようになったのは、最近の5~10年くらいである。そのような大規模データを、1台のパーソナルコンピュータのメモリに載せて分析することは、つい10年前までは難しく、したがって、大型計算機を用いて基本統計量を算出するなどの単純な作業しか行えなかった。しかし、この10年間ほ

どで、計算機の性能が飛躍的に進歩して、これまで困難だった大規模データのハンドリングなどが、1台のパーソナルコンピュータ上で比較的手軽に行えるようになってきた。

また、それに伴って、マーケティング・サイエンスの分野などにおいて、大規模データを分析するための洗練された手法群がソフトウェアに実装されてきた。特に、データマイニング用のソフトウェアであるClementine (SPSS Inc., 2005) やVisual Mining Studio 5.0 (株式会社数理システム, 2005) など、データの前処理・加工・分析・探索などを統合的に行えるプラットフォームの登場は大きいだろう。大規模データの分析は、主にデータの前処理・加工に多くの時間と労力が費やされるからである。

## 2.2. 欠測値の問題

また、センター試験データの書式から由来する分析の困難さがある。というのは、2005年センター試験は6教科32科目の中から、受験生は、受けた科目を選択するアラカルト方式が採用されているからである。注釈1) すなわち、非常に複雑な欠測パターンが出現する。そして、その欠測パターンは、完全ランダム欠測 (missing completely at random, MCAR; e.g., Little & Rubin, 1987; 岩崎, 2002) を仮定して、リストごとの削除あるいはペアごとの削除を行うなどの欠測処理が許されない。それらは、MCARを仮定できないとき統計量の推定値にバイアスが生じるからである。

なぜ、MCARを仮定できないか? それは、一般的に、よくできる (教科学力が高い) 受験生は、5教科以上を受ける傾向があり、選択科目数が少ない受験生は、教科学力が低い傾向がある。このようなとき、他の教科の得点の高低が欠測したか否かに関する情報を持っている可能性があるため、MCARを仮定できない。したがって、MCARのもとでは許されるいかなる方法も用いることができない。少なくともランダム欠測 (missing at random, MAR) のもとでの分析を行う必要がある。なお、MARのもとで不偏な方法である完全情報最尤法 (full-information maximum likelihood, FIML; e.g., Arbuckle, 1996; Bock, Gibbons & Muraki, 1994; Jöreskog & Moustaki, 2001; Lee, Poon & Bentler, 1990; Muraki & Carlson, 1995), EMアルゴリズム (Dempster, Laird & Rubin, 1977; 渡辺・山口, 2000) を用いた補完法, 多重代入法 (Rubin, 1987) などは、市販のソフトウェアに実装されるようになり

つつある。

無論、欠測が、情報のある欠測 (informative missing, IM) である可能性もある。しかし、現実にそのときの欠測構造を同定することは難しく、また、多変量データの場合には、変数の数が大きいほど、欠測の情報を他の観測された変数の値が保持している期待が高まるので、MARの過程は必ずしも強すぎる仮定とは考えられない。つまり、IMのもとで分析する必要性を頭の片隅に持ちつつも、ソフトウェアの実装状況を考えてMARを仮定するのが現実的ではなからうか。

しかしながら、MARのもとで不偏な手法群も理論的には、1970年代から行われているが、ソフトウェアで手軽に行うことができるようになったのは、やはり最近の10年くらいだと思われる。したがって、センター試験の詳細な結果がこれまでにあまり報告されなかったのは、MCARのもとでの結果しか報告できなかった当時の計算機状況を考えると、バイアスがあると分かりきった情報を提出することに二の足を踏んでいたことにも起因していたと思われる。

## 3. 目的

本論文では、2005年の大学入試センター試験データ (以降、NCT2005と略記) における教科学力構造を自己組織化マップ (self-organizing maps, SOM; e.g., Kohonen, 1995; 徳高・岸田・藤村, 1999; 徳高・藤村・山川, 2002) を用いて探索的に分析した後に、完全情報最尤法 (FIML) を用いて、主要科目の教科学力の相関構造を報告する。その際、現役生・健常者・本試験受験者のデータを対象とする。

なぜ、このようなデータを取り上げ、上述したような分析法により検討を行うのかについて、以下で説明する。

### 3.1. なぜ、現役生・健常者・本試験受験者が対象か

まず、なぜ、現役生を対象とするかである。これは、高等学校あるいは文部科学大臣が認める高等学校卒業相当の学生たちが、いま、現実にどのような教科学力構造であるのかを最初に同定することを優先する必要があると考えたからである。浪人生の教科学力構造は、現役生のそれとはまた異なる構造を有していることは十分に考えられる。したがって、現浪を一緒にするよりも現浪別に分析したほうがよいと思われた。無論、NCT2005では、全受験志望者のうち、4分の1の

およそ14万人ほどが浪人生であるので、浪人生集団そのものも興味の対象になりうる。また、両集団を分析した上で、現浪比較を行うことは大変有意義であるので、今後の課題となりうる。なお、一般的には、浪人の方が各教科科目において成績がよいことが知られている。

また、健常者のデータに限定するのも上記と同様の理由であり、まずは、健常者の教科学力構造を明らかにしたいと考えたからである。なお、身体障害者は全体の受験志望者のうち、400名程度おり、当該集団の教科学力構造を探索することも急務であろう。

また、本試験受験者を対象とするのも、追・再試験と本試験が同じテストでないため、基本的に比較不可能であることが理由である。注釈2) ただし、追・再試験受験者は全体のうち100名程度であり、このデータを除外することによる影響はほとんどない。

### 3.2. なぜ、SOMを用いるか

SOMは、個々の標本(入力ベクトル)を2次元格子上に布置していくクラスタリング手法の1つである。テストデータにSOMを適用した例を見あたらないが、本研究でSOMを用いる理由は、いくつかある。それらを列挙すると、①探索的データ解析に適していること、②モデルが柔軟であること、などである。

SOMは、基準変数(あるいは、教師信号、出力変数)のないニューラルネットワークモデルと言われることもあり、モデルが前提とする仮定が緩やかであるために、モデルを柔軟にデータに適合させることができるので、探索的データ解析に向いている。また、多次元の情報を2次元に圧縮して可視化するという性質を持っているため、データからの知識抽出が容易である。このようにSOMを適用する理由は、データの中でどのようなことが起こっているかについて、分析者がアプリオリに用いるモデルによってなるべく限定されない知識を抽出したかったからである。

なお、SOMにおける欠測値の取り扱い、入力ベクトルのうち、欠測箇所を入力信号がない状態として処理している。

### 3.3. なぜ、FIMLによる相関構造か

2.2節で先述したが、ペアワイズ削除など、暗黙にMCARが仮定されたもとで報告される相関構造は、MCARを仮定することが困難だと思われる欠測構造を持つNCTデータにおいては、結果を誤って解釈する恐

れがある。したがって、MCARの仮定を緩和させて、MARの仮定のもとで相関構造を検討する必要がある。FIMLは、MARの仮定のもとでは、平均・分散・共分散・相関などの推定値が不偏であることがわかっている。

大学入試センター試験が発表している各科目ごとの平均点や標準偏差(SD)は、当該科目を受験した人数から算出されている。しかしながら、一方で、我々は、「仮に全受験者が当該科目を受験したときの平均点やSD」を知りたい。そのとき、仮に欠測構造にMARを仮定できるなら、FIMLを用いることは、それらを知るために有効である。なお、仮に、欠測構造がMCARを仮定できるのであれば、そのとき、FIMLを用いてMARのもとで分析した結果と、MCARのもとでペアワイズ削除などを用いて報告された結果は理論的に一致する。また、先述したように、本質的に欠測構造がIMである可能性を頭の片隅においておいたほうがよい。

SOMは、柔軟であるがゆえに豊かな解釈が可能な結果を得ることができるが、分析的に詳細な結果を返すことには不得手である。しかしながら、FIMLによる分析は、モデルが前提とする分布などの諸仮定を通して観察されるので、数学的仮定の明証性が高いために、精緻な結果を得ることができる。したがって、多変量正規分布が仮定され、また、MARが仮定された下で得られる相関構造は、SOMの結果と相補的に利用することによって教科学力という現象を立体化することができると思われる。

## 4. 基本統計量と科目選択パターン

本研究で用いたNCT2005データの6教科32科目の基本統計量を表1に示す。ただし、教科の欄は、センター試験の時間割に即して9つの分類としている。当該データは、NCT2005のうち、1科目以上受験した現役生・健常者・本試験受験者の385,892人を対象としている。なお、以降、NCT2005といった場合、当該データを示すものとする。また、もともと圧倒的に健常者と本試験受験者が多い集団であるので、受験生の分類を最も規定している属性が現役生となっている。したがって、以降、対象とする受験者集団を単に「現役生」などと言ったりする。

表1中、混乱のない範囲で教科名・科目名を省略している。たとえば、『数①』は正しくは『数学①』であり、『国Ⅰ・Ⅱ』は正しくは『国語Ⅰ・国語Ⅱ』などである。

表1：6強化32科目の基本統計量  
(現役生・健常者・本試験受験者)

教科	科目	code	度数	Min	Max	Mean	SD
国語	国Ⅰ	P1	33,834	0	200	115.0	36.4
	国Ⅰ・Ⅱ	P2	331,487	0	200	117.8	32.5
地理歴史	世史A	Q1	1,432	6	100	43.2	18.6
	世史B	Q2	72,150	0	100	61.5	20.1
	日史A	Q3	3,598	10	100	54.2	17.1
	日史B	Q4	117,781	0	100	58.0	16.2
	地理A	Q5	5,024	5	100	64.7	15.3
	地理B	Q6	79,160	0	100	69.0	15.0
公民	現社	R1	148,962	0	100	69.2	12.4
	倫理	R2	35,134	0	100	65.9	14.3
	政経	R3	47,910	0	100	62.9	16.6
数①	数Ⅰ	S1	10,201	0	100	46.9	22.9
	数Ⅰ・A	S2	269,873	0	100	67.5	22.9
数学②	数Ⅱ	T1	9,321	0	100	38.7	15.7
	数Ⅱ・B	T2	234,362	0	100	50.8	18.3
	工業数理	T3	72	6	100	60.5	22.3
	簿記	T4	969	10	98	55.5	17.7
	情報	T5	428	9	100	54.1	18.5
理科①	総合理科	U1	63,070	0	100	47.6	14.5
	物理ⅠA	U2	1,187	7	100	65.4	15.5
	物理ⅠB	U3	98,600	0	100	57.9	20.2
理科②	化学ⅠA	V1	4,808	7	100	62.5	17.7
	化学ⅠB	V2	143,464	0	100	63.7	22.1
	地学ⅠA	V3	2,574	12	100	54.4	16.4
	地学ⅠB	V4	12,428	0	100	64.0	20.5
理③	生物ⅠA	W1	5,929	9	100	55.6	17.3
	生物ⅠB	W2	131,157	0	100	50.2	14.8
外国語	英語	X1	382,729	0	200	113.3	37.7
	独語	X2	36	29	200	136.6	47.9
	仏語	X3	94	47	200	138.9	40.2
	中語	X4	263	38	200	174.7	36.5
	韓語	X5	54	56	197	161.0	31.3

また、codeとは、科目名をコンパクトに分類するために便宜的に割り当てたものである。科目によって、受験者数のバラツキがあることがわかる。また、国語と外国語は、200点満点となっている。表1の結果は、受験した現役生たちの平均点とSDを正しく報告しているが、「仮に全受験生が当該科目を受験したならば、そのような平均とSDにならない」ことに注意してほしい。

次に、表2に科目選択パタンの集計結果を示す。実際には、385,892人の現役生たちは、さまざまな科目選択方略をとっており、実に、4,907通りの科目選択パターンが観察された。ここでは、最もよく観察された上位20パターンを挙げておく。

表2：上位20通りの科目選択パターン

科目選択パターン	度数	%	累積%	Mean
P2 Q4 X1	40,218	10.42	10.42	54.3
P2 Q2 X1	20,890	5.41	15.84	57.0
P2 Q6 S2 T2 U3 V2 X1	19,248	4.99	20.82	68.9
P2 Q6 R1 S2 T2 U3 V2 X1	15,767	4.09	24.91	66.2
P2 Q4 R1 S2 T2 W2 X1	11,139	2.89	27.80	63.1
P2 X1	9,302	2.41	30.21	46.8
P2 R1 S2 T2 U3 V2 X1	8,113	2.10	32.31	64.8
P2 Q4 R3 S2 T2 W2 X1	8,095	2.10	34.41	62.8
P2 R3 S2 T2 U3 V2 X1	7,418	1.92	36.33	64.7
P2 Q2 R1 S2 T2 W2 X1	7,416	1.92	38.25	64.8
P2 R3 X1	6,829	1.77	40.02	52.5
P2 Q6 S2 T2 V2 W2 X1	6,706	1.74	41.76	67.7
P2 S2 X1	5,902	1.53	43.29	48.7
P2 R2 S2 T2 U3 V2 X1	5,506	1.43	44.71	68.9
S2 T2 V2 X1	5,408	1.40	46.12	56.4
S2 T2 U3 X1	5,381	1.39	47.51	50.5
P2 Q4 S2 T2 U3 V2 X1	5,288	1.37	48.88	68.5
P2 Q6 R1 S2 T2 V2 W2 X1	5,005	1.30	50.18	64.8
P2 Q2 R3 S2 T2 W2 X1	4,510	1.17	51.35	65.0
P2 Q4 R2 S2 T2 W2 X1	4,488	1.16	52.51	64.4

表2中、それぞれ、左から9教科に対応して選択された科目コードが示されている。空白は、当該科目を受験しなかったことを表している。たとえば、385,892人の現役生のうち、40,218人は、『国語Ⅰ・国語Ⅱ (P2)』『日本史B (Q4)』『英語 (H1)』の3科目だけを受験したということである(志願ではなく、実際に解答したパターン)。また、Meanは、当該科目選択パタンの科目数の影響を除いた得点の平均値である。たとえば、一番上の行の科目選択パターンを選択した40,218人は、3科目平均が54.3点だったということである。その際、『国語』と『外国語』は100点満点で換算してある。科目によって難易度が異なるので、直接比較可能ではないが、一般的に、科目選択数が多いパターンほど、平均点が高い傾向が見て取れる。

このような受験者たちの科目選択行動は、大学・短期大学の進路選択と密接にかかわっていると考えられる。たとえば、センター試験を利用している私立大学・短期大学のうち、センター試験のいくつかの科目を指定している学部・学科が少なくない。科目選択パターンと大学・短期大学進路選択の関連を調べることも今後の1つの研究の方向となりうるが、本研究では、教科学力の相関構造を探ることを目的としているため、これ以上は立ち入らない。ここでは、表2のような4,907通りの欠測パターンが生じており、その欠測パターンが教科学力の高低に密接にかかわっている可能性があることを示唆しておくにとどめる。

### 5. 自己組織化マップの結果

本節では、SOMによる結果を示す。SOMでは、欠測があるときは、各格子点における参照ノードとデータとの距離を計算するときに、欠測箇所を省略した次元で距離を計算する。単純な欠測処理であるが、大きなマップを用意したとき、類似した欠測パターンの類似したデータは、マップ上で近い位置を占めるという基本は不変である。また、高次元の情報を2次元に圧縮してコンパクトに視覚的に表現してくれる利点がある。

ソフトウェアは、Viscovery SOMine Plus 4.0 (Eudaptics Software GmbH, 2001) を用いた。2次元格子は、六角形のユニットを用いた $131 \times 151 = 19,705$ の比較的大きいマップであった。多くの受験者をマップ上に布置していくためであるが、20,000個以上のノードをSOMineでは仮定することができないこともある。また、50,000件をランダムサンプリングして入力データとした。これは、SOMineの制約であるが、およそ、38万件のデータに対して5万件の標本抽出は、一般的と考えられる。張力 (tension) は初期設定の0.5とした。張力とは、学習効果を及ぼす近傍の半径のことであり、値が大きいほど、ノード間の参照ベクトルが影響し合い平均化してくる。これらの値は、何度かの試行錯誤の後に決定したものである。46回の最適化計算サイクルの後、SOMにおけるWard法によってソフトウェアが支持する最適なクラス数は3のとき群を抜いて適合がよかった。得られたマップを図1に示す。

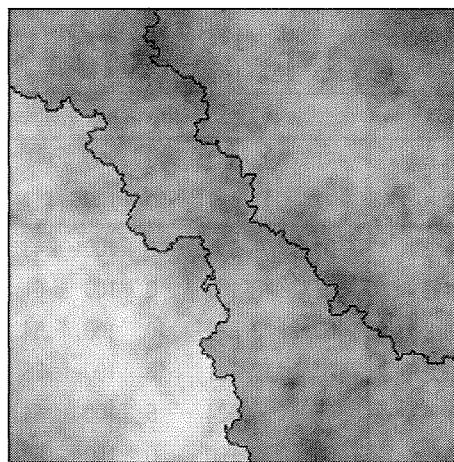


図1：層数3の結果 (SOM-Ward法による)

図における肌理は、濃いほど受験者の密度の高さを表している。また、左下から右上に行くに連れて、いわゆる「教科学力」が高・中・低グループとなっている。このクラスタ数が最も支持されたことによって、あたかも一般学力なるものがあって、それによって、現役生を層別化できそうな期待が出てくるが、38万人 (データは5万件だが) の現役生を3つのクラスタに分けることは、いささか乱暴である。したがって、ほぼ次点で同順位であった、ソフトウェアが支持したクラスタ数である4, 7, 8, 9のうち、説明の明証性と解釈可能性を考慮してクラスタ数=9の結果を選択し、図2に示す。また、便宜上、クラスタにI~IXの番号を割り当てた。

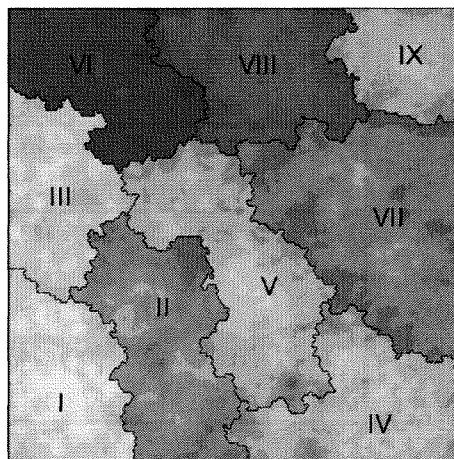


図2：層数9の結果 (SOM-Ward法による)

ところで、ここでいうクラスタ化とは、1つの集団を異なる小集団に分類するというよりは、同質な1集団を上位から下位へと「層別化」するとの意味合いが強い。

実際、この場合には、明示していないが全体の分散のかなりの部分を相関分散によって説明することができる。したがって、以降では、クラスタのかわりに層(layer)という言葉を用いることにする。図1の左下・中・右上の層を、図2の結果では、それぞれ、(I, II, III), (IV, V, VI), (VII, VIII, IX) に買い分類されていることが分かる。以下からは、図2にしたがって見ていくことにす

るが、図2におけるそれぞれの層は、一体どのような集団であるかについて考察していく。各科目と層の関係を、図3に示す。図中の色は、下のカラースケールに対応した参照ベクトルの値である。たとえば、国語Iの図における右上の青色あたりに位置づけられる受験者は、国語Iの得点が30点ほどだということを示している。ただし、32科目分の出力は冗長であるために、表1を参考に、受験者数が多かった主要な16科目に限定した。図3から読み取れる事実を簡単に要約すると、次のようになる。

- (1) 左下から右上にかけて、「教科学力の総合力」が高いものから低いものに並んでいる。
- (2) 「 $y=x$ 」の対角線より左上には、相対的に文系の学生が集中している。
- (3) 「 $y=x$ 」の対角線より右下には、相対的に理系の学生が集中している。
- (4) 層IIは、全科目万能型であるので、上位一般型と呼ぶことにする。
- (5) 層IIIは、ある程度全科目でき、かつ、『地理B』『数学I・A』『化学IB』などができる集団である。したがって、上中位一般・上位理数型と呼ぶことができる。
- (6) 層IIIは、ある程度、全科目できるが、特に、『地歴』の3科目『世界史B』『日本史B』『地理B』および『現代社会』に強い集団である。いわゆる上位文社型である。ただし、ここでも、「文社」とは、いわゆる人文社会科学を指し示すものとする。
- (7) 以上のような解釈を同様に行うと、層IVは、中位理数型である。
- (8) 同様に、層VIは、中位一般型である。
- (9) 同様に、層VIIは、中位文社型である。
- (10) 同様に、層VIIIは、下位理数型である。
- (11) 同様に、層VIIIは、下位文社型である。
- (12) 同様に、層IXは、下位一般型である。
- (13) 『倫理』は、現役生を (I, III, VI), (II, IV, V), (VII, VIII, IX) と、大きく3つの層に分けるのに有効である。

(14) 『日本史B』『政経』『地学IB』『生物IB』は、現役生を (I, II, III), (IV, V, VI), (VII, VIII, IX) と、大きく3つの層に分けるのに有効である。

(15) 『国語I・II』『数学I・A』『英語』は、できる受験生とそうでない受験生がばらけている。

特に (15) より、一般に教科学力が低い集団の中にも『国語I・II』『数学I・A』『英語』の科目ができる受験生がいることは、これらの科目が主要科目中の主要科目であり、優先的に教科学力を上げようとする受験生たちの意図があると思われる。逆に言えば、余力のある受験生が、『地歴』『公民』『理科』などの教科を勉強し、総合的な教科学力を身につけていくと考えることができる。また、層IIについては、理数型でありながら『地歴』もできるゆえに上中位一般・上位理数型と命名したが、この層では、一方では、『国語』『公民』の教科ができないと考えることもできる。したがって、『地歴』などの比較的短答式の色彩が強い教科については点数がよいが、『国語』や『公民』など、ある程度の長文を読み、解釈を伴う解答を必要とする科目は苦手する受験者たちで層が形成されている可能性もある。これは、昨今言われているような読解力の低下問題と関連があるかもしれない、今後の検討が必要であろう。

さて、5万件のデータを用いて、図2のように同定されたSOMマップをもとに、一体、385,892人の現役生たちの各人が、どの層に属するかについて興味があるだろう。そこで、図2で得られたマップから、各現役生が、どの層に属する可能性が高いかについて、参照ベクトルとの距離から予測した。また、層ごとに、選択科目パターンも集計したが、これらを合わせて表3に示す。

一般的に、上位であるほど、選択科目が多い。また、どの層にも『国語』から1科目(多くは『国語I・II』)、『地歴』から1科目(『世界史B』『日本史B』『地理B』のいずれか)、『外国語』から1科目(ほとんどが『英語』)の3科目受験を行っているものが多い。理数型と命名した層にも、上記の3科目受験者はいる。SOMによるメンバーシップは、科目選択パターンと同時に、受験者同士の教科得点の距離(本分析ではユークリッド距離)の近さによって、受験者を布置しているためにこのようなことが起こっている。また、理数型・文社型の別を、名義的なものではなく、連続的な変化の強弱から分類しているためである。



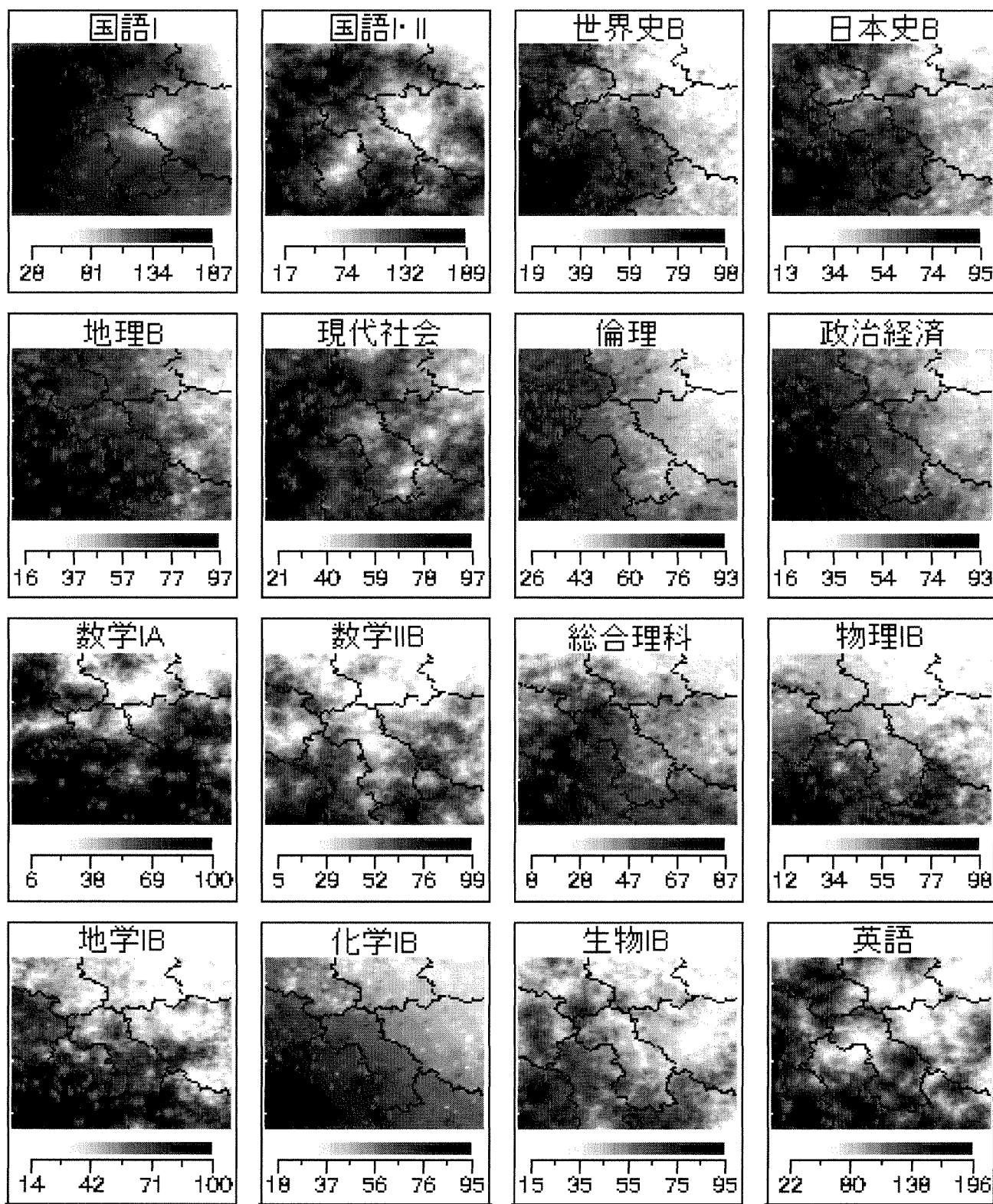


図3：主要16科目と層との関係



表3：層ごとの度数と層ごとの上位10通りの科目選択パターン

I：上位一般型	39,211	II：上中位一般・上位理数型	41,924	III：上位文社型	34,936
P2 Q6 S2 T2 U3 V2 X1	3,686	P2 Q4 X1	4,021	P2 Q4 R1 S2 T2 W2 X1	2325
P2 Q6 R1 S2 T2 U3 V2 X1	2,608	P2 Q6 S2 T2 U3 V2 X1	3,651	P2 Q4 X1	2247
P2 Q4 X1	2,372	P2 Q6 R1 S2 T2 U3 V2 X1	3,118	P2 Q4 R3 S2 T2 W2 X1	1576
P2 Q2 X1	1,643	P2 Q2 X1	2,418	P2 Q2 R1 S2 T2 W2 X1	1485
P2 R1 S2 T2 U3 V2 X1	1,623	P2 Q3 X1	1,250	P2 Q2 X1	1168
P2 R2 S2 T2 U3 V2 X1	1,513	P2 Q3 R2 S2 T3 U2 X1	1,233	P2 Q4 R2 S2 T2 W2 X1	1038
P2 R3 S2 T2 U3 V2 X1	1,343	P2 R1 S2 T2 U3 V2 X1	1,204	P2 Q6 S2 T2 U3 V2 X1	959
P2 Q6 S2 T2 V2 W2 X1	1,110	P2 Q6 S2 T2 V2 W2 X1	1,018	P2 Q2 R2 S2 T2 W2 X1	944
P2 Q4 R1 S2 T2 W2 X1	1,058	P2 R2 S2 T2 U3 V2 X1	973	P2 Q2 R3 S2 T2 W2 X1	878
P2 Q4 S2 T2 U3 V2 X1	929	P2 X1	911	P2 Q6 R1 S2 T2 U3 V2 X1	868
IV：中位理数型	42,895	V：中位一般型	49,356	VI：中位文社型	37,876
P2 Q4 X1	3,336	P2 Q4 X1	6,218	P2 Q4 X1	3,277
P2 Q6 R1 S2 T2 U3 V2 X1	3,228	P2 Q2 X1	2,918	P2 Q4 R1 S2 T2 W2 X1	2,224
P2 Q6 S2 T2 U3 V2 X1	3,194	P2 Q6 S2 T2 U3 V2 X1	2,614	P2 Q2 X1	1,871
P2 Q2 X1	2,036	P2 Q6 R1 S2 T2 U3 V2 X1	2,107	P2 Q4 R3 S2 T2 W2 X1	1,585
P2 R1 S2 T2 U3 V2 X1	1,684	P2 Q4 R1 S2 T2 W2 X1	1,212	P2 Q2 R1 S2 T2 W2 X1	1,542
P2 R3 S2 T2 U3 V2 X1	1,497	P2 Q4 R3 S2 T2 W2 X1	1,179	P2 Q6 S2 T2 U3 V2 X1	1,054
P2 Q4 S2 T2 U3 V2 X1	1,150	P2 X1	1,116	P2 Q6 R1 S2 T2 U3 V2 X1	888
P2 Q6 S2 T2 V2 W2 X1	1,100	P2 Q6 S2 T2 V2 W2 X1	1,082	P2 Q2 R3 S2 T2 W2 X1	848
P2 R2 S2 T2 U3 V2 X1	1,055	S2 T2 V2 X1	1,044	P2 Q4 R1 S2 T2 U1 W2 X1	844
P2 Q4 R1 S2 T2 U3 V2 X1	1,039	P2 R3 X1	910	P2 Q4 R2 S2 T2 W2 X1	837
VII：下位理数型	71,524	VIII：下位文社型	39,631	IX：下位一般型	28,538
P2 Q4 X1	9,431	P2 Q4 X1	5,245	P2 Q4 X1	4,071
P2 Q2 X1	4,306	P2 Q2 X1	2,568	P2 Q2 X1	1,962
P2 Q6 S2 T2 U3 V2 X1	3,079	P2 Q4 R1 S2 T2 W2 X1	1,342	P2 X1	1,837
P2 X1	3,034	P2 X1	1,158	P2 S2 X1	1,263
P2 Q6 R1 S2 T2 U3 V2 X1	2,369	P2 S2 X1	928	S2 T2 U3 X1	727
S2 T2 U3 X1	1,865	P2 Q4 R3 S2 T2 W2 X1	885	P2 R3 X1	688
P2 S2 X1	1,749	S2 T2 U3 X1	809	P1 X1	606
P2 R3 X1	1,656	P2 Q2 R1 S2 T2 W2 X1	770	X1	600
P2 R1 S2 T2 U3 V2 X1	1,478	P2 R3 X1	722	P2 S2 T2 U3 X1	426
P2 S2 T2 U3 X1	1,394	P2 Q6 S2 T2 U3 V2 X1	635	S2 T2 V2 X1	398

## 6. FIMLによる教科学力構造

本節では、FIMLを用いて得られた相関構造を報告する。また、合わせて平均構造・共分散構造も示す。前述したように、本節で報告する平均・共分散・相関構造は、「欠測構造がMARであると仮定したもとの、仮に、全現役生が、各科目に解答したならば、そこで得られるであろう平均・共分散・相関構造」である。したがって、データがMCARの仮定を満たさないのであれば、平均やSDなどが、表1で得られたものと異なってくるはずである。

なお、相関については報告していないが、表1で得られた平均およびSDは、「各科目に解答した2005年の現役生の平均とSDが単にどれくらいであったかを示すだけならば、絶対的に正しい」ことに注意されたい。同様に、大学入試センターが毎年ウェブサイトで報告している各科目の平均点およびSDも、「解答した受験生の平均とSDを報告しているものとしては絶対的に正しい」ものである。つまり、分析するにあたり、抛って立つ前提が異なっている。我々は、欠測構造がMARであると仮定したもとの平均・共分散・相関構造に興味があるために、そのもとのバイアスのないFIMLを用いている。

結果を表4に示す。大きな行列であるため、主要16科目に限って報告している。ソフトウェアは、AMOS 5.0 (Smallwaters, 2003) を用いた。相関係数の強弱によって網掛けを施してある。また、比較のために表1における平均とSDも付した。

まず、平均とSDの推定値が表1と大きく変わることから、やはり本データの欠測構造がMCARであると仮定することは難しいことが分かった。たとえば、表1の『国語I』の平均点は115.0点であり、表4におけるFIMLにもとづく『国語I』の平均点は126.9点であった。表1の115.0点の解釈は「実際に『国語I』を受験した33,834人の現役生の平均点」であるのに対し、表4の126.9点は「欠測構造がMARと仮定したもとの、仮に385,892人の全現役生が『国語I』を受験したら得られるであろう平均点」である。このような前提の違いから由来する結果の相違である。したがって、どちらが優劣しているというものではない。しかし、両者の『国語I』の平均点を比較することにより、教科学力の低い集団が『国語I』を選択している可能性が示唆される。

また、分散構造から、表1のSDと比べて、表4のSDの方がわずかに大きくなっている。これも、表1のSD

が、「各科目を受験した現役生の得点の散らばり」であるのに対し、表4のSDが「欠測構造がMARと仮定した下で、仮に385,892人の全現役生が受験したときの得点の散らばり」であるがために、後者の方が散らばりが大きいのである。

さらに、相関構造より、大学入試センター試験のアラカルト方式により、同一教科内の科目は、同一時間に実施されるために、同一教科内科目間の相関を計算することができない。したがって、相関を計算できない箇所が空白となっている。相関行列に構造的な欠測があることが、多変量正規分布に基づく因子分析などの諸モデルを行うことを難しくしている。しかしながら、それでも大体において以下の事柄が示唆される。

- (1) 『国語』『地歴』および『数学』『理科』という相関が高い2つの変数群が存在することから、大きく分けて、理数系の因子と文社系の2因子になりそうである。
- (2) 『英語』はどの科目とも相関が高いことから、一般因子、あるいは上記の2因子の双方から影響を受けていることが示唆される。いわゆる、文系学生・理系学生もまず勉学に励む主要科目の中でも随一の主要科目だからである。
- (3) 『数学』と『理科』の教科内相関の方が、『国語』『地理歴史』『公民』の教科内相関よりも高い傾向がある。いわゆる、文社系の教科・科目の方が、学力達成度を測定する困難さが改めて浮き彫りになっている。
- (4) 『地学IB』が科目横断的に相関が高いことから、理系学部志望の現役生だけでなく、文系学部志望の現役生にも科目選択の受け皿となっており、一般因子、あるいは上記2因子から影響を受けているといえる。
- (5) 『数学I・数学A』『数学II・数学B』『総合理科』が人文社会系科目と相関が低く、理数系科目と相関が高いことから、理数系因子の弁別的妥当性に貢献しそうである。一方で、『物理IB』『化学IB』が『公民』の3科目と意外と相関が高い。逆に、『国語I』『国語I・国語II』は、文社系因子の弁別的妥当性に貢献しそうである。

ただし、表4の結果は、主要16科目で計算したものであるため、全32科目で計算すれば結果がわずかに異なるであろうと考えられる。このように、モデルが前提とする諸条件—ここでは、16変量正規分布か

32変量正規分布カーがかわれば、それに伴って結果も変わるのがモデルに基づく分析の特徴でもあり、時として短所・長所が分かれるところである。その点、表1の結果は不変である。

また、前節で層別化した9層ごとに、16の主要科目についてFIMLによって推定された平均値をプロットしたものを図4に示す。『国語I』『国語I・国語II』および『英語』は、100点満点に換算してある。また、図5は、図4にもとづいて順位によって並べ替えたものである。図4、5より、一般型の層I（上位一般型）、層II（上位理数型）、層V（中位一般型）、層IX（下位一般型）は、科目によってそれほど順位が変更しない。一方、層II（上中位一般・上位理数型）、層IV（中位理数型）、層VII（下位理数型）は、『数学』の2科目、および『理科』

の5科目が、他の科目と比べて相対的に得点・順位が高い。なお、層IIが上中位一般型としたわけは、層V（中位一般型）との順位相関が高いことにもよっている。また、層III（上位文社型）、層VI（中位文社型）、層VIII（下位文社型）の3層は、『国語』『地理歴史』『公民』の得点・順位が高い。なお、明示していないが、各層内の分散および共分散は小さく、層間分散が大きかった。これは、得点の散らばりが、層内の散らばりよりも層間の散らばりの方が大きいことを示すものであり、前述したが、クラスタではなく、層と呼ぶ方が好ましい根拠である。

表4：欠測構造がMARと仮定したもとのFIMLによる主要16科目の平均・共分散・相関構造  
(対角要素および下三角行列部分は共分散構造、上三角行列部分は相関構造)

	P1	P2	Q2	Q4	Q6	R1	R2	R3	S2	T2	U1	U3	V2	V4	W2	X1
国 I	1406.3		0.449	0.428	0.495	0.537	0.548	0.446	0.387	0.383	0.393	0.451	0.431	0.520	0.446	0.541
国 I・II		1074.2	0.509	0.486	0.520	0.593	0.611	0.514	0.394	0.394	0.423	0.471	0.471	0.505	0.489	0.588
世史 B	340.8	337.4	409.6			0.538	0.595	0.452	0.446	0.442	0.389	0.542	0.574	0.582	0.490	0.627
日史 B	263.8	261.7		269.8		0.515	0.564	0.513	0.409	0.410	0.388	0.517	0.559	0.537	0.472	0.578
地理 B	284.8	261.5			235.0	0.557	0.553	0.599	0.450	0.449	0.464	0.555	0.589	0.545	0.520	0.573
現社	259.4	250.3	140.3	109.1	109.9	165.9			0.385	0.376	0.454	0.472	0.471	0.527	0.492	0.551
倫理	308.5	300.6	180.9	139.1	127.4		225.7		0.409	0.404	0.431	0.519	0.551	0.538	0.525	0.600
政経	281.5	283.8	222.1	169.7	154.6			283.5	0.456	0.452	0.493	0.561	0.604	0.418	0.556	0.594
数 I・A	335.4	298.5	208.3	155.0	159.3	114.5	141.9	177.4	533.7	0.756	0.468	0.576	0.661	0.538	0.541	0.572
数 II・B	272.9	245.4	170.0	128.1	130.9	92.1	115.5	144.8	332.0	361.8	0.452	0.638	0.661	0.535	0.546	0.566
総合理科	225.1	211.7	120.4	97.4	108.8	89.3	98.9	126.9	165.3	131.3	233.6		0.503	0.536	0.557	0.434
物理 IB	343.0	312.8	222.4	172.0	172.5	123.2	158.0	191.6	316.6	265.3		410.8	0.703	0.625	0.585	0.630
化学 IB	363.1	346.8	279.6	206.6	203.2	136.5	186.2	228.7	343.2	282.9	173.0	320.5	505.6		0.587	0.658
地学 IB	414.1	351.0	250.1	187.0	209.8	144.1	193.8	219.6	263.6	215.9	173.8	269.4		450.2	0.600	0.566
生物 IB	255.6	244.5	151.5	118.5	121.6	96.7	120.4	143.0	190.9	158.5	129.9	181.1	201.4	195.2	233.0	0.550
英語	765.0	726.3	478.2	358.0	331.0	267.7	339.9	376.7	498.0	405.6	250.2	488.3	556.2	452.6	316.7	1420.8
Mean	126.9	115.9	58.4	58.0	66.7	67.4	60.4	61.4	65.3	47.1	47.7	51.8	57.6	57.6	49.3	113.1
SD	37.5	32.8	20.2	16.4	15.3	12.9	15.0	16.8	23.1	19.0	15.3	20.3	22.5	21.2	15.3	37.7
Mean(表1)	115.0	117.8	61.5	58.0	69.0	69.2	65.9	62.9	67.5	50.8	47.6	57.9	63.7	64.0	50.2	113.3
SD(表1)	36.4	32.5	20.1	16.2	15.0	12.4	14.3	16.6	22.9	18.3	14.5	20.2	22.1	20.5	14.8	37.7

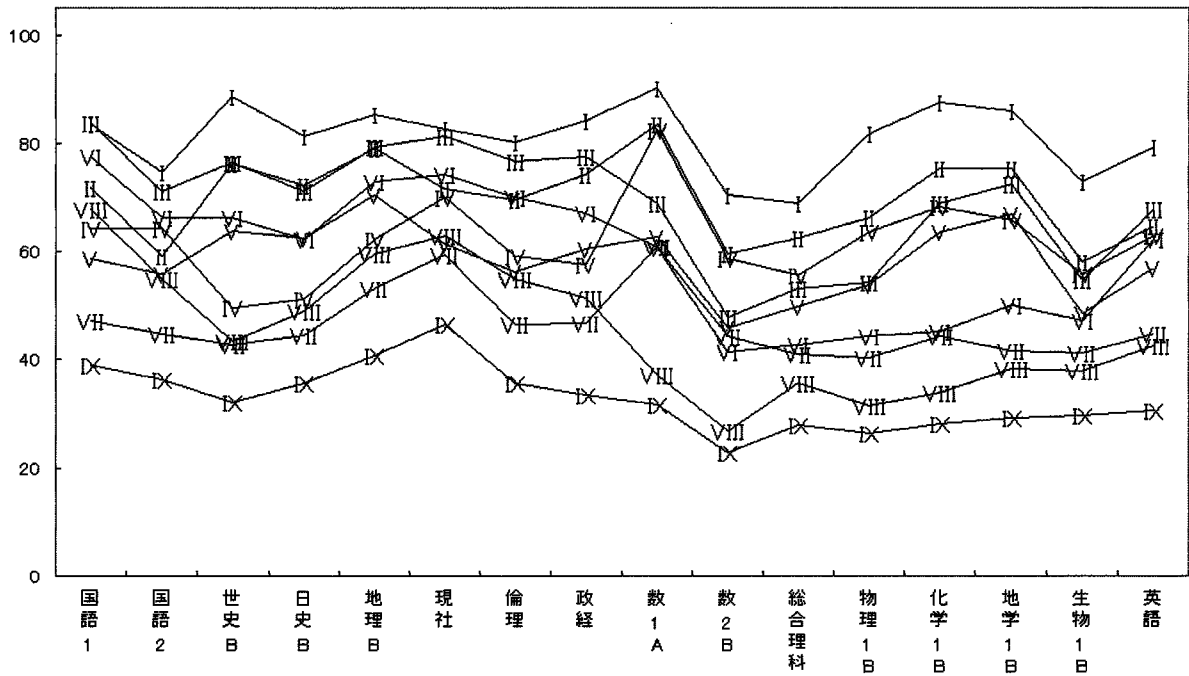


図4：FIMLによる各9層の教科別平均点（『国語』『外国語』は100点満点換算）

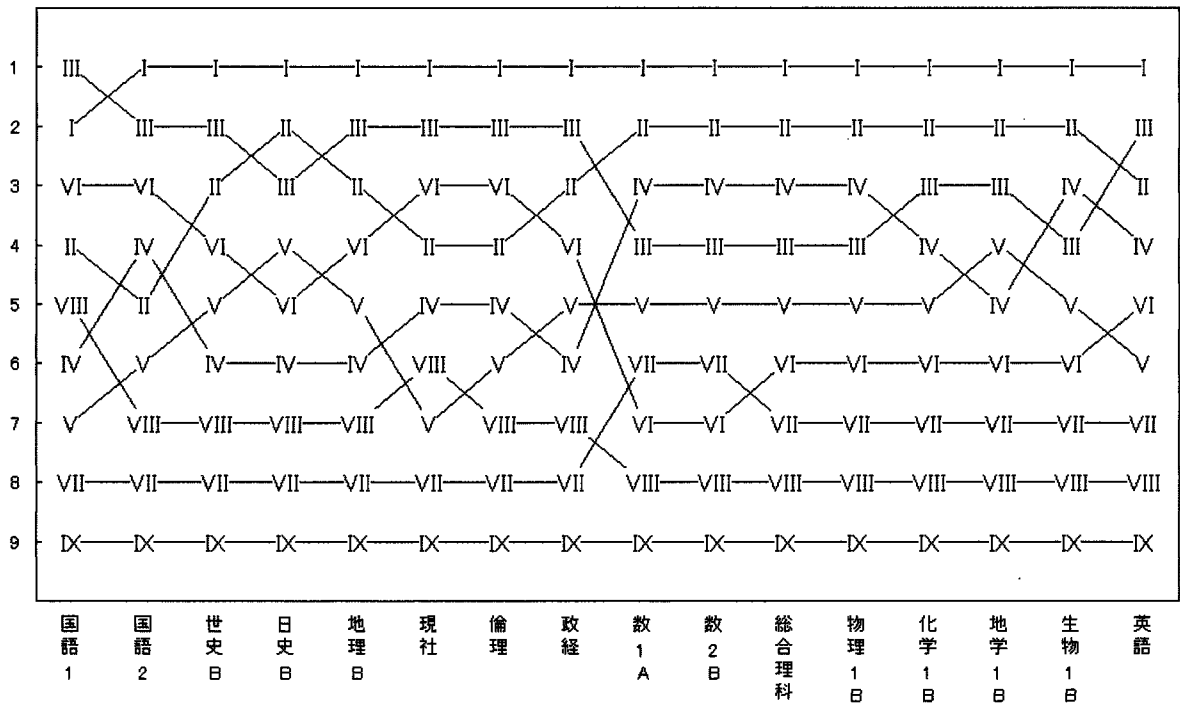


図5：各9層の主要教科別平均点の順位

## 7. まとめ

本研究では、2005年の大学入試センター試験における現役生・健常者・本試験受験生の教科学力構造を自己組織化マップ(SOM)およびFIMLにもとづく平均・共分散・相関構造の両面からアプローチした。

SOMによって、明らかになったことは、学力をクラスタリングすると、ほぼ、総点の高さによって層別化されることが分かったことである。また、大まかには、一般型・理数型・文社型に3分され、さらに上中下位に3分類されることが分かった。それら、9(=3×3)層のグループは、それぞれ、科目選択パターンが異なり、現役生たちは、多様な入試戦略を採用していることが明らかになった。これらについては、入試に携わる者、受験産業の関係者にとっては、経験的に知られていたことであるが、今回、SOMを用いて改めて実証されたといえよう。

続いて、FIMLを用いて、多変量正規分布にもとづいて、また、欠測構造がMARであると仮定したもとの平均・共分散・相関構造が示された。そこでは、MCARが仮定されている表1の結果と、MARが仮定されている表4の結果が比較された。また、相関構造および図4、5より、現役生が多様な入試方略を採用していることが示唆された。

今後の分析の方向性としては、モデルに基づく統計分析をより詳細に検討することで、データからより抽象化された知見を抽出することである。たとえば、探索的因子分析や確認的因子分析(Jöreskog, 1969)を用いるなどである。SOMの結果から、あるいは、表4の相関行列をざっと見る限り、2因子のCFAなどは有望なモデルである。ただし、表4の結果を見て分かるとおり、大学入試センター試験のアラカルト方式により、同一教科内の科目は、同一時間に実施されるために、同一教科内科目間の相関を計算することができない。仮に計算できたとしても、相当強い欠測構造を仮定する必要がある。あるいは、欠測パターンごとに母集団を便宜的に設定して、多母集団同時分析によるCFAを行うことは理論的に可能である。ただし、上述したとおり、欠測パターン(科目選択パターン)は4,907通りあるので、現実的には非常に困難である。もう1つの有効な手段は、擬最尤法(Gong & Samaniego, 1981; 狩野, 2003)に基づいて識別性が確保される変数群から分析し、得られた推定値を利用して全体のモデルを識別させることであろう。

このことが、冒頭で述べたように、センター試験データをモデルに基づいて統計分析することの本質的な難点である。ほとんどの多変量解析は、多変量正規分

布にもとづいている。特に、テストデータに対しては、一般に、多変量正規分布を仮定してもあまり無理はないと考えられる。それにもかかわらず、センター試験のデータにおける欠測構造がモデルに基づく分析を困難にしており、この点については、今後の課題となろう。

本研究では、2005年のセンター試験を受験した385,892人の現役生・健常者・本試験受験者の教科学力構造に着目したが、一口に38万人といっても、そこには、多種多様な学習履歴、あるいは国公立志願の別、男女の別など、さまざまな集団が混在しているであろう。SOMとFIMLによる結果からは、38万人の受験者全体では、おおむね2因子構造が指示されるかもしれないが、ある特定の属性変数に着目して、より分析的な検討を行うと、結果は異なってくるかもしれない。また、SOMは平均にもとづくクラスタ化の傾向があるので、分析手法を越えて得られる知見であるかの確認も必要である。今後の課題として、具体的な、科目選択行動や出願行動などによって受験者を層別化した上で、グループごとの教科学力構造を明らかにすることであると思われる。

### 注釈

- 1) なお、平成19年度より、6教科9グループ28科目となる。詳細は、[http://www.dnc.ac.jp/center\\_exam/19exam/syutudai.html](http://www.dnc.ac.jp/center_exam/19exam/syutudai.html)で確認されたい。
- 2) 等化、あるいは得点換算表を作成して、比較を行う手段もある。

### 参考文献

- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced Structural Equation Modeling: Issues and Techniques*. Lawrence Erlbaum Associates, Inc. (pp. 243-277).
- Bock, R. D., Gibbons, R., & Muraki, E. (1994). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- 大学入試センター (2005a). 平成 17 年度センター試験利用大学。  
([http://www.dnc.ac.jp/center\\_exam/7exam/riyou](http://www.dnc.ac.jp/center_exam/7exam/riyou).)

- html).
- 大学入試センター (2005b). 平成 17 年度センター試験  
利用短期大学.  
([http://www.dnc.ac.jp/center\\_exam/17exam/riyou\\_tanki.html](http://www.dnc.ac.jp/center_exam/17exam/riyou_tanki.html)).
- 大学入試センター (2005c). 平成 17 年度センター試験  
志願者数及び受験者数等.  
([http://www.dnc.ac.jp/center\\_exam/17exam/17shigan.html](http://www.dnc.ac.jp/center_exam/17exam/17shigan.html)).
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977).  
Maximum likelihood from incomplete data via  
the EM algorithm (with discussion). *Journal of  
the Royal statistical Society, Series B*, 39, 1-38.
- Gong, G. & Samaniego, F. J. (1981).  
Pseudo-maximum likelihood estimation: Theory  
and applications. *The Annals of Statistics*, 9,  
861-869.
- 岩崎 学 (2002). 不完全データの統計解析 エコノミスト  
社.
- Jöreskog, K. G. (1969). A general approach to  
confirmatory maximum likelihood factor analysis.  
*Psychometrika*, 34, 183-202.
- Jöreskog, K. G. & Moustaki, I. (2001). Factor  
analysis of ordinal variables: A comparison of  
three approaches. *Multivariate Behavioral  
Research*, 36, 347-387.
- 狩野 裕 (2003). 擬最尤法と識別性 統計関連学会連合  
大会講演報告集, 499-500.
- 株式会社数理システム (2005) *Visual Mining Studio  
5.0*.  
(<http://www.msi.co.jp/>).
- 勝田守一 (1962). 学力とは何か—意見をよんで 国土社  
『教育』, p. 24.
- Kohonen, T. (1995). *Self-organizing maps*. Springer.
- Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1990). Full  
maximum likelihood analysis of structural  
equation models with polytomous variables.  
*Statistics & Probability Letters*, 9, 91-97.
- Little, R. J. A. & Rubin, D. B. (1987). *Statistical  
analysis with missing data*. John Wiley & Sons.
- 松原達哉 (2004). 知能の考え方の多様性と見方 イア  
ン・ディアリ (著) 繁榊算男 (訳) 知能 岩波書店,  
pp.161-174. (Deary, I. J. (2001) *Intelligence: A  
very short introduction*, Oxford University  
Press.).
- Muraki, E. & Carlson, J. E. (1995). Full-information  
factor analysis for polytomous item responses.  
*Applied Psychological Measurement*, 19, 73-90.
- 中畝菜穂子・吉村 幸・荘島宏二郎 (2005). 大学入試セ  
ンター試験『国語』『数学』『英語』の正答率・無答率か  
らみた特徴と作問への示唆 大学入試研究ジャーナ  
ル, 15, 55-62.
- Rubin, D. B. (1987). *Multiple imputation for  
nonresponse in surveys*. John Wiley & Sons.
- 高浦勝義 (2005). 学力の測定から学力を育てる評価  
へ国立教育政策研究所 (編) 学力の総合的研究, pp.  
8-20.
- 徳高平蔵・岸田悟・藤村喜久郎 (1999). 自己組織化マッ  
プの応用—多次元情報の 2 次元可視化— 海文堂.
- 徳高平蔵・藤村喜久郎・山川烈 (監修) (2002). 自己組織  
化マップ応用事例集—SOM による可視化情報処理  
— 海文堂.
- SmallWaters Corporation (2003). *AMOS 5.0*.  
(<http://www.smallwaters.com/>).
- 荘島宏二郎・吉村 幸・中畝菜穂子 (2005). 項目反応理  
論から見た大学入試センター試験『国語』『数学』『英  
語』の特徴 大学入試研究ジャーナル, 15, 63-66.
- SPSS Inc. (200?) *Clementine 7.0*.  
(<http://www.spss.co.jp/>).
- Eudaptics Software GmbH (2001) *Discovery SOMine  
Plus 4.0*. (国内連絡先: 高松市, マインドウェア).
- 渡辺美智子・山口和範 (2000). EM アルゴリズムと不完全  
データの諸問題 多賀出版.
- 吉村 幸・荘島宏二郎・杉野直樹・野澤健・清水裕子・齋藤  
栄二・根岸雅史・岡部純子・フレイザー, サイモン  
(2005). 大学入試センター試験既出問題を利用した  
共通受験者計画による英語学力の経年変化の調査  
日本テスト学会誌, 1, 51-58.

# 自己組織化マップによる平成18年度大学入試センター試験の教科学力構造

荘島 宏二郎

独立行政法人 大学入試センター

本研究では、平成18(2006)年度の大学入試センター試験データを分析した。分析対象は、389,235人の健常な現役生であった。まず、34科目の基礎統計量と科目選択パターンを示した。続いて、完全情報最尤法を用いて主要19科目の平均・共分散・相関構造を同定した。また、教科学力構造を自己組織化マップ(SOM)を用いて探索的に分析した。SOMによって8つのクラスター(層)が抽出された。また、性差について検討した。

**キーワード:** 教科学力, 大学入試センター試験, 自己組織化マップ, 完全情報最尤法

Scholastic Achievement Structure of National Center Test 2006

by Self-Organizing Map

Kojiro Shojima

The National Center for University Entrance Examinations

The scholastic achievement structure in the data of the National Center Test 2006 was exploratorily analyzed using a self-organizing map (SOM). Our target population was a group of 389,235 able-bodied high school graduates (12th graders). First, the simple statistics of all 34 subjects. We then used the full-information maximum likelihood method to identify the mean, covariance, and correlation structure of the main 19 subjects. Second, we did an exploratory analysis of the scholastic achievement structure of NCT2006 by SOM, and extracted eight clusters (layers). We also examined the distinction of genders in scholastic achievement.

**Keywords :** scholastic achievement, The National Center Test, self-organizing map, full-information ML

## 1. はじめに

本研究では、平成18年度(2006年1月実施)の大学入試センター試験(The National Center Test for University Admissions, NCT)について、自己組織化マップ(self-organizing maps, SOM; e.g., Kohonen, 1995; 徳高・岸田・藤村, 1999)を用いた教科学力(荘島・大津・田栗, 2006; 荘島, 2006)の構造を明らかにする。なお、これより、2006年度のNCTデータをNCT2006などと表現する。また、教科学力とは、簡単に言うと、客観式テストで測定しやすい旧来型学力の

一群である。

SOMによるNCTの分析は、荘島ら(2006)や荘島(2006)に見られる。ここでは、SOMによって9つの学力層を取り出し、各層ごとの科目選択行動や得意科目を明らかにした。

しかし、平成18(2006)年度から、NCTの科目編成が変わり、中でも「理科」の科目編成が大きく変更した。特筆すべきは「英語リスニング」が加わったことである。つまり、荘島ら(2006)によるNCT2005の結果との比較の中で、本研究においてNCT2006における教科学力構造を明らかにすることは有意義である。

---

Correspondence concerning this article should be sent to: Kojiro Shojima, The National Center for University Entrance Examinations, Komaba 2-19-23, Meguro-ku, Tokyo, Japan 153-8501. E-mail: shojima@rd.dnc.ac.jp

This research was supported by a Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science

(C) No. 16650220.

また、そもそもNCTに関する科目横断的な研究は多くない。そういった意味で、本研究はNCT2006に関する資料的な価値が高い。

## 2. 平成18(2006)年度大学入試センター試験

大学入試センター(2006)によると、NCT2006の利用大学・短期大学は、表1のとおりである。また、本試験受験者数は、506,241人であった。

表1：センター試験利用大学・短期大学(H18)

国立大学	公立大学	私立大学	
82大学	72大学	440大学	1,179学部
公立短期大学		私立短期大学	
13短期大学	32学科	120短期大学	288学科

NCT2006には、実に様々な受験者が混在し、実際には昭和20~30年代卒業の受験者も存在している。しかし、本研究では、平成18年度の17-18歳人口(12年生)がどのような教科学力構造および科目選択行動を行っているかに焦点化する。また、先述した科目編成の変更により、浪人生は旧課程の科目を、現役生は新過程の科目を選択する傾向が高く、現浪比較が行いにくいという事情もある。ただし、NCT2005における現浪比較は、荘島(2006)が報告している。要約すると、一般的に浪人生のほうが現役生より教科学力が高いが、それは1~3浪までであり、必ずしも長く浪人を経験した者が高いテスト得点をとるわけではなかった。この事実は、我々の経験を裏付けるような結果である。

本研究でもNCT2006データのうち、本試験を受験した健常な現役生389,235人を分析対象とする。以降、彼らを単に現役生と表現することがある。また、特別な断りがなくNCT2006と言う場合は、当該の現役生に関するデータをさすものとする。

表2に基本統計量を示す。受験者が最も多い科目は、「英語(筆記およびリスニング)」次に、「国語」であった。国語は、近代以降の文章、古文、漢文の問題から構成されているので、その内訳も示した。ただし、「古典」「漢文」は、そもそも選択していない者も含めて算出しているが、その選択/非選択のフラグがデータからはわからないので、それらを選択していない現役生の得点も0点として扱っている。注意されたい。

「国語」と「外国語」が200点満点、「英語リスニン

グ」は50点満点、その他の科目はすべて100点満点であった。また、Codeは、科目をコンパクトに示すために用意した記号である。荘島・大津・田栗(2006)とおおむね一致しているが、科目編成が変更したため、一部の科目で一致していない。

Shojima, Otsu, Taguri, Mayekawa, & Yanai(2006)が指摘するように、NCTデータを分析するためには2つの困難がある。1つは、構造的な欠測パターン、もう1つは非構造的な欠測パターンである。

構造的な欠測パターンは、NCTの実施プログラムが原因で起こる。NCT2006は、英語リスニングも1つと数えると、全34科目が実施された。それらは、大きく6つの教科「国語」「地理歴史」「公民」「数学」「理科」「外国語」に分類されている。ただし、「数学」は「数学①」「数学②」、「理科」は「理科①」「理科②」「理科③」に分かれており、全部で9つの教科単位に下位分類されている。その9つの教科単位が、NCTの2日間の日程の中で9コマに配置され、実施されている。したがって、同一教科単位内の科目は、同じコマで実施されているため、複数受験することができない。たとえば、「公民」には「現代社会」「倫理」「政治・経済」の3科目があるが、2つ以上同時に受験することができない。このような、NCTの科目編成が構造的な欠測を生じさせている。

また、非構造的な欠測パターンとは、NCTのアラカルト方式から由来している。というのは、NCTは、プレースメントテストではなく、選抜テストであるため、ある科目を受けるも受けないも受験者に委ねられている。したがって、1科目しか受験しない者もいれば、最大で9科目を受験する者まで実にさまざまである。

表3には、現役生約39万人の科目選択パターン(欠測パターン)を示した。表中の記号は表2のCodeに準じている。たとえば、「国語」「日史B」「英語(筆記とリスニング)」を選択した現役生は、35,012人いたことになる。また、すべての科目が100点満点だとして、1科目あたりの平均点を受験者ごとに算出し、科目選択パターンごとに、それらの最小値、最大値、平均、SDを計算したものを表3に付した。科目ごとに難易度が異なるものを一緒にくたにしている。厳密には比較できないが、一見して、科目選択数が多いグループほど、平均点が高いことがわかる。実際、約39万人の受験者の科目数と1科目あたりの平均点との相関は0.493であった。



表2：NCT2006の基本統計量（現役生389,235人）

		Code	N	Min	Max	Mean	SD	Skew	Kurt
国語	国語	P1	371,216	0	200	123.55	36.18	-0.460	-0.572
	近代以降	PJ	371,216	0	100	66.44	17.27	-0.461	-0.189
	古文	Pcl	371,216	0	50	22.79	12.43	-0.162	-0.733
	漢文	PCh	371,216	0	50	34.32	16.69	-1.015	-0.333
地理歴史	世界史A	Q1	1,101	11	100	44.05	16.77	0.871	0.296
	世界史B	Q2	72,421	0	100	65.04	21.30	-0.338	-0.901
	日本史A	Q3	3,501	7	100	56.84	15.39	0.199	-0.376
	日本史B	Q4	116,214	0	100	53.38	16.95	0.099	-0.664
	地理A	Q5	4,340	10	100	61.83	15.55	-0.140	-0.512
	地理B	Q6	83,956	6	100	64.05	16.07	-0.295	-0.516
公民	現代社会	R1	174,707	0	100	56.75	13.43	0.090	-0.299
	倫理	R2	28,826	10	100	67.86	15.80	-0.522	-0.334
	政治・経済	R3	48,190	5	100	59.52	15.57	-0.194	-0.570
数学	数学I	S1	10,578	0	100	54.31	27.00	-0.150	-1.005
	数学I・A	S2	272,153	0	100	61.07	21.42	-0.317	-0.625
	数学II	T1	10,153	0	100	34.47	18.92	0.548	-0.180
	数学II・B	T2	238,354	0	100	55.56	22.87	-0.078	-0.775
	工業数理	T3	76	10	100	59.50	18.56	-0.593	0.213
	簿記	T4	984	8	100	57.18	18.38	0.108	-0.683
	情報関係基礎	T5	437	7	100	59.66	19.09	-0.195	-0.605
理科	理科総合B	U1	14,446	4	100	66.10	14.06	-0.477	0.070
	生物I	U2	139,413	0	100	68.62	17.14	-0.713	0.087
	理科総合A	V1	29,759	3	100	65.34	12.85	-0.064	-0.091
	化学I	V2	144,980	0	100	62.12	22.83	-0.212	-1.027
	物理I	W1	103,922	0	100	71.84	20.23	-0.624	-0.406
	地学I	W2	19,368	0	100	58.11	22.01	-0.156	-1.241
外国語	英語（筆記）	X1	384,036	0	200	125.10	39.20	-0.173	-0.854
	ドイツ語	X2	33	54	200	171.85	31.23	-2.036	5.463
	フランス語	X3	81	55	195	144.96	35.03	-0.573	-0.560
	中国語	X4	297	42	200	175.94	31.50	-2.296	5.122
	韓国語	X5	57	51	195	155.58	30.64	-1.230	1.644
	英語リスニング	Y1	379,491	0	50	36.09	8.24	-0.477	-0.187

荘島ら（2006）およびShojima et al.（2006）は、この事実を、NCTにおける欠測メカニズム（Rubin, 1976; Little & Rubin, 1987; 岩崎, 2002）がランダム欠測（missing at random, MAR）である1つの有力な傍証としている。

また、Schafer（1999）は、MARが成り立つための経験則をいくつか例示しているが、その1つに“double

sampling”という状況がNCT2006に当てはまる。Double samplingは、一部の変数は受験者の全てにデータがあればよく、NCT2006では、「英語」「国語」「数1A」の科目選択率が高いので、おおむねdouble samplingの状況が成立していると考えられる。

つづいて、表2における主要科目について、各教科

表3：科目選択パターン（現役生389,235人）

	Pattern	科目数	N	%	Min	Max	Mean	SD
1	P1 Q4 X1 Y1	4	35,012	9	5.75	76.2	44.9	11.16
2	P1 Q6 R1 S2 T2 V2 W1 X1 Y1	9	20,511	5.27	17.4	86.7	61	11.69
3	P1 Q4 R1 S2 T2 U2 X1 Y1	8	18,326	4.71	2	85.67	55.13	10.8
4	P1 Q2 X1 Y1	4	17,676	4.54	10	78	48.99	12.11
5	P1 Q6 S2 T2 V2 W1 X1 Y1	8	17,060	4.38	2	87.78	60.69	12.94
6	P1 Q2 R1 S2 T2 U2 X1 Y1	8	13,130	3.37	0	86.56	58.67	11.21
7	P1 R1 S2 T2 V2 W1 X1 Y1	8	10,798	2.77	12.5	88.22	57.55	12.9
8	P1 Q4 R3 S2 T2 U2 X1 Y1	8	9,103	2.34	8.5	84.22	56.23	10.98
9	P1 Q6 R1 S2 T2 U2 V2 X1 Y1	9	8,547	2.2	10.25	85.5	57.64	10.84
10	P1 R3 X1 Y1	4	7,099	1.82	12.5	74.1	43.65	10.7
11	P1 X1 Y1	3	7,023	1.8	2	73.38	36.31	11.14
12	P1 R3 S2 T2 V2 W1 X1 Y1	8	6,813	1.75	16.75	86.83	60.29	12.99
13	P1 Q6 S2 T2 U2 V2 X1 Y1	8	6,741	1.73	12.57	88	57.95	12.04
14	P1 R1 S2 T2 U2 V2 X1 Y1	8	6,024	1.55	10.67	85.39	55.8	12.34
15	P1 Q2 R3 S2 T2 U2 X1 Y1	8	5,200	1.34	12	86.78	59.55	11.43
16	P1 Q4 R1 S2 T2 V2 W1 X1 Y1	9	4,987	1.28	19.25	87.8	60.34	11.98
17	P1 Q4 R1 X1 Y1	5	4,602	1.18	0	79.17	45.78	10.66
18	P1 Q4 R2 S2 T2 U2 X1 Y1	8	4,494	1.15	14.33	85.83	59.64	10.52
19	P1 R2 S2 T2 V2 W1 X1 Y1	8	4,384	1.13	14	87.17	64.76	12.63
20	P1 Q2 R2 S2 T2 U2 X1 Y1	8	4,231	1.09	17.5	86.94	62.56	11.23
21	P1 Q4 S2 T2 V2 W1 X1 Y1	8	4,075	1.05	4.5	86.94	60.35	13.56
22	P1 Q6 R1 S2 T2 U2 X1 Y1	8	3,692	0.95	18.25	85.39	55.83	11.54
23	P1 Q4 R1 S2 T2 U2 V2 X1 Y1	8	3,640	0.94	16.77	85.5	55.56	11.06
24	P1 Q2 R1 S2 T2 V2 W1 X1 Y1	9	3,560	0.91	11.7	86.8	60.59	12.74
25	S2 T2 W1 X1 Y1	9	3,511	0.9	1.5	77.75	39.56	11.7
26	P1 S2 X1 Y1	5	3,371	0.87	8.33	76	37.59	10.44
27	P1 S2 T2 W1 X1 Y1	6	3,287	0.84	7.5	75.17	43.12	11.42
28	P1 R3 S2 T2 U2 V2 X1 Y1	8	3,227	0.83	15.5	85.78	57.48	12.47
29	S2 T2 V2 X1 Y1	5	3,225	0.83	5.75	81.5	42.21	12.56
30	P1 Q2 R1 X1 Y1	5	2,826	0.73	8	78.83	48.59	11.44
	周辺統計量	7.25	389,235	100	0	88.22	52.95	14.11

の同時選択率を表4、ペアワイズ削除にもとづく分散共分散行列と相関行列を表5に示す。いくつかの科目間で同時選択率が0であるが、これは先述した構造的欠測パターンによる。また、すべての科目間において、同時選択率がばらついていることは、アラカルト方式から由来する非構造的欠測パタンのためである。また、共分散行列や相関行列において、穴抜きが観測されるのは、構造的欠測パタンのためである。

表6には、完全情報最尤法（full-information

maximum likelihood, FIML; Arbuckle, 1996; Bock, Gibbons & Muraki, 1994; Muraki & Carlson, 1995; 荘島ら, 2006) による平均・共分散・相関構造を示した。「国語」の平均構造および、他科目との共分散・相関構造は、「国語現（近代以降）」「国語古」「国語漢」の平均・共分散・相関構造と、それらと他科目との共分散・相関構造から算出した。というのは、「国語」および「国語現」「国語古」「国語漢」は線形独立でないからである。ソフトウェアはAMOS 5.0 (Smallwaters,

表4：主要科目の同時選択率

	国語	世史B	日史B	地理B	現社	倫理	政経	数1	数1A	数2	数2B	物1	化1	生1	地学I	理総A	理総B	英語	リス
国語	371,216	0.186	0.298	0.215	0.447	0.074	0.123	0.025	0.662	0.025	0.581	0.252	0.354	0.348	0.050	0.076	0.037	0.945	0.936
世史B	72,237	72,421	0.000	0.000	0.087	0.021	0.022	0.005	0.117	0.004	0.104	0.021	0.034	0.080	0.014	0.020	0.007	0.185	0.184
日史B	115,837	0	116,214	0.000	0.135	0.020	0.037	0.007	0.170	0.009	0.148	0.030	0.063	0.121	0.019	0.030	0.010	0.296	0.294
地理B	83,647	0	83,956	0	0.143	0.007	0.015	0.003	0.200	0.004	0.190	0.118	0.156	0.076	0.012	0.016	0.014	0.215	0.214
現社	174,059	33,912	52,430	55,675	174,707	0.000	0.000	0.012	0.383	0.013	0.350	0.155	0.220	0.207	0.029	0.052	0.027	0.446	0.445
倫理	28,743	8,174	7,913	2,889	0	28,826	0.000	0.003	0.065	0.003	0.060	0.017	0.028	0.041	0.007	0.008	0.003	0.074	0.073
政経	47,890	8,402	14,362	5,934	0	0	48,190	0.003	0.088	0.004	0.079	0.022	0.034	0.053	0.011	0.012	0.005	0.123	0.122
数1	9,711	1,788	2,771	1,280	4,550	1,049	1,362	10,578	0.000	0.010	0.004	0.003	0.004	0.013	0.002	0.005	0.003	0.026	0.026
数1A	257,771	45,654	66,059	77,760	148,986	25,309	34,384	0	272,153	0.016	0.608	0.284	0.366	0.331	0.046	0.071	0.033	0.695	0.691
数2	9,588	1,730	3,496	1,556	5,189	1,065	1,562	3,857	6,202	10,153	0.000	0.003	0.006	0.016	0.003	0.006	0.002	0.028	0.025
数2B	226,236	40,447	57,430	74,111	136,059	23,185	30,673	1,407	236,660	0	238,354	0.261	0.357	0.287	0.041	0.063	0.028	0.610	0.608
物1	98,207	7,981	11,507	45,771	60,444	6,799	8,533	1,019	102,704	1,060	101,672	103,922	0.224	0.018	0.000	0.011	0.014	0.266	0.265
化1	137,957	13,371	20,559	60,606	85,814	10,989	13,203	1,663	142,513	2,163	138,941	87,045	144,980	0.120	0.009	0.000	0.015	0.371	0.369
生1	135,513	30,980	47,036	29,724	80,476	15,885	20,620	5,097	128,871	6,080	111,706	7,057	46,583	139,413	0.017	0.056	0.000	0.356	0.354
地学I	19,301	5,507	7,410	4,763	11,327	2,857	4,248	894	17,979	1,100	15,875	0	3,346	6,438	19,368	0.011	0.011	0.050	0.050
理総A	29,498	7,896	11,824	6,129	20,407	3,174	4,704	1,752	27,553	2,260	24,342	4,359	0	21,936	4,122	29,759	0.008	0.076	0.076
理総B	14,313	2,671	3,944	5,391	10,383	1,212	2,047	1,003	12,814	856	10,874	5,602	5,791	0	4,397	3,182	14,446	0.037	0.037
英語	367,982	71,958	115,389	83,586	173,791	28,648	47,747	10,137	270,398	10,002	237,443	103,386	144,327	138,462	19,307	29,645	14,356	384,036	0.975
リス	364,269	71,562	114,498	83,356	173,402	28,542	47,346	9,953	268,852	9,895	236,593	103,079	143,716	137,714	19,269	29,607	14,330	379,467	0.975

表5：主要科目の共分散行列（下三角）と相関行列（上三角）

	国語	国語現	国語古	国語漢	世史B	日史B	地理B	現社	倫理	政経	数1	数1A	数2	数2B	物1	化1	生1	地学	理総A	理総B	英語	リス
国語	1309.2	0.754	0.780	0.807	0.498	0.492	0.501	0.535	0.602	0.533	0.444	0.446	0.343	0.390	0.533	0.528	0.520	0.431	0.234	0.395	0.611	0.572
国語現	470.8	298.1	0.373	0.321	0.400	0.416	0.374	0.445	0.463	0.485	0.335	0.330	0.258	0.276	0.422	0.398	0.431	0.342	0.243	0.352	0.484	0.467
国語古	350.9	80.1	154.4	0.561	0.432	0.414	0.392	0.393	0.473	0.413	0.352	0.351	0.272	0.322	0.401	0.421	0.384	0.343	0.118	0.230	0.499	0.447
国語漢	487.5	92.5	116.4	278.7	0.336	0.318	0.407	0.392	0.460	0.351	0.379	0.376	0.279	0.328	0.431	0.427	0.395	0.317	0.147	0.311	0.450	0.422
世史B	361.6	142.5	109.3	109.8	453.5			0.559	0.596	0.634	0.474	0.459	0.438	0.460	0.543	0.607	0.516	0.521	0.238	0.350	0.615	0.487
日史B	293.8	119.8	84.2	89.7		287.3		0.529	0.595	0.628	0.423	0.430	0.392	0.407	0.511	0.570	0.498	0.474	0.283	0.381	0.576	0.471
地理B	250.9	100.8	71.3	78.8			258.3	0.447	0.439	0.524	0.385	0.440	0.310	0.428	0.530	0.585	0.498	0.507	0.296	0.439	0.542	0.489
現社	226.2	100.2	59.8	66.2	163.4	116.9	86.4	180.4			0.369	0.339	0.334	0.298	0.425	0.448	0.466	0.486	0.291	0.397	0.511	0.457
倫理	268.2	114.6	81.6	72.1	181.3	159.4	117.0		249.7		0.529	0.454	0.406	0.437	0.525	0.579	0.555	0.524	0.271	0.375	0.602	0.508
政経	294.1	125.8	78.7	89.6	218.4	167.6	139.2			242.4	0.466	0.478	0.429	0.444	0.565	0.592	0.556	0.475	0.378	0.468	0.587	0.511
数1	441.6	160.3	117.2	164.1	260.1	181.1	166.3	128.5	215.9	193.0	728.9		0.762	0.686	0.575	0.584	0.564	0.422	0.324	0.409	0.578	0.485
数1A	312.4	117.6	88.7	106.0	201.9	151.0	137.3	91.8	142.7	153.2		458.8	0.667	0.747	0.658	0.653	0.531	0.396	0.284	0.330	0.592	0.490
数2	225.7	85.4	60.6	79.6	172.7	114.8	94.9	80.6	125.0	124.6	394.4	228.7	357.8		0.512	0.568	0.482	0.335	0.259	0.319	0.514	0.390
数2B	278.5	104.1	84.8	89.6	205.8	147.3	147.5	88.9	149.0	148.1	360.8	349.0		523.2	0.658	0.670	0.492	0.369	0.239	0.253	0.578	0.460
物1	345.5	141.0	93.9	110.5	238.0	176.9	158.3	105.3	153.4	169.2	340.8	266.4	202.8	300.4	409.2	0.695	0.211		0.485	0.433	0.631	0.554
化1	375.9	151.6	109.9	114.4	300.6	221.5	203.4	127.5	200.6	203.7	330.8	297.1	211.7	341.8	307.9	521.0	0.535	0.243		0.411	0.656	0.551
生1	282.2	122.2	75.8	84.1	172.3	138.7	137.5	105.8	128.7	139.4	271.2	186.3	154.5	174.4	119.1	209.7	293.9	0.188	0.366		0.534	0.478
地学	273.9	122.8	82.3	68.8	218.1	167.2	178.8	145.8	151.9	135.6	234.4	182.6	125.0	177.1		76.1	46.7	484.2	0.288	0.395	0.458	0.399
理総A	88.9	50.7	16.8	21.4	56.0	53.2	59.2	48.0	48.7	66.2	114.7	69.0	62.2	58.8	123.9		69.6	63.3	165.0	0.560	0.203	0.212
理総B	172.4	82.4	36.1	53.8	102.4	86.1	92.7	71.9	76.8	94.8	173.9	93.1	85.9	72.2	110.1	119.3		93.8	104.6	197.6	0.312	0.321
英語	856.2	324.0	240.7	291.5	504.1	367.3	319.1	251.6	343.6	354.1	637.1	480.4	365.6	490.8	491.1	562.1	338.8	363.1	90.0	160.0	1536.6	0.759
リス	167.0	65.4	44.9	56.7	82.5	63.4	56.5	46.7	57.9	63.9	117.4	82.1	61.1	80.5	88.8	97.1	62.5	64.9	19.9	34.6	244.4	67.9

表6：FIMLによる平均構造（下段）・共分散構造（下三角）・相関構造（上三角）

	国語	国語現	国語古	国語漢	世史B	日史B	地理B	現社	倫理	政経	数1	数A	数2	数B	物1	化1	生1	地学	理総A	理総B	英語	リス
国語	1319.9				0.511	0.516	0.538	0.570	0.673	0.530	0.446	0.471	0.415	0.436	0.513	0.528	0.549	0.480	0.287	0.446	0.616	0.580
国語現		299.7	0.377	0.324	0.407	0.435	0.399	0.472	0.504	0.496	0.329	0.348	0.302	0.310	0.422	0.407	0.453	0.379	0.307	0.404	0.490	0.475
国語古			81.3	0.563	0.433	0.428	0.427	0.435	0.529	0.402	0.351	0.365	0.351	0.372	0.420	0.403	0.379	0.165	0.288	0.504	0.454	
国語漢				93.9	0.366	0.352	0.437	0.425	0.546	0.339	0.367	0.391	0.341	0.365	0.400	0.412	0.424	0.367	0.182	0.337	0.456	0.430
世史B					397.7	151.0	115.5	131.2	459.3		0.564	0.636	0.623	0.485	0.530	0.612	0.529	0.539	0.275	0.398	0.626	0.501
日史B						323.4	129.9	91.9	101.6		0.588	0.628	0.630	0.441	0.528	0.582	0.519	0.518	0.324	0.435	0.597	0.487
地理B							327.2	115.6	89.2	122.4		0.529	0.561	0.560	0.427	0.489	0.440	0.477	0.349	0.508	0.575	0.505
現社								288.5	114.0	99.1	168.5	136.1	123.4	194.2		0.411	0.407	0.391	0.378	0.456	0.552	0.488
倫理									421.8	150.5	113.6	157.6	166.7	161.9		0.547	0.518	0.491	0.502	0.460	0.644	0.560
政経										300.5	133.9	78.2	88.4	243.2		0.450	0.486	0.469	0.471	0.494	0.590	0.516
数1											431.4	151.8	116.2	163.4		0.762	0.614	0.586	0.348	0.412	0.579	0.480
数A												373.5	131.7	99.2	142.7		0.723	0.770	0.440	0.359	0.608	0.507
数2													289.6	103.7	82.8	113.1		0.436	0.327	0.389	0.572	0.466
数B														378.1	128.0	104.3	145.8		0.328	0.386	0.606	0.480
物1															382.4	149.8	95.1	137.4		0.525	0.613	0.534
化1																451.5	165.9	123.1	0.492	0.660	0.555	
生1																	354.1	139.1	89.0	0.418	0.565	0.510
地学1																		397.0	149.4	0.330	0.450	0.447
理総A																			137.8	0.291	0.289	
理総B																				235.7	0.392	0.388
英語																					880.0	0.761
リス																						174.4
平均(FIML)																						122.75
SD(FIML)																						36.33
平均(表2)																						123.55
SD(表2)																						36.18

2003)を用いた。

表2と表6における平均・SDの違いは、表2のそれらは、単に観測された得点から算出したものに過ぎないが、表6のそれらは、欠測メカニズムがMAR、分布を多変量正規分布と仮定した下で、「仮に全員が当該科目を受験したら得られるであろう平均とSD」である。したがって、表6の平均およびSDのほうが、集団全体の統計量を知る上で有意義であり、少なくとも表2を参考するよりは正しい周辺統計量である。

たとえば、「数学I(表中,数1)」や「数学II(表中,数2)」など、表2の平均と比べて、FIML平均のほうが高い科目は、もともと教科学力が低い集団が当該科目を選択しているため、仮に全員を受験したときの指標であるFIML平均のほうが得点が高く推定されている。逆に、「倫理」「数学I・A(表中,数1A)」「数学II・B(表中,数2B)」など、FIML平均のほうが低い科目は、そもそも教科学力が高い集団がそれらの科目を選択しているといえる。なお、「日本史B(表中,日史B)」のFIML平均と表2の平均が小数点2桁までたまたま一致しているが、小数点3桁以下は異なっている。

FIMLによるSDは、仮に全員を受験したときのSDなので散らばりが大きくなる。したがって、表2で得られたSDより大きくなっている。同様に、表5と表6における共分散・相関の違いは、表5のそれらは、それぞれの2つの科目を同時に選択した受験生から単純に算出した共分散・相関であるが、表6のそれらは、欠測メカニズムがMAR、分布を多変量正規分布と仮定した下で、「仮に全員が全科目を受験したときの共分散・相関行列」である。したがって、少なくとも集団全体の周辺統計量を知る上では、表5よりも表6のほうが正しい。

注目すべきは、平成18(2006)年度から新たに導入された「英語リスニング(表中,リス)」と他科目との相関である。とりわけ、「英語(筆記)(表中,英語)」との相関は興味深く0.761であった(表6)。したがって、互いの説明率は58%ほどとなる。この大きさは、「数学I・A」と「数学II・B」の相関0.770、「数学I」と「数学II」の相関係数0.762に次いで大きい。さらに、「英語(筆記)」「英語リスニング」の2科目にとって、お互いに最大の相関係数の値であった。このことは、それら2科目が「英語力」という心理量を測定する上での弁別的妥当性と同時に、収束的妥当性の高さを物語っているとさえ言えよう。つまり、筆記テストとリスニングテストは、ほどよく棲み分けていると言えるだろう。

「英語」と「リスニング」の相関が高いことをもっ

て、リスニング不要論が一部で起こってはならない。相関係数はリスニングテストの要否の論材にはならない。テストは、社会の公具(public tool)であり、ただ存在しているだけで、子どもたちの学校生活や日常生活に影響を与える。したがって、わが国の子どもたちが、将来、英語を使って外国人と柔軟なコミュニケーションが取れるようになることを期待するならば、英語リスニングテストは続行しなければならない。リスニングはコミュニケーションの基本だからである。

### 3. 自己組織化マップによる分析

本節では、2006年のNCT2006における教科学力構造を自己組織化マップ(SOM)を用いて分析した結果を報告する。その際、性差について詳細に見ていくことにする。SOMは、クラスタリング手法の1つであり、マーケティング・金融分野など、さまざまな応用分野で用いられている。テストの分野では、荘島ら(2006)によるNCT2005の分析がある。SOMは、もともとは人間の視覚野の情報処理プロセスを模式化したモデルであり、教師信号なしのニューラルネットワークモデルなどとも言われ、高次元の情報を2次元に圧縮してコンパクトに視覚的に表現してくれる。

分析対象は、ここでも現役生389,235人であるが、用いたソフトウェアViscovery SOMine Plus 4.0(Eudaptics Software Gmbh, 2001)の制約により、50,000件をランダムサンプリングして入力データとした。2次元格子は、19,705(=131×151)の六角形のユニットを用いた。また、張力(tension)はデフォルトの0.5とした。

図1は、表2における全科目を入力変数としたSOMの成分マップである。ただし、「国語」は「国語現」「国語古」「国語漢」の3変数に分解して入力変数とした。観測変数によって単位が異なるので、データは標準化(variance rescaling)して入力変数とした。

得られたマップから、SOMにおけるWard法によってマップを分類した結果、荘島ら(2006)と同じく、SOMineが最も支持するクラスタ数は3であった。それは、学力層の上・中・下位に対応している。し

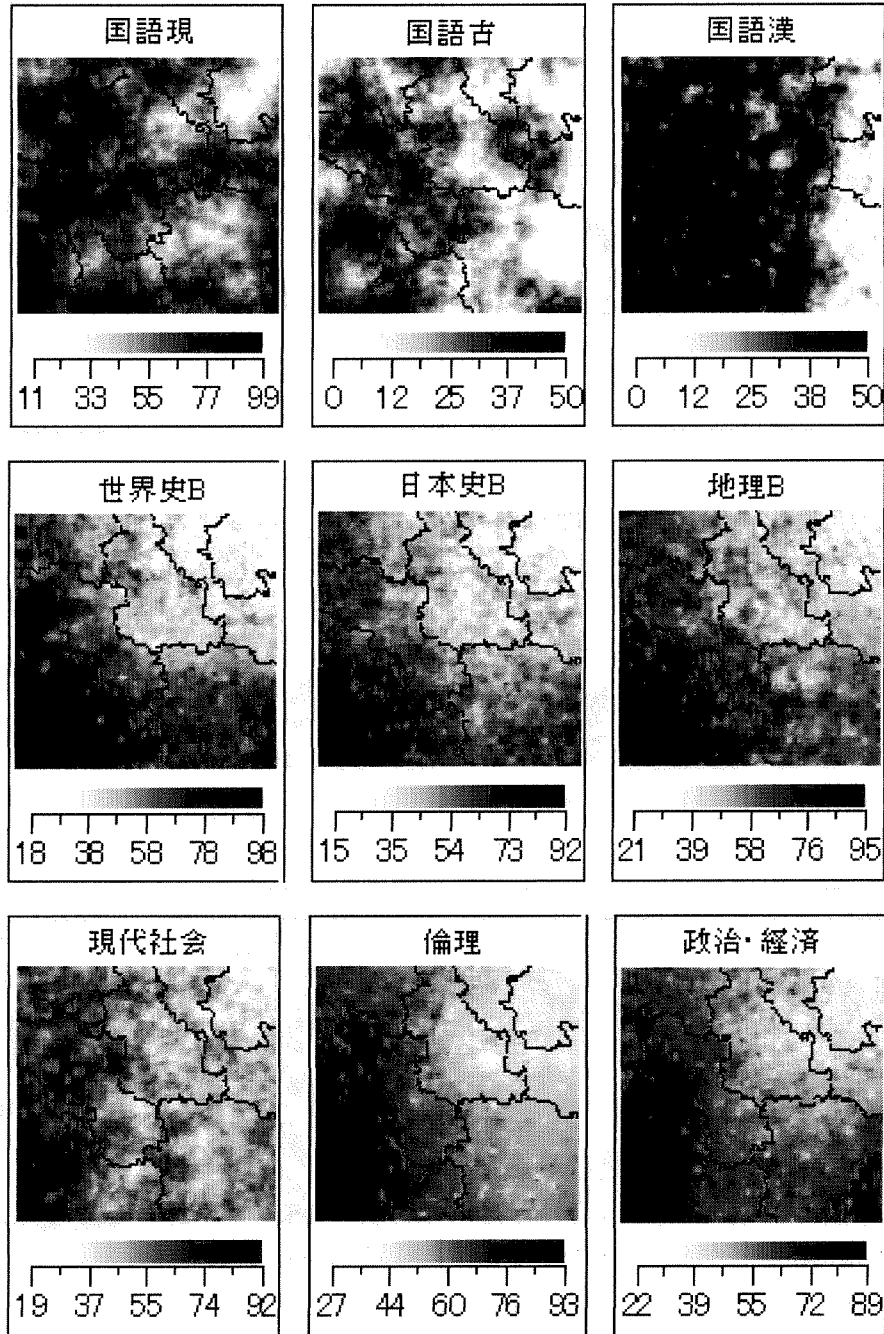


図1：主要科目の成分マップ（その1）

かし、39万人の受験者（入力は5万件であるが）を3分類するのは乱暴であるので、説明の都合上、クラスタ数を8とした。実際は、クラスタ数が5である解が次点であったが、クラスタ数が5以上の解の印象はいずれもそれほど変わらず、マップの左下から右上にかけての対角線から対称的にクラスタが順次分離していく展開であった。そのような展開は、莊島ら（2006）で観測

されたものと同一であった。

まず、図1について目に付くのは、「国語漢」の得点が二極化しており、多くの受験者は50点満点であったことがわかる。しかし、そもそも漢文を必要としないため、漢文に解答していない受験生をデータから識別することができない（推測はできる）。したがって、漢文で0点であった受験生の多くは、漢



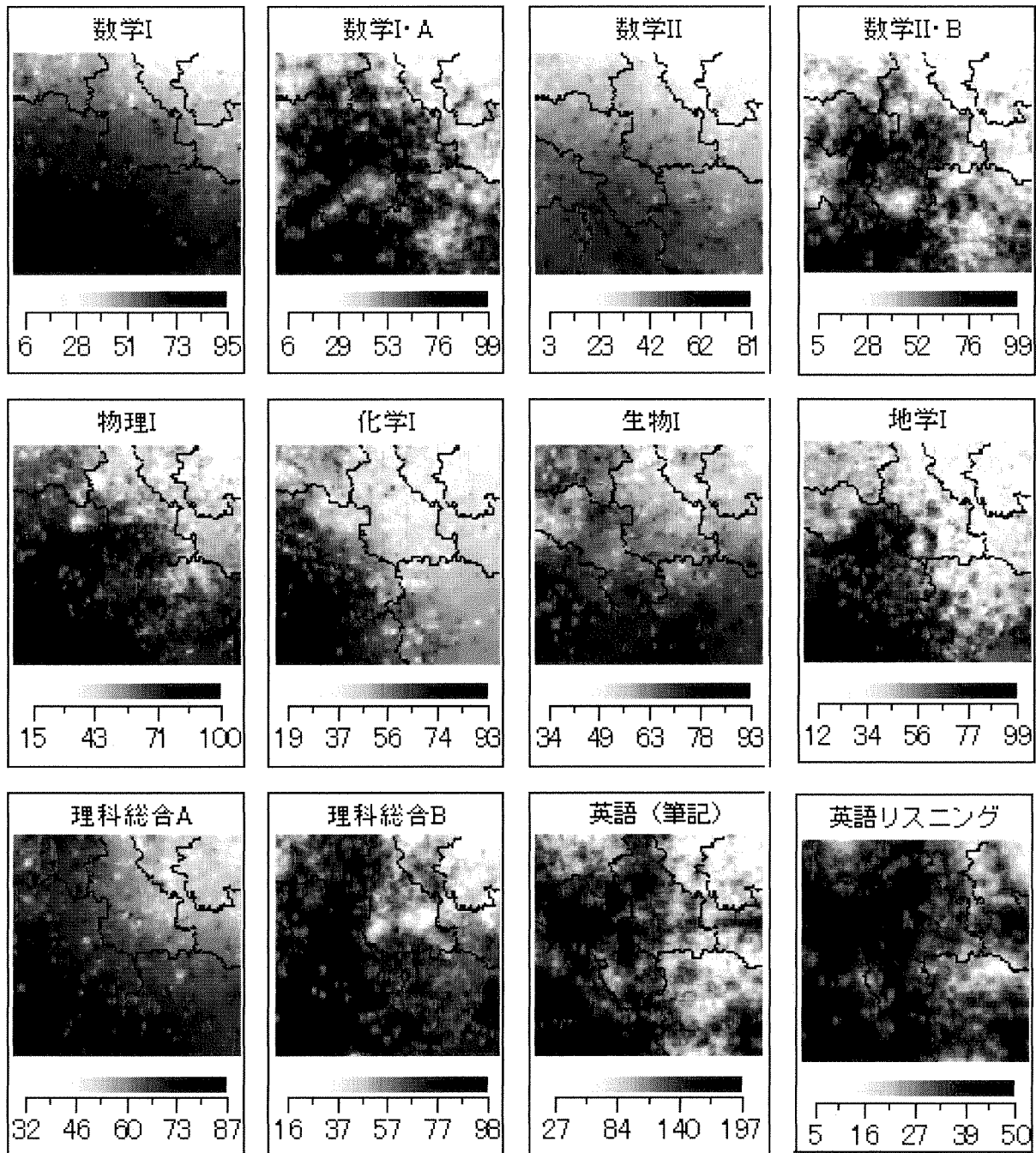


図1：主要科目の成分マップ（その2）

文に解答していない者を多く含むと思われる。また、「数学II」で60点以上をとった受験者は稀であったこともわかる。

続いて、各成分マップの左下には、ほとんどの科目について得点が高い、いわゆる学力が高い受験生たちが集中しており、逆に、マップの右上には学力層の低い受験者が集中している。つまり、左下から

右上にかけての対角線には、「一般教科学力」とも呼べるような軸が存在している。その対角線を境に左上には文系科目に強い受験生が集中し、右下には理系の学生が集中している。これは、莊島ら（2006）でも観察されたことである。

莊島ら（2006）で得られたマップと本研究で得られたマップにおいて、互いのマップの位相が異なる



ため、厳密な比較はできないが、彼らの結果と比較して異同がいくつかある。まず、「生物I」は、ほとんど文理の区別がなくなっていることが共通している。ただし、科目編成が変更されたので、NCT2006の「生物I」に対応するのは、NCT2005では「生物IB」である。また、「地学I」「理科総合A」「理科総合B」は、文理ともに分布しているが、荘島ら（2006）では、「地学IA」と「総合理科」が文理の別がない。しかし、「世界史B」「日本史B」「地理B」の「地理歴史」教科は、NCT2006では文理にわたって分布しているが、NCT2005では文系に集中して分布していた（荘島ら、2006）。しかし、「現代社会」「倫理」「政治・経済」の「公民」教科は、主に対角線の左上に分布していることは共通していた。

また、「国語現」「国語古」「数学I・A」「数学II・B」「英語」「リスニング」のマップの勾配が滑らかではない。これは、学力層の高低に限らず、どの受験者も真っ先に勉強する主要科目中の主要科目であるからである。したがって、最終的にマップの左下に位置することができた受験生は、余力を残してしっかりと他科目の得点を高めることに成功した総合的に学力の高い受験生ということが出来る。

便宜のため、図2のように、クラスタ名を名づける。ただし、これらのクラスタは、学力層の上位から下位に層別化している意味合いが強いので、層I…層VIIIと呼ぶことにする。また、図3(A)は、マップに分布する男性をプロットしたものであり、図3(B)は、女性をプロットしたものである。

図3(A)(B)の結果より、性差による特徴が見えてくる。まず、NCT受験者全体の約55%、つまり、20人に11人は男性であるので、男性のほうが全体に密度が高い。しかし、それを踏まえたとしても、全体にマップの底辺（層I, II, IVの下部）に女性が少ない。このエリアは、図3より、「数学I・A」「数学II・B」「物理I」「化学I」を得意とする集団と位置づけられ、男性と比べて相対的に女性がそれら理系科目を好まないからである。反対に層III, VIの上部は、女性に比べて男性の密度が疎である。女性の中には、「英語」「国語現」「国語古」「国語漢」の語学科目を特別に選好する集団が知られており、特に女性で顕著であるが、男性でもそれらの科目を選好する集団がいる。

したがって、彼らが当該エリアに分布している集団に該当すると思われる。また、最も学力層が低い集団だと推測される層VIIIにも男性が集中している

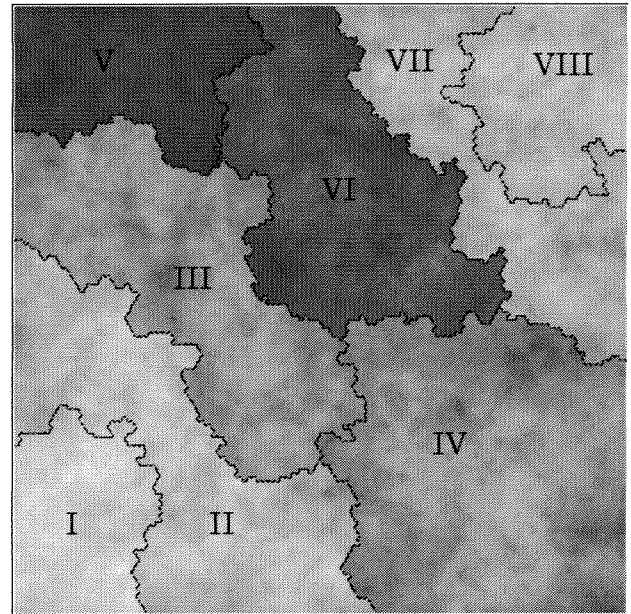


図2：SOM Ward法による8つの層

ことがわかる。つまり、男性のほうが、女性に比べて相対的に学力層の格差が大きい。特に下位層は男性が集中している。

続いて、前期日程で国立大学の試験に合格した集団について見ていくことにする。現役生全体では、233,775人が前期日程で受験し、そのうちの62,429+6,641人が合格した。ただし、6,641人は受かったが入学しなかった。

入力データとして用いた5万件のデータのうち、27,458人の現役生が前期日程で受験し、合格した7,426+758人のSOMマップ上での配置を見てみる。なお、758人は、受かったが入学しなかった人数である。全国の国立大学の学部を「人文社会系」「教育系」「文理系」「理工系」「農水産系」「医歯薬系」「保健系」「家政系」「芸術系」の分類し、それらの学部合格した男女をマップに布置したものが図4である。男性は0、女性は1で表現されている。

まず、全体的に左下に分布が偏っているのは、彼らが試験に受かるような教科学力が高い人たちだからである。また、全般的に学部のカテゴリの仕方にもよるが人文社会系と理工系に人が多い。これは、全国にそのような学部が多いためである。また、人文社会系の密度が高いエリアはマップの左下から右上の対角線上から左上に位置するのに対し、理工系学部の合格者は、マップの左下と中央の2箇所主に分布している。左下に分布している一群は、文理間

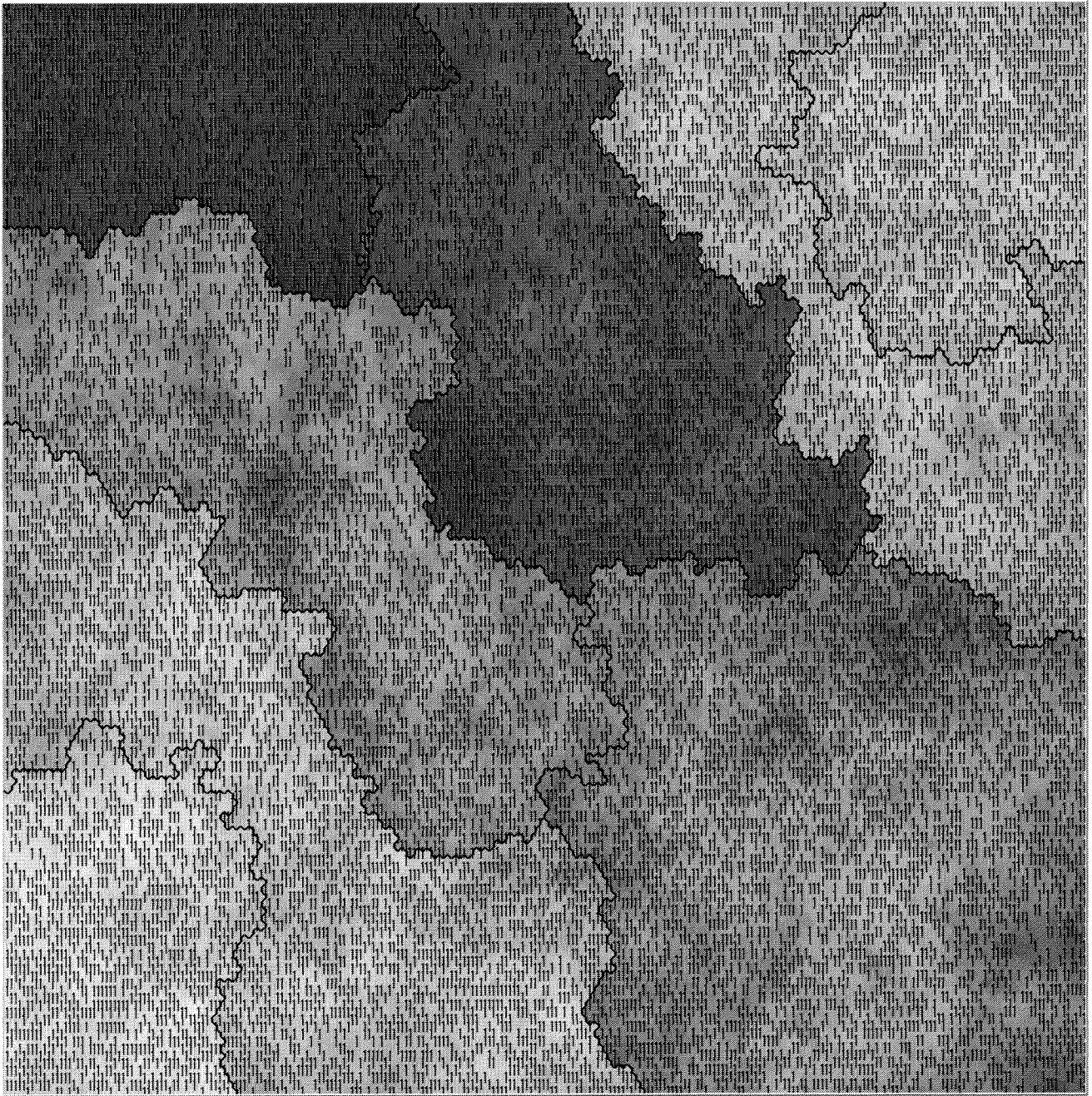


図3(A) : 男性のプロット

わず全ての科目について教科学力が高い、いわば一般学力の高い集団である。マップの中央に位置する集団は、図1より、「数学I・A」「数学II・B」「物理I」などの理系科目は得意とするものの、あまり文系科目に強くない集団ということがわかる。また、同様に図1の成分マップと参照しながら考察すると、文理系は、教養学部などを含んでおり、文理にバランスの取れているといえよう。また、農水産系・医歯薬

系は、「数学I・A」「数学II・B」「生物I」「英語」をほとんどトレースするように分布している。

性差については、やはり理工系は男性に集中している。しかしながら、同じ理系学部である農水産系・医歯薬系・保健系などは、男女のバランスがとれている。保健系・医歯薬系では、女性のほうが多い結果が得られている。保健系では、幅広い文理能力をもった女性が保健系学部合格したことが分かる。



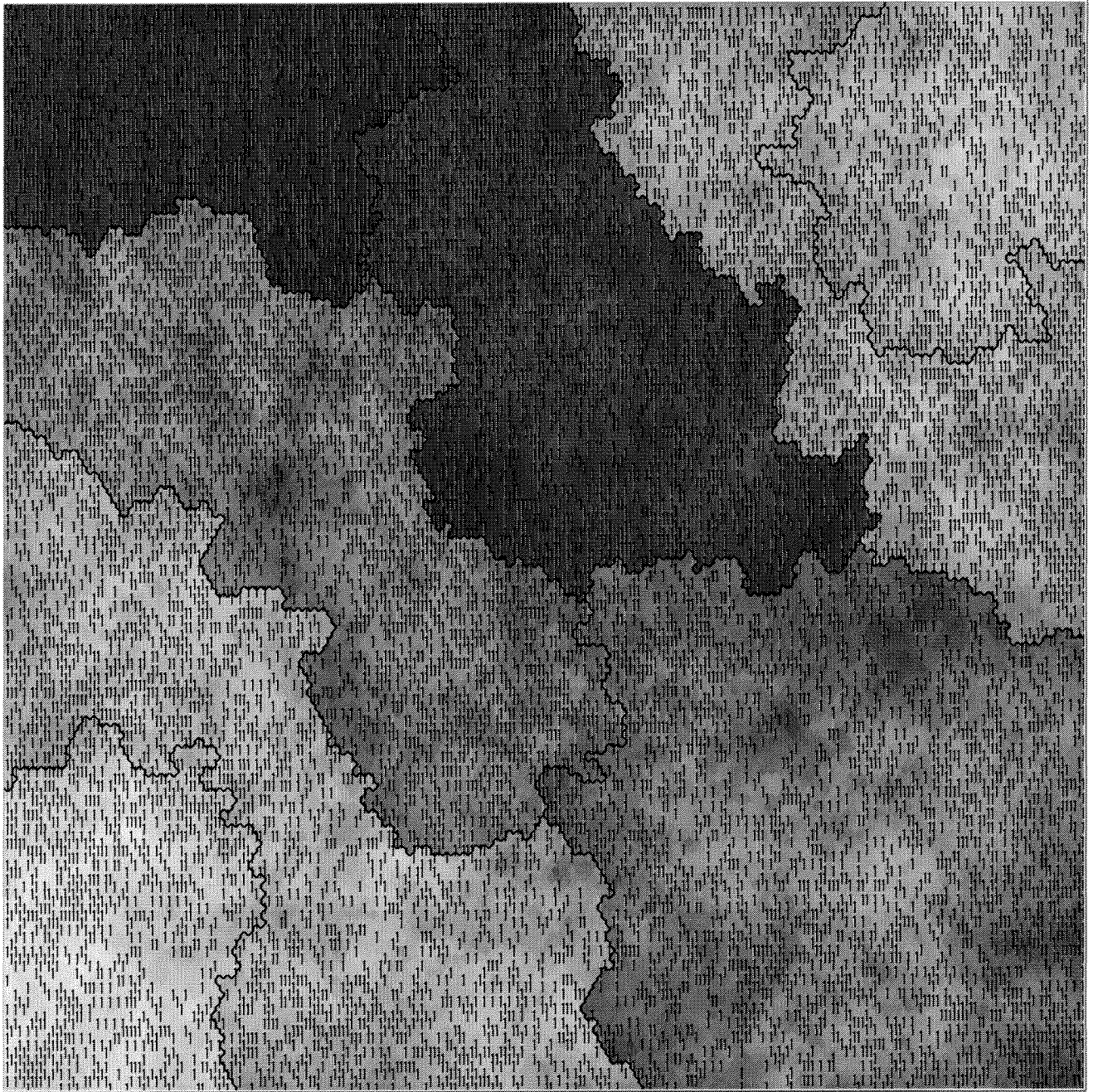
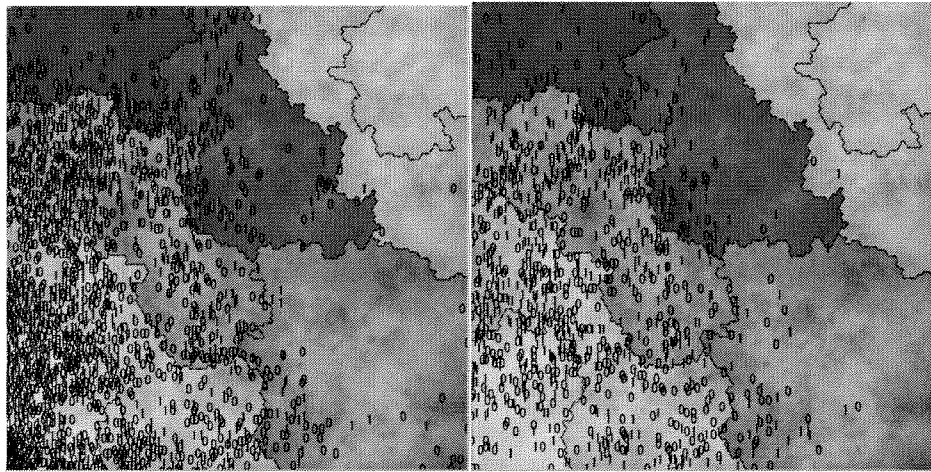


図3(B)：女性のプロット

特に、医歯薬系では、女性が男性よりも広範囲に分布しているが、図の左下の学力の比較的高い層は男性と女性の割合が均衡しているが、図の左下から離れるにつれて女性が分布している。今回の分析では、医歯薬系とひとまとめしたが、医学系・歯学系・薬学系と詳細に分析をすることで性差の違いがより見えてくるだろう。が医歯薬系学部に入學している。したがって、昨今話題となっている「女子中高生の

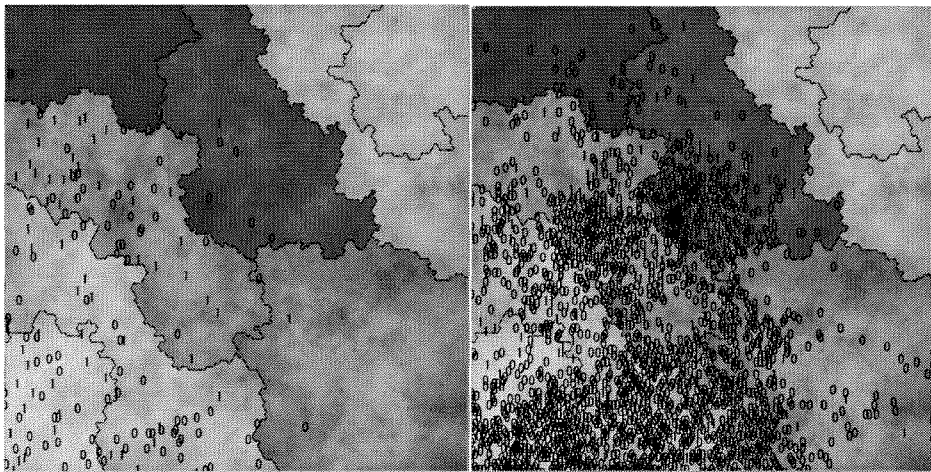
理系離れ」における理系とは、理工系のことに他ならない。また、家政系・芸術系学部には女性の入學者が多い。

ただし、人文社会系といっても、文学・外国語・国際文化・法・経済・経営学部などが含まれている。したがって、それらの学部でも性差は確認されるだろう。同様に、医歯薬系についても、医学部・歯学部・薬学部で性差があると思われる。



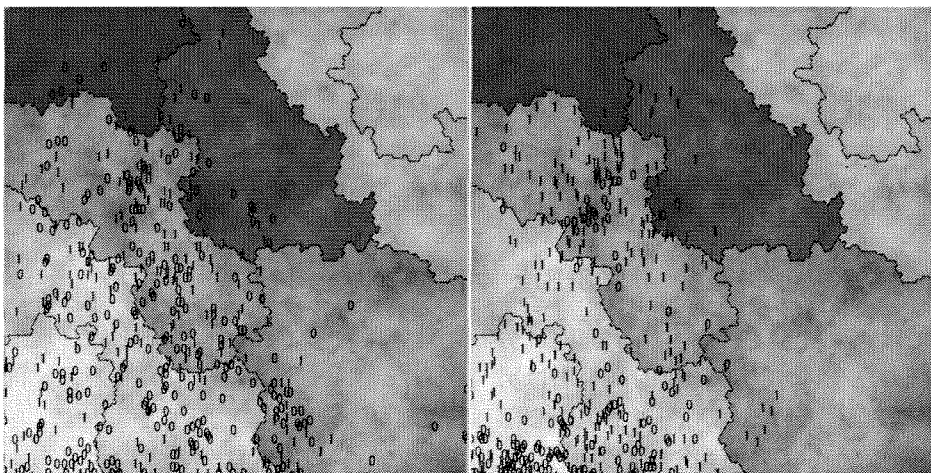
(A) 人文社会系

(B) 教育系



(C) 文理系

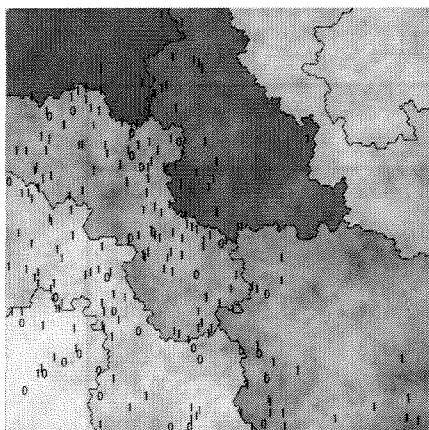
(D) 理工系



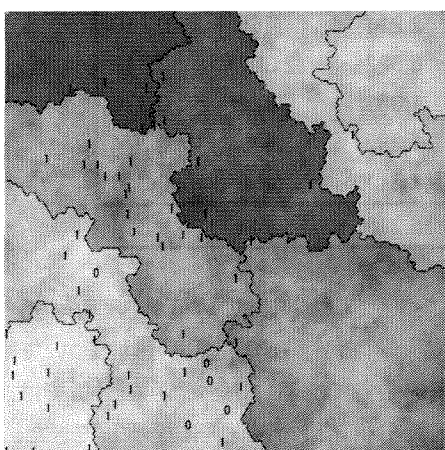
(E) 農水産系

(F) 医歯薬系

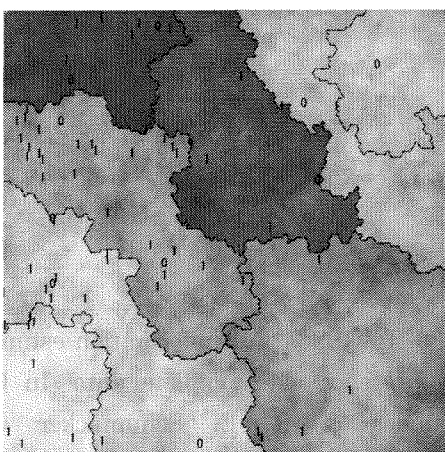
図4：学部系統別の性別プロット



(G) 保健系



(H) 家政系



(I) 芸術系

図4: 学部系統別の性別プロット (つづき)

参考として表7には, NCT2006における前期日程の合

格者を学部カテゴリ×性別のクロス表で示した. 表7では, より詳細に学部系統を14分類している. いま, 学力を問題にしているのだから, 前期日程の受験者・合格者・合格率を示しているが, 合格した者のすべてが入学手続きをとったわけではないことに注意されたい. やはり, 学部によって性差はある. 注目すべきは, 理学系学部も工学系学部も合格率における性差は小さいということである. しかしながら, 受験者数・合格者数の男女差は大きいといわざるを得ない. なお, 大津・橋本・荘島・石塚(2006)も平成16年度センター試験の分析から同様な結果を得ている.

#### 4. まとめ

本研究では, 平成18(2006)年度大学入試センター試験(NCT2006)における現役生・健常者・本試験受験生の教科学力構造をFIMLによる平均・共分散・相関構造を示した後, 自己組織化マップ(SOM)によって分析した結果を示した. NCT2006において最大の注目であったリスニングについては, FIMLを用いて得られた他科目との相関構造より, 良い結果を示したといえる. また, SOM分析では, 性別による比較を中心に行い, 昨今言われているような女性の理科離れについては, 特に国立大学の理工系学部受験者・合格者における男女差が大きかった.

これらの性差は, 現前として存在するものである. ただし, 図3のマップを見れば一目瞭然であるが, 理系が得意な女性もいれば, 古典・漢文が得意な男性も少なくない. 一般に男性のほうが教科学力の高いものから低いものまでのバラツキが大きく(Shojima, et al., 2006), 特に学力の上位層と下位層は男性が多いことが知られているが, 図3(B)の層Iと層VIIIより, 女性にも非常に教科学力の高い受験生から低い受験生まで, 実にさまざまである. したがって, これらの性差によって男性がどうであるとか, 女性がこうであるといった短絡的な二分法は厳に慎むべきであろう. そういった二分法は, 進路選択や就職活動などにおいて性ステレオタイプな雰囲気を醸成する可能性もあり, 最終的には社会における適切な人材配置に悪影響を及ぼすことになる.

また, 示してきた性差のうち, 性の格差というよりも性の特徴といったほうが適切なものもある. た



表7：前期日程合格者の人数（H18現役生）

	男性			女性		
	受験者	合格者	合格率	受験者	合格者	合格率
人文学系	5,814	2,028	0.349	11,853	4,375	0.369
教育系	11,697	3,851	0.329	16,204	5,631	0.348
社会科学系	18,570	6,278	0.338	9,213	3,576	0.388
理学系	6,511	2,715	0.417	2,451	975	0.398
工学系	32,925	14,928	0.453	4,068	1,769	0.435
農水産系	6,174	2,420	0.392	3,892	1,555	0.400
医歯薬系	7,756	1,597	0.206	8,754	2,649	0.303
保健系	1,093	257	0.235	5,146	1,582	0.307
家政系	234	77	0.329	1,554	504	0.324
芸術系	665	112	0.168	2,211	465	0.210
文科総合系	5,636	1,785	0.317	6,622	2,307	0.348
理科総合系	2,157	789	0.366	1,651	621	0.376
文理総合系	5,819	2,381	0.409	1,282	479	0.374
二部	1,938	698	0.360	1,311	456	0.348
	106,989	39,916	0.373	76,212	26,944	0.354

たとえば、家政系学部・保健系学部に女性が多いのは、それすらも後天的に構築されたものといえなくもないが、少なくとも女性の選好が大きく働いている。同時に、男性が家政・保健学部にな少ないのも性差というよりも性の特徴といったほうがよい。

少なくとも、図3で明示されている事柄の最たるものは、男女の文理の能力差は、大きくないということである。つまり、教科学力という心理量での性差はなくはないが決して大きくない。これは、間違いない。しかしながら、行動レベルで、すなわち、国立大学前記日程の学部選択行動を見たとき、性の選考を抜きにしても、女性の理工系離れによる男女格差は大きいといわざるを得ないだろう。

また、Shojima et al. (2006)の結果より、男性のほうが文理共に能力のバラツキが大きく、特に理系上位者には男性が多いのは事実である。しかしながら、依然として理系上位の女性も多い。したがって、学力の高い大学の理工系学部男子が偏っているのは、学部選択の選好によるものが主たる原因であると思われる。ただし、文理能力の性差に対するより詳細な検討は、センター試験が測定している能力よりも高いレベルに対して識別力をもつテストの結果などを参考にしないと明らかにならない。

この点については、NCTの分析から追跡することが難しい私立大学における学部選択行動についても報告が待たれるところである。また、今回の分析では、SOMを中心とした分析結果を視覚的に提示してきた。SOMは視覚的に有意義な知見を得ることができるが、より精緻に検討するには、性差に関する情報など、詳細な統計量を検討することも重要である。

#### 謝辞

本論文を作成するにあたり、腰越滋先生（東京学芸大学）のご意見が大変参考になりました。厚くお礼申し上げます。

#### 参考文献

- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced Structural Equation Modeling: Issues and Techniques*. Lawrence Erlbaum Associates, Inc. (pp. 243--277).
- Bock, R. D., Gibbons, R., & Muraki, E. (1994). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.

- 大学入試センター (2006). 平成 18 年度センター試験 (実施済). ([http://www.dnc.ac.jp/center\\_exam/18exam\\_index.html](http://www.dnc.ac.jp/center_exam/18exam_index.html)).
- Eudaptics Software GmbH (2001) *Viscovery SOMine Plus 4.0*. (国内連絡先: 高松市, マインドウェア).
- 岩崎 学 (2002). 不完全データの統計解析 エコノミスト社.
- Kohonen, T. (1995). *Self-organizing maps*. Springer.
- Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Muraki, E. & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, 19, 73-90.
- 大津起夫・橋本貴充・荘島宏二郎・石塚智一 (2006) 平成 16 年度大学入試センター試験前期日程における志願状況と科目別得点の状況について 大学入試センター研究紀要, 35, 109-129.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
- 徳高平蔵・岸田悟・藤村喜久郎 (1999). 自己組織化マップの応用—多次元情報の 2 次元可視化— 海文堂.
- Schafer, J. L. (1999) *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.
- SmallWaters Corporation (2003). *AMOS 5.0*. (<http://www.smallwaters.com/>).
- 荘島宏二郎 (2006) 2005 年度センター試験における教科学力構造 山森光陽・荘島宏二郎 (編著) 学力—いま, そしてこれから— ミネルヴァ書房 pp.119-134.
- 荘島宏二郎・大津起夫・田栗正章 (2006). 自己組織化マップと完全情報最尤法から見た平成 17 年度センター試験の教科学力構造 日本テスト学会, 2, 75-89.
- Shojima, K., Otsu, T., Mayekawa, S., Taguri, M., & Yanai, H. (2006) Factor structure of National Center Test 2005 by full-information pseudo-maximum likelihood method. Paper submitted for publication.

# ピア・アセスメント支援システムを利用した 紙媒体レポートの相互評価の実践

藤原 康宏<sup>†</sup> 西久保 健太<sup>‡</sup> 永岡 慶三<sup>‡</sup>

<sup>†</sup> 岩手県立大学 〒020-0193 岩手県滝沢村滝沢字巢子 152-52

<sup>‡</sup> 早稲田大学 〒359-1192 埼玉県所沢市三ヶ島 2-579-15

E-mail: <sup>†</sup> fuji@iwate-pu.ac.jp,

**あらまし** 実践的な課題やグループワークの場面で、学習者間の相互評価を導入した実践がよく行われている。本稿では、レポートの改善のために、ピア・アセスメント支援システムを利用した相互評価の実践について述べる。ピア・アセスメント支援システムを使用する場合、レポートも電子媒体であることが多かったが、レポートの内容によっては、画面上では読みにくい場合がある。本実践では、紙媒体のレポートの読み易さと、システムを利用した即時フィードバックの特徴を組み合わせた。実践が終了した後に行ったアンケートでは、多くの学習者が相互評価は有意義であると考えていることが分かった。

**キーワード** 相互評価, 教育評価, ピア・アセスメント支援システム, 教育実践

## A practice of peer assessment on paper based reports using peer assessment support system

Yasuhiro FUJIHARA<sup>†</sup> Kenta NISHIKUBO<sup>‡</sup> and Keizo NAGAOKA<sup>‡</sup>

<sup>†</sup> Iwate Prefectural University 152-52 Takizawa-sugo, Takizawa-mura, Iwate, 020-0193 Japan

<sup>‡</sup> Waseda University 2-579-15 Mikajima Tokorozawa-shi, Saitama, 359-1192 Japan

E-mail: <sup>†</sup> fuji@iwate-pu.ac.jp

**Abstract** Peer assessment is well known as the effective evaluation way for project based learning, group work and so on. We practiced peer assessment for formative evaluation in a university class. In the case of using peer assessment support system, the evaluation objects are mostly electronic files. It is often hard to read reports on screen. We used peer assessment support system on paper based report. As a result of the practice, it was found that the learners felt peer assessment effective.

**Keyword** peer assessment, educational evaluation, peer assessment support system, educational practice

### 1. はじめに

近年、学習者が同じ学習コミュニティに属する他の学習者の成果物を評価する相互評価が盛んになっている。教育評価における相互評価とは、「学級集団などにおいて、生徒同士に、互いに級友を評価させる方法」をいう[1]。植野[2]は、学習者同士で評価を行うことの利点として「学習者間の評価は、最も自然な評価体系であり、評価結果を受け入れやすくし、学習者の内省を引き出し、結果として知識の内化や深化を導くことや、学習動機の向上が期待される」と述べている。相互評価の教育的効果を高めるためには、学習者が評価結果を受け入れやすくするためのフィードバックが重要である。

学習者間の相互評価が注目されるようになった背景には近年の学力観の変化、e-ラーニングの普及による電子会議室での議論や協調学習の環境の整備がある。

e-ラーニングの普及によって、電子会議室での議論や協調学習が行われるようになった(例えば, Edelson et al. [3], 中原ほか[4], 緒方ほか[5])。このような学習場面では、狭義の知識獲得にとどまらず、学習者が学習活動を通して実践的能力を獲得することを想定しており従来のような形式のテストではなく、課題などの提出物を課して、学習過程に評価を埋め込むことが行われている。

学習者間の評価を支援するシステムやそれらを利用した実践が試みられている。これまで、ピア・アセスメント支援システムを使用する場合、レポートも電子媒体であることが多かったが、実際の現場で課せられるレポートは紙媒体で提出されることが多い。本稿では、紙媒体で提出されたレポートに対して、ピア・アセスメント支援システムを利用して行った相互評価の実践について報告する。ピア・アセスメント支援シ



システムを使用することにより、評価結果を即時に参照することができる。今回の実践では、最終成果物の途中段階のレポートを相互評価することにより、レポートの改善に資することを目的とする。そのため、学習者が評価結果を受け入れやすくするためのピア・アセスメント支援システムの利用法について述べる。

## 1. 関連研究

相互評価に関する取り組みとしては、プレゼンテーションスキルを題材とした実践(例えば、大倉ほか[6]、山口ほか[7])、外国語教育でコミュニケーションを題材とした実践(例えば、大倉・村尾[8])、情報教育を題材とした実践(例えば、柴田・小川[9])、プロジェクト学習での実践(例えば、松本[10]、Sluijsmans et al. [11]、Akahori & Kim[12])が行われており、それぞれ教育効果をあげたと報告されている。多くの分野で相互評価が成功している事例があるが、それらの実践で、相互評価のどのような要素が学習効果につながっているかは、ほとんど言及されていない。

コンピュータを用いた様々な相互評価ツールが開発、実践され、その教育的効果が報告されている。柴田・小川[9]は、情報関連科目の作品制作で、Webベースの相互評価システムを用いてお互いに評価させた後に、その結果にもとづいて評価レポートを作成させた。その結果、他者のコメントに基づいて多くの学生が作品の改良を行うことができたとして述べている。布施ら[13]は、ビデオオンデマンドとウェブデータベースを利用した相互・自己評価システムを開発し、プレゼンテーションの授業で、学習者の意欲が向上したと述べている。天野・下村[14]は、共同での作品制作の中間段階で、Webベースの相互評価システム(下村ほか[15])を用いた実践を行い、シンプルな評価観点を学習コミュニティで共有することで、作品制作途中での意見の交流が活発になったと述べている。藤原ら[16]は、プレゼンテーションソフトウェアの利用法の授業で相互評価の実践を行った。その結果、学習者がお互いに評価しあう場合は、そうでない場合に比べて、高い評価値をつけることが分かり、この現象を「お互い様効果」と名づけた。また、評価する相手に評価されない場合の方が、より適正な評価を行うこともわかった。評価者を選択する必要がある場合、お互い様効果を考慮した評価者の割り付けについて考慮する必要がある。

相互評価にコンピュータを利用する場合と、そうでない場合を比較した研究が行われており、対面で相互評価を行った場合の方が評価後の学習者同士のコミュニケーションを促進するといわれている(Liu & Sadler, [17], Figl et al.[18])。コンピュータを用いた協調学習支援システムに、単に課題提出機能や相互評価等を付

加するだけでは、対面で行う相互評価と同様の教育的効果が得られないことも考えられる。コンピュータを使って相互評価を行う際に、どのような支援を行うかについて考える必要がある。

## 2. 実践内容

### 2.1. 授業科目の概要

A大学人間科学部で行っている「実験調査研究法」で、相互評価を取り入れた授業実践を行った。この授業の目標は、「教育測定・評価の代表的手段であるテストについて、項目の作成とテストの設計・作成、実施・データ収集、分析・考察を体験し、教育データ解析の手法と意味を把握すること」である。この授業の受講生は30名である。

授業内容を表1に示す。授業中、受講者はデータ収集のため実地調査を行う。また、データ分析や相互評価はコンピュータ端末室で行った。受講者は、5人で1グループの計6グループに分けられ、グループで協力して課題を進める。授業で収集・分析するデータはグループ内で同一のものとなるが、受講者は提出レポートをそれぞれ独自に作成する。本稿では、2007年12月17日に行った相互評価について述べる。なお、受講生は、この相互評価で指摘された箇所を修正して、2008年1月25日まで修正後のレポートを提出した。

表1 授業内容

講義：テストの作成法，テスト理論，S-P表
テスト作成，作成したテストの修正
データ収集計画立案，データ収集の実施
テスト得点データ分析実施
レポート提出・相互評価(相互評価後、指摘された部分について修正し再提出)
発表会
レポート再提出(最終版レポート)

### 2.2. レポート課題及び評価項目

課題ではテストの作成を行う。作成の前に、「多くの人が関心を持ち、被験者も受験する気になるようなテスト」を目指して、グループごとに話し合っ、テストの実施目的を決定させた。作成するテスト項目は15項目以上、被験者は20人以上を目安とした。得られたデータから、信頼性係数の推定、S-P表分析を行い、レポートは個別に作成させた。

相互評価では、まず、初回の授業で配布したレポートの要件から作成した9項目の評価項目を提示し、適切、不適切の2段階で評価させた。9つの項目は、テストの設計について2項目、信頼性係数について1項目、S-P表に関して5項目に従い、レポートの体裁に

関して1項目から構成される。不適切を選んだ場合は、その理由の記入を義務付けた。次に、全体の印象を5段階で評価させ、その評価をつけた理由を記入させた。最後に、改善のためのアイデア、そのレポートを見た感想や疑問を自由記述させた。

### 1.1. 相互評価の方法

今回の授業では、紙でレポートを提出させた。今回、相互評価の実施に使用したピア・アセスメント支援システムは、電子ファイルでのレポート提出に対応している。しかし、今回の課題の場合は画面上では、画面の解像が低い読みにくいこと、課題では一部手書きでの作成を認めていることなどから、紙で提出され

たレポートを、そのまま評価に使用した。提出されたレポートの表紙をはずし、レポートの1ページ目上部にIDを記入することで、レポートを識別し、評価の際に誰のレポートであるか分からないようにした。

レポートの素材となるデータは、グループごとに作成したため、他のグループの受講生のレポートを評価させることとした。受講生は、所属していないグループから、それぞれ1人分のレポートの評価を行った。それぞれの受講生は、表2に示す受講生が作成されたレポートの評価を行う。表中の学習者IDは、今回の評価のために付された番号で、この番号は前述の提出されたレポートの1ページ目に付した番号を表している。ピア・アセスメント支援システムでは、このID

表2 評価の割当て  
評価対象

班名	A					B					C					D					E					F				
	学習者ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
A	1					1					2					3					4					5				
	2					2					3					4					5					1				
	3					3					4					5					1					2				
	4					4					5					1					2					3				
	5					5					1					2					3					4				
B	6		1								2					3					4					5				
	7		2								3					4					5					1				
	8		3								4					5					1					2				
	9		4								5					1					2					3				
	10		5								1					2					3					4				
C	11			1				2								3					4					5				
	12			2				3								4					5					1				
	13			3				4								5					1					2				
	14			4				5								1					2					3				
	15			5				1								2					3					4				
D	16			1				2				3									4					5				
	17			2				3				4									5					1				
	18			3				4				5									1					2				
	19			4				5				1									2					3				
	20			5				1				2									3					4				
E	21				1				2				3							4						5				
	22				2				3				4							5						1				
	23				3				4				5							1						2				
	24				4				5				1							2						3				
	25				5				1				2							3						4				
F	26	1					2					3					4							5						
	27	2					3					4					5							1						
	28	3					4					5					1							2						
	29	4					5					1					2							3						
	30	5					1					2					3							4						

※表中の数字は、それぞれの学習者が行う評価の順序

を用いて、次に評価すべきレポートを受講生に指示している。表中の数字は、それぞれの学習者が行う評価の順序を表しており、数字の入っていない受講生のレポートの評価は行わない。このように割り当てることにより、すべてのグループで、グループ外の5種類のレポートが評価されることになり、相互評価の時間中に他のグループとレポートのやり取りをすることなく、グループの中で決まった順序で評価済みのレポートを次の人に渡せばよいことになる。

次に、今回使用したピア・アセスメント支援システムについて述べる。今回使用したシステムは、お互い様効果を除去する機能をもったピア・アセスメント支援システム[16]を拡張したものである。システムを利用した相互評価の画面例を図1に示す。受講生は、画面上に指示された番号のレポートに対して、前述した項目について評価を行う。評価対象のレポート群は、予めグループごとに配布されており、グループに配布されたレポートを指定された順番に従って評価を行う。

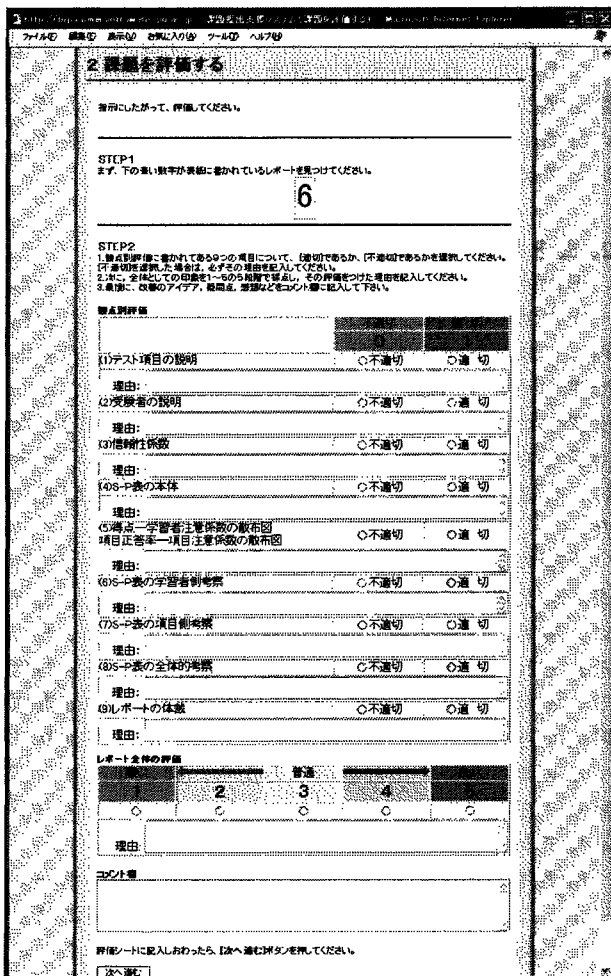


図1 ピア・アセスメント支援システムの画面例

相相互評価の結果は、他の受講生のレポートの評価を終わったあとに参照することができ、各評価者の入力した結果が一覧表にて表示される。なお、相互評価の実施時間には、相互評価に先立って、相互評価の意義、評価項目、システムの使用方法について、講義を行った。

## 1. 実践の結果

### 1.1. 相互評価の結果

本科目の受講生30名、全員がレポートを期日までに提出した。相互評価を行った日の授業には、29人が出席し、欠席者の分のレポートも評価の対象とした。1人当たり5人分のレポートを評価することを予定していたが、1人分のレポートが10枚を超えるなど、評価に時間がかかったため、表2中の1~4に相当するレポートを評価対象とし、1人あたり4人分のレポートの評価を行うこと変更した。

29人の出席者全員が4人分のレポートの評価を終えた。9つの評価項目別の評価結果を集計したものを表3に示す。不適切と評価された項目が、全体の39.4%あり、その理由とともにレポート作成者にフィードバックされた。レポート全体の評価(5段階で、5=良い~1=悪い)の平均は3.077、標準偏差は1.100であった。標準的なレポートを3と評価していたこと、適度に差をつけて評価をしていたこと、評価の理由が適切であると思われることから、適正な評価が行われたことが推測される。自由記述式のコメント欄には、116件中108件の記述が見られ、そのほとんどはレポートの改善の参考となる建設的なものであった。

表3 評価の結果

評価項目	適切	不適切	不適切の割合
テスト項目の説明	77	39	0.336
受験者の説明	68	48	0.414
信頼性係数	65	51	0.440
S-P表本体	97	19	0.164
正解数と注意係数の散布図	104	12	0.103
S-P表学習者側考察	34	82	0.707
S-P表項目側考察	58	58	0.500
S-P表全体的考察	58	58	0.500
レポートの体裁	72	44	0.379
合計	633	411	0.394

### 1.2. 事後アンケートの結果

今回使用したシステムの使いやすさについてアンケートを行い、5段階で回答させた。結果を図2に示す。「システムは使いにくかった」という問いに対し、

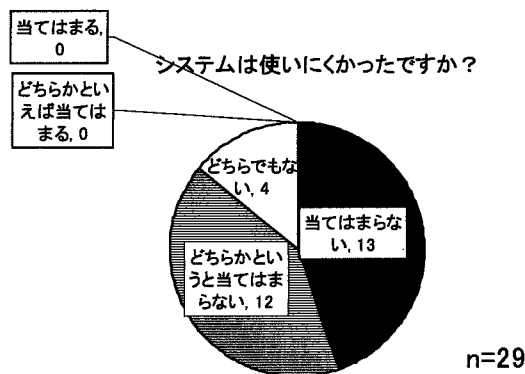


図2 システムの使いやすさ (アンケートの結果)

「当てはまる」、または、「どちらかといえば当てはまる」と回答した者はいなかった。このことから、システムの使いやすさに問題はなかったと考えられる。

相互評価の印象に関する8項目を5段階で回答させた (n=29)。評価値をそれぞれ「1: 当てはまらない」、「2: どちらかという当てはまらない」、「3: どちらでもない」、「4: どちらかといえば当てはまる」、「5: 当てはまる」として集計した結果を表4に示す。相互評価の意義について肯定的な結果が得られた。また、他者に評価された結果に対しても肯定的に捉えられていた。一方、他者のレポートを評価することは難しいという結果が得られた。自由記述項目への回答から、評価する時間をもう少し長く欲しいという学習者も存在した。そこで、今後は評価時間について検討する必要があるといえる。

表4 相互評価の印象 (アンケートの結果)

質問項目	平均値	標準偏差
自分のレポートにつけられた評価結果を確認することは有意義である	4.66	0.60
他人のレポートを評価することは、自分のレポートの改善のために有意義である	4.59	0.72
自分に対して付けられた評価結果は、参考になった	4.48	0.72
学生間での評価を導入した授業を、また受けてみたい	4.48	0.81
他の人が、自分に対して行った評価結果は、おおむね納得できた	4.10	0.92
他人のレポートを評価することは	3.86	1.01

難しかった		
評価項目 (観点別評価, 総合評価, コメント) の設定は適切だった	3.76	0.90
他人のレポートを適切に評価することが出来なかったと思う	2.72	0.87

### 1.1. 評価人数

今回の授業では、受講生1人につき5人分のレポートを評価する予定であったが、実践中に4人分の評価でも同等の効果があると判断し、時間の関係上、評価人数を変更した。実践の終了後にアンケートで、何人分のレポートを評価してみたいか、尋ねたところ、その平均は5.21人分であり、4人以下と答えた人数は、29人中5人であった。また、最大で何人分のレポートを評価できるか尋ねたところ、8.38人であったことから、意欲的に他の受講生の評価に取り組んだことが推測される。

同様に、何人の受講生の評価を受けたいか尋ねたところ、その平均は7.90人であり、10人以上と答えた人数は、29人中15人であった。また、少なくとも何人の評価を受けたいか尋ねたところ、その平均は4.79人であった。自分が受けた評価のうち、参考になった数を尋ねたところ、その平均は3.13人分であり、多くの評価が役に立ったことから、多くの評価を受けたいと考えていると推測される。

### 2. まとめ

形成的評価を目的として、紙媒体で提出されたレポートに対して、ピア・アセスメント支援システムを利用して相互評価の実践を行った。実践の結果、多くの受講生が積極的に評価に参加したことが確認できた。受講生は、レポートの改善のために、多くの参考となる意見を得ることができ、多くの受講生がアンケートに、今後も学生間の相互評価を受けて見たいと回答した。

今回の実践では、授業の欠席者が1名であったが、多数の欠席者が出た場合は、今回の実践で行ったように、事前に評価の割り当て (表2) を決めておくと、それぞれの受講生が受ける評価の人数にばらつきが出ることが考える。また、レポート提出時間と評価実施時間が近い場合は、未提出者のことを考慮して割り当てる必要がある。これらのことから、課題提出者、出席者に応じて割り当てを自動的に行うアルゴリズムが必要である。

紙媒体で評価を行ったため、評価者や対象レポートによって、1つのレポートの評価にかかる時間にばらつきがあり、それが他の受講生に影響を与えることがあった。今回の実践では、グループの全員が、手元に

あるレポートを評価し終えてから、次のレポートの評価を一斉に開始した。そのため、他の受講生の評価が終るのを待っている時間が長い場面や、逆に、もっと時間をかけて評価をしたいという要望もあった。また、前に評価したレポートの評価をやりなおしたいとの希望もあったが、既に他の受講生が評価を行っていることから、進行上の理由で、評価し終わったレポートの再評価はできないこととした。これらの問題については、例えば、提出させるレポートを2部にするなど、運用面の工夫で改善の余地があるため、その運用方法について検討することが今後の課題である。

## 文 献

- [1] 東洋, 梅本堯夫, 芝祐順, 梶田叡一, 現代教育評価事典, 金子書房, 1998.
- [2] 植野真臣, 先端的 e-learning の理論と実践, 教育心理学年報, no.44, pp.126-137, 2005.
- [3] Edelson D. C., O'Neill D. K., Gomez L. M., & D'Amico, L, A design for effective support of inquiry and collaboration, Proceedings of CSCL'95 , pp.107-111, 1995.
- [4] 中原淳, 西森年寿, 杉本圭優, 浦島憲明, 永岡慶三, 議論を通じた共同的な問題解決を支援する CSCL 環境の開発, 日本教育工学雑誌, vol. 24, no.Suppl., pp.97-102, 2000.
- [5] 緒方宏明, 松浦健二, 矢野米雄, WWW を利用した開放型グループ学習支援における Knowledge Awareness Map, 教育システム情報学会誌, vol.17, no.3, pp. 263-274, 2000.
- [6] 大倉孝昭, 高村博正, 奥田アレックス・H, 英語による効果的なプレゼンテーションと相互評価--連続性・発展性の視点から, 教育福祉研究, no.30, pp.20-30, 2004.
- [7] 山口顕司, 大塚茂, 森田慎一, 松本至, 矢壁正樹, 早水庸隆, 大塚宏一, チームでプロジェクト活動を行う科目での教育評価-学生の相互評価と教員の評価観点-, 工学教育, vol.55, no. 1, pp.41-46, 2007.
- [8] 大倉孝昭, 村尾敏彦, 映画の音声と解答を同期させ相互評価により聴解力を高めるCALLシステムの開発と学習者評価, 日本教育工学会論文誌, vol.27, no. 3, pp.315-324, 2003.
- [9] 柴田好章, 小川亮, 相互評価システムの開発と大学情報科目における利用, 日本教育工学会論文誌, vol.25, no.Suppl., pp. 33-38, 2001.
- [10] 松本重男, チームでプロジェクト活動を行う科目での教育評価: 学生の相互評価と教員の評価観点, 日本教育工学会論文誌, vol.24 ,no.1, pp.93-98, 2000.
- [11] Sluijsmans, D. M. A., Moerkerke, G., van Merriënboer, J. J. G., & Dochy, F. J. R. C., Peer Assessment in Problem Based Learning, Studies in Educational Evaluation, 27, pp.153-173, 2001.

平成18年度～平成20年度 文部科学省科学研究費補助金 研究報告書  
萌芽研究

「グループ学力の評価法開発」

発行 早稲田大学 人間科学学術院  
〒359-1192 埼玉県所沢市三ヶ島 2-579-15

Tel:04-2947-6848(代表)

平成 21 年 3 月

印刷 大学生協 u-poc