2006－CG－122（7）
2006／2／20

択一式試験問題の分析のための解答分布可視化

斎藤 隆文*，宮村（中村）浩子*，横内 文香**

* 東京農工大学 大学院生物システム応用科学教育部
** 東京農工大学 工学部 情報コミュニケーション工学科

我々は，択一式試験問題の結果を，問題の分析や改善に役立たせることを目指して，解答状況の可視化手法の開発を進めている．今回は，個々の設問について，正答および各誤答の選択状況を，総得点との相関を含めて可視化する手法を提案する．従来の設問解答分布図に誤答の選択肢を加え，移動平均で得られた曲線で色を塗り分ける．それによって，各設問の難易度，識別力，誤答の選択比率などといった基本的な特性が直観的に示されるとともに，成績階層による思考過程の違いを推測することが可能となった．

Visualization of Answer Distributions for Selection-Type Test Analysis

Takafumi Saito*, Hiroko Nakamura Miyamura*, Fumika Yokouchi**

* Graduate School of Bio-Applications and Systems Engineering,
Tokyo University of Agriculture and Technology
** Department of Computer, Information and Communication Sciences,　Faculty of Engineering,
Tokyo University of Agriculture and Technology

We are going to develop the visualization techniques for answer distributions of selection-type tests.　In this paper, we propose a new visualization method, which shows the relationship between the ratio of each selection and the total score.　With this method, basic features of each question, such as the degree of difficulty, power of discernment, and ratio of incorrect selections, are presented intuitively.　Moreover, it is possible to guess the difference of thinking between each score level.

１．はじめに

試験を効果的に行うためには, 試験の目的や受験者の能力に合った問題を作成することが必要である．より適切な問題を作成する上で，事前の検討とともに, 結果を分析して次回の試験に反映させることが重要となる. 択一式試験問題の場合, 分析方法として，古典的テスト理論，項目反応理論など，さまざまな統計的手法が研究され，実際に用いられている．これらの手法によって，各設問の難易度や受験者の識別性能など, 種々の数値指標が得られる．しかしながら，これらの数値指標は, 問題の適否を統計的に判断するためには有効であるが, 問題の中身の良否を示すものではない．そのため，作問者が多数の指標を見ても，各設問に対する受験者の思考過程を探るための手がかりは掴みにくく, 問題の改善に直接活かすことは難しい.

本研究では，問題に対する受験者の反応を，結果だけでなく思考過程も含めて, 作問者が直観的に知り, 問題の分析や改善に役立たせられるような，解答状況の可視化手法の開発を目的とする．今回は，個々の設問について，正答および各誤答の選択状況を, 総得点との相関を含めて可視化する手法を提案する.

以下，2 節に択一式試験の分析に使われるテスト理論について述べる．3 節では提案する可視化手法の考え方と計算方法を紹介する. 実際の試験結果に適用した例を 4 節に示し，5 節で考察を行う．

２．択一式試験の分析のためのテスト理論

択一式試験の結果を分析するために, さまざま統計的手法が研究され, また実際に使われている. 本節では，古典的テスト理論（2.1, 2.2 項）および現代テスト理論（2.3 項）から，代表的な解析手法を紹介する［1, 2]．また，本研究の目的に対

して, それらの手法だけではなぜ不十分であるかを述べる.

2.1 平均と標準偏差

試験の成績処理においては, 平均や標準偏差が古くから使われている. 点数の平均は, 母集団(受験者全体) に対する問題の難易度を示す. 各受験者の点数と平均点との差の 2 乗和が分散であり, 分散の平方根が標準偏差となる. これらの指標は, 母集団における点数のばらつきを示す.

平均と標準偏差とがわかれば, 難易度や得点のばらつきの大小による影響を排除した, いわば標準得点が得られ, それによって, 母集団の中での成績位置を示すことができる. 標準得点の一つとしてしばしば用いられるものに, 偏差値がある. 偏差値は,

50 ＋ (素点－平均点) × 10 ／標準偏差

によって求める. 偏差値 50 がちょうど平均であり, 偏差値 60 は平均よりも標準偏差分だけ点数が高いことを意味する.

2.2 識別力の分析

個々の設問には, 能力が上位の受験者と下位の受験者とで正解率が大きく異なるものと, あまり変わらないものとがある. 一般に前者の方が, 能力の違いが点数に反映されやすい, 言い換えれば, 試験結果から能力の違いを識別しやすいため, 良い設問であると考えられる. このような識別のしやすさを示すために, 弁別指数とよばれる指標が用いられる. 弁別指数は通常, 以下の方法で求められる. まず, 全受験者を総得点によって, 上位 27%, 中位 46%, 下位 27%の 3 グループに分ける. このとき, 各設問の弁別指数は, 上位グループの正答率から下位グループの正答率を引いたものとなる.

能力の違いと正答率との関係をより詳しく示す方法として, 設問解答分析図が使われることがある. これは, 横軸に総得点, 縦軸に正答率をとり, その関係をグラフで示すものである. たとえば, 全受験者を総得点によって上位から 20%ずつの 5 つのグループに分け, それぞれのグループの正答率を折れ線グラフで示せばよい. 多くの場合, このグラフは右上がりとなるが, その位置が上か下かによって, 問題が易しいか難しいかがわかる. また, グラフの傾きが大きいほど, 識別力が高いことがわかる. 図 1 にいくつかの例を示す. (a)は良い設問のグラフの一例である. (b)は易しすぎ, (c)は難しすぎる設問である. (d)は総得点にかかわらず正答率の変化がほとんどなく, 識別力が低い. (e)は上位層の識別に優れた設問, (f)は下位層の識別に優れた設問であるといえる. (g)や(h)のように右下がりの部分がある場合, 何らかの不適切な要因による可能性がある.

2.3 項目反応理論

古典的テスト理論では, 前述した種々の数値指標が母集団である受験者の集合に依存し, 絶対的な評価が得られない, という問題点がある. これを解決するために, 項目反応理論が用いられている. 項目反応理論では, まず測定したい受験者の能力を, 絶対的な指標 θ を用いて表す. 各設問(項目) の特性は, 能力 θ と正答率との関係として, 関数で表す. 各設問に対して, 試験結果からこの関数のパラメータを推定することにより, 設問の絶対的な特性(難易度や識別力など)が得られる. それによって, 母集団にかかわらず, 各受験者の絶対的な能力値 θ を推定することができる.

2.4 解決すべき課題

前述した各手法だけで十分な分析が行えるのは, いくつかの条件を満たす場合に限られる. しかし, 現実の試験においては, 以下のような理由から, 統計的な数値指標だけは, 満足な解析はできない.

(1) 総合力の測定への対処

弁別指数や項目反応理論を適用するには, その前提として, 測るべき受験者の能力が単一でなければならない. また, 弁別指数や設問解答分布図では, その測るべき能力の値を試験の総得点で代用できることを仮定している. しかし, 実際の試験では, 単一の能力ではなく, 多様な能力の組み合わせである総合力を測りたいことも多い. その場合, 好ましくない数値指標が出たとしても, 問題が不適切であるとは言い切れない. 問題の適否を判断するには, 受験者の思考過程, 特に誤答を導いた過程が, 作問者の意図に反したものでないかなど, より内容に踏み込んだ検討が必要となる.

(2) 大問形式への対応

項目反応理論では, 各設問が独立した小問形式での出題が前提となる. 小問形式での作問は, 英語の TOEFL や TOEIC に代表されるような, スキルを測る試験においては, 比較的容易である. しかし, 大学入試センター試験をはじめとする多くの試験では, 種々の場面での応用力や総合力を測ることが求められるため, 大問形式での出題が避けられないことがある. 大問形式では, ある設問の解答が次の設問の解答に強く影響することがありうるため, 有効な分析を行うには, 複数の設問の間の相関を調べる必要がある.

(3) 問題作成へのフィードバック

より良い作問を行うには, 単に問題の適否を判断するだけでなく, 不適切な解答分布となった理由を解析し, 次回の作問に活かすことが重要である. そのためには, 統計的な数値指標だけでなく, 受験者の思考過程にまで踏み込んだ検討ができるような，何らかの情報を得る必要がある.

３．解答分布の可視化

本稿では，受験者の思考過程を探るために，設問解答分布図を拡張した可視化手法を提案する. 本節では，可視化の考え方と，描画曲線の計算方法について述べる.

3.1 設問解答分析図を拡張した可視化

設問解答分布図では, 成績上位から下位までの各層ごとの，正答率の違いを示すことができる. しかし，どのように間違えたかはわからない. そこで, 各階層について, 正答, 誤答にかかわらず, すべての選択肢の解答比率を示すことを試みる. このとき，選択肢ごとに比率を曲線（折れ線）で示す方法もあるが, ここでは各選択肢の比率を縦軸方向に積み上げた, 選択肢を色によって区別する．このとき，正答を最も下に配置することで, 正答の曲線は従来の設問解答分布図に相当するものとなる.

3.2 解答分布曲線の計算法

より見やすい情報提示のために, 以下の 2 点を改善した可視化を行う.

(1) 解答分布を滑らかな曲線で描く

(2) 全体の面積比を解答比率に一致させる

いま, 受験者全体を成績が昇順になるように並べ，その順位を 0（最下位）から 1（最上位）の範囲に正規化して考える．ただし，受験人数は十分多いものとする.

このとき，2.2 項で説明した設問解答分析図の例は, 順位が [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1] の範囲について解答の選択率を計算し，それぞれ横軸の 0.1, 0.3, 0.5, 0.7, 0.9 の位置にプロットし，折れ線でつなぐことに相当する．受験

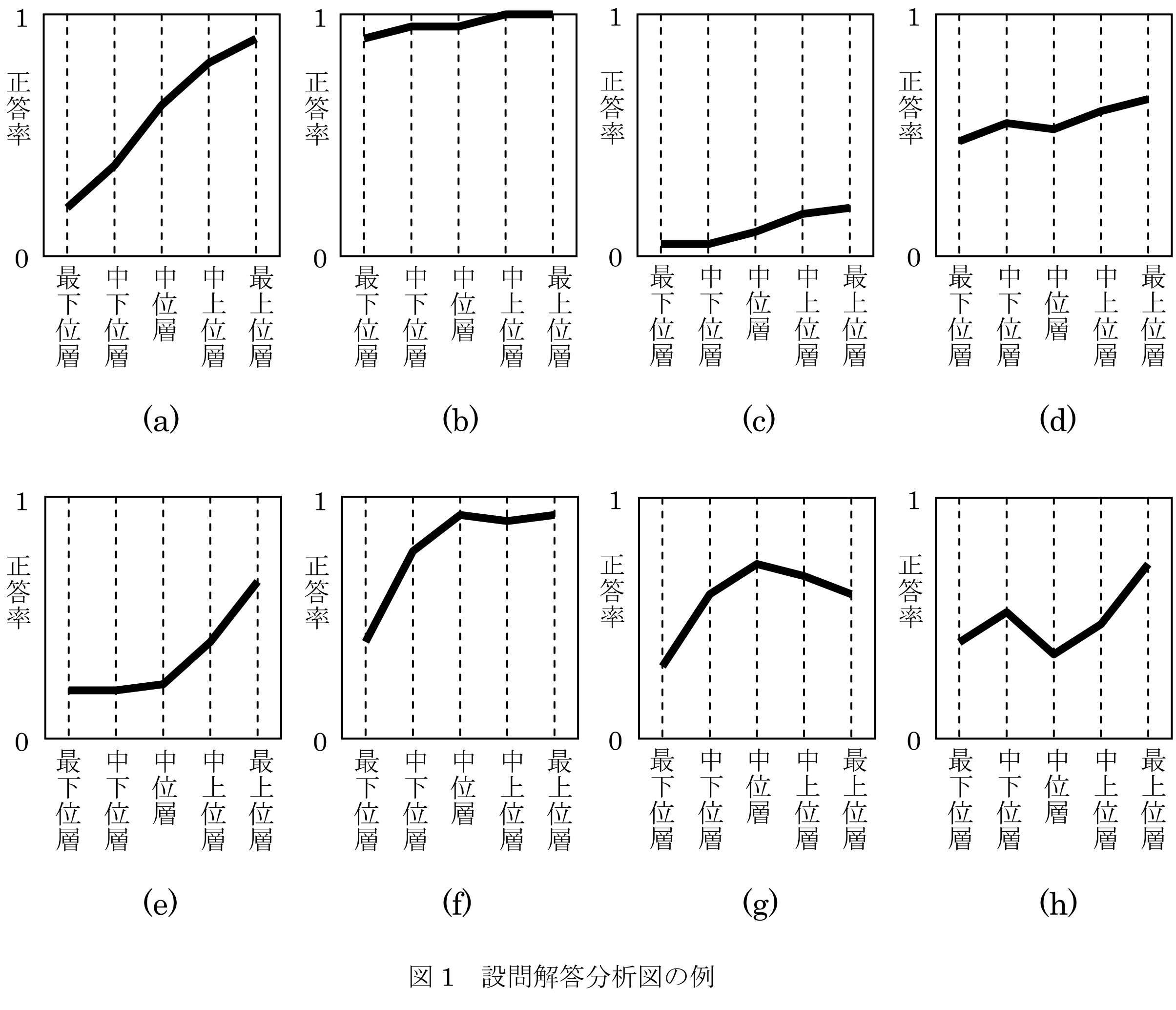

図１　設問解答分析図の例

者を 5 分割ではなく，もっと細かく分割すれば，折れ線から曲線に近づく．しかし，選択率を計算する際の対象人数が少なくなるため，ばらつきの影響が強くなり，変動の大きな曲線となってしまう．

そこで，移動平均を用いて描画する．たとえば，(受験人数×$2d$) について正答率の移動平均をとる場合，横軸 x における移動平均値は，順位が $[x-d, x+d]$ の範囲について計算すればよい．すなわち，

$$\overline{f}(x)=\frac{1}{2d}\left(\int_{x-d}^{x+d} f(x)dx\right) \quad (d \le x \le 1-d \quad \text{のとき})$$

ここで，$f(x)$ は順位 x の受験者の正答率を示し，$\overline{f}(x)$ は求める移動平均値を示す．ただし，$0 \le x < d$ あるいは $1-d < x \le 1$ の場合は，移動平均が計算できない．これらについては，以下のような重みつき平均を用いる．

$$\overline{f}(x)=\frac{1}{2d}\left(\int_{0}^{d-x} 2f(x)\,dx+\int_{d-x}^{x+d} f(x)dx\right) \quad (0 \le x < d \quad \text{のとき})$$

$$\overline{f}(x)=\frac{1}{2d}\left(\int_{x-d}^{2-(x+d)} f(x)\,dx+\int_{2-(x+d)}^{1} 2f(x)dx\right) \quad (1-d < x \le 1 \quad \text{のとき})$$

このとき，

$$\int_0^1 \overline{f}(x)\,dx=\int_0^1 f(x)\,dx$$

となり，曲線より下の部分の面積が，全受験者の正答率に一致する．誤答の各選択肢に対する平均値も同様の方法で計算し，下から積み上げて曲線を描き，着色すればよい．

４．適用例

本節では，提案手法を実際の試験結果に適用した例を示す．今回用いたデータは，かつて CG-ARTS 協会（画像情報教育振興協会）が行った，ある実験的な試験の結果である．この試験は，47 問の設問からなり，それぞれの設問は 3〜10 個の選択肢（A, B, C, ...）から一つを選ぶ形式である．受験者数は約 700 名である．

我々は，全 47 問の設問のすべてに対して，提案手法の適用実験を行った．図２は，それらの中から特徴的な結果が見られる 12 個の設問について，おおむね易しい順に並べたものである．移動平均の範囲は，$d=0.1$，すなわち，受験者全体の 20%の平均をとった．この場合，横軸の 0.1, 0.3, 0.5, 0.7, 0.9 に相当する位置（図２の白い縦線の位置）での正答率は，2.2 項で説明した 5 グループによる設問解答分布図のものと一致する．

まず，正答率の線（最下部領域の上端の白線）に着目すると，設問解答分布図と同様の見方により，たとえば以下のような性質がわかる．

1) 易しい問題： (1), (2)
2) 難しい問題： (12), (11)
3) 識別力が高い問題： (4), (8)
4) 識別力が低い問題： (1), (9)
5) 上位の識別に優れた問題： (7), (10)
6) 下位の識別に優れた問題： (3), (2)

また，誤答の各選択肢の分布を見ることにより，それぞれの選択肢がどのように機能したかがわかる．

7) 多くの選択肢が機能している： (11), (7)
8) 少数の選択肢のみ機能している： (3), (12)

正答や誤答の分布が特異なものについては，成績階層ごとの特徴を見るとともに，選択肢の中身を確認することにより，受験者の思考過程をある程度推測できることがある．以下にその例を示す．

[設問(5)]　正答：D
誤答の選択率は上位層ほど少なくなるが，E だけは上位層でもあまり減らない．選択肢の中身を照らし合わせたところ，E はある基本的な事項を見落としたことによる誤答であった．上位層については，その事項を知らないとは考え難いため，ケアレスミスによるものが多いと考えられる．

[設問(6)]　正答：C
誤答のうち，A は上位層でもあまり減らない．中身を確認したところ，A は C とかなり類似した用語であった．このことから，A と答えた受験者，特に上位層では，説明の内容は理解しているが，用語を確実に知らないために間違えたと考えられる．

[設問(10)]　正答：B
最上位層を除いて正答率が低い．最上位層については，誤答のうち A，D は減るが，C はほとんど減らない．中身を確認すると，A や D を解答する人は，ある基本事項について誤解をしているようである．最上位層とそれ以外とでは，この基本事項の理解度に差があると考えられる．

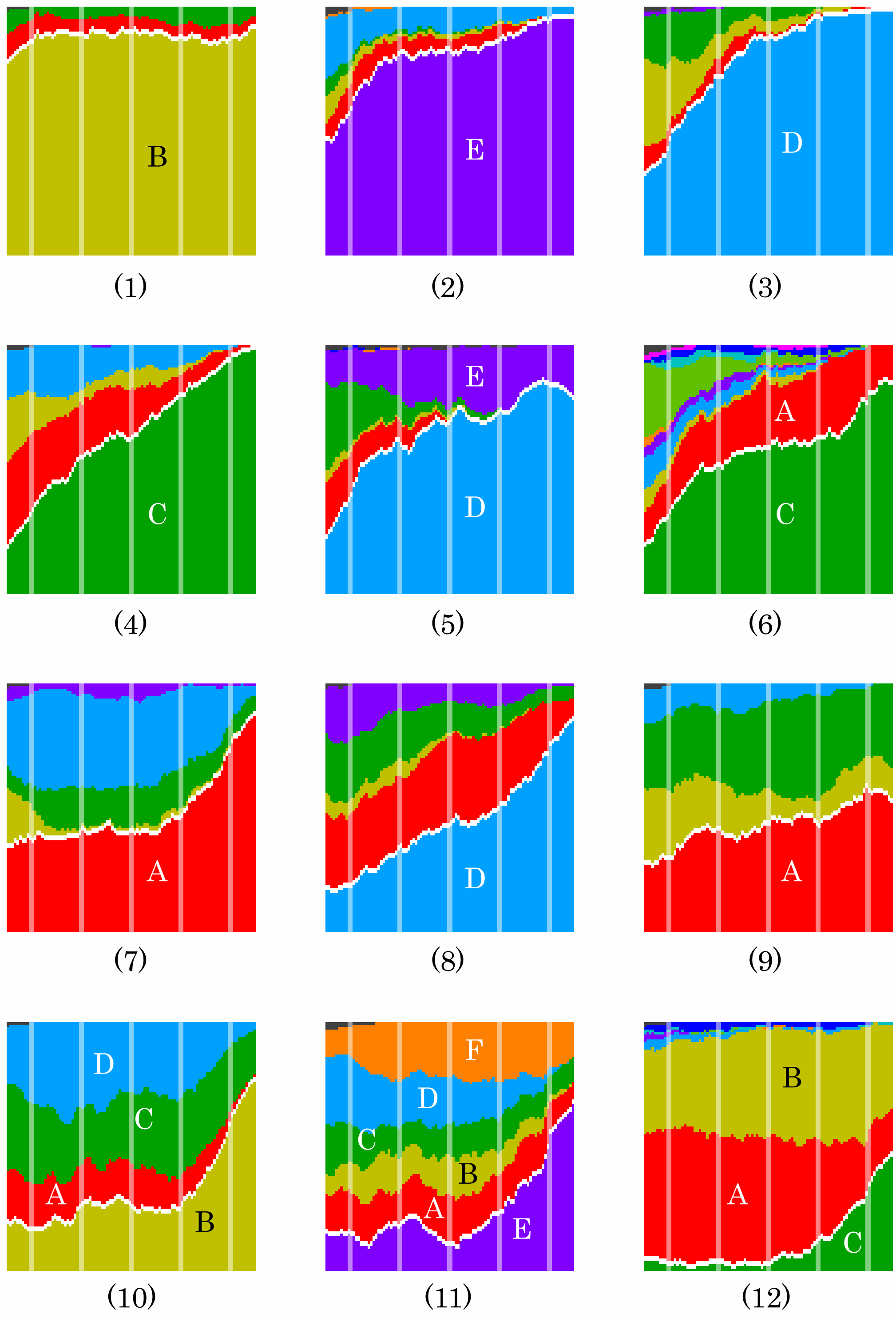

図 2　提案手法による可視化結果の例

[設問(11)]　正答：E

正答率が低い．誤答のうち，F は上位層でも減らない．この問題の中身に照らし合わせたところ，ある一つの事項が理解できれば，他の選択肢 A，B，C，D のいずれもが誤りとして却下できることがわかった．このことから，上位層だけはその事項を理解できる者が増えるが，E と F との区別は上位層でも難しいと考えられる．

[設問(12)]　正答：C

正答率が極めて低い．選択肢の中で，A，B，C の３個だけが機能しており，他の選択肢はほとんど選ばれていない．中位層以下では，A，B が半数ずつで，C はほとんどない．上位になるにつれて，C が増え A は減るが，B の比率は成績によらずほとんど不変である．すなわち，A，B，C のうち，中上位層以下のほとんどの者が C を棄却し，A，B から選択している．一方．最上位層では A を棄却し，B，C から選ぶ傾向にある．選択肢の内容を確認したところ，C の選択肢は一見不自然な点があるために棄却されたと考えられる．それが不自然でないことを確信できるのは，かなり能力の高い人に限られるようである．A の選択肢は，ある点に気がつけば，明らかに誤りとわかるはずである．最上位層では，その点に気づいて A を棄却した人が多かったと考えられる．

５．考察

提案手法は，本質的には，従来の設問解答分布図に，誤答も含むすべての選択肢を付加したものと考えることができる．このように誤答を加えることにより，成績階層ごとの正解率の違いに加えて，思考過程の違いや，基本事項の理解度の違いを推測することが可能となった．これを一般化して考えると，全受験者を何らかの観点でグループに分け，それぞれの解答パターンに違いが出れば，そこから有益な情報が引き出せる可能性がある．今回は，総得点をもとにグループ分けしたが，他のグループ分けでも有効性があるかどうか，検証する必要がある．

可視化の工夫として，提案手法では，折れ線のかわりに移動平均による曲線で表示した．これは，単に見やすいというだけでなく，より多くの情報を提示できる．たとえば，成績上位層の方が正解率が低いという逆転現象が見られた場合，折れ線による表示では，データの単なるバラツキによるものなのか，本質的な逆転なのか，わかりづらい．移動平均による曲線では，任意の成績を中心とした移動平均がわかるので，そこからバラツキの程度が示される．そのため，本質的な逆転を見分けるための一つの手がかりとなりうる．また，全体の色の面積比が各解答の選択比率に一致するように両端部を処理しているため，誤解の恐れが少ない．

６．おわりに

択一式試験問題の結果を，問題の分析や改善に役立たせることを目指して，解答状況を可視化するための新しい手法を提案した．従来の設問解答分布図に誤答の選択肢を加え，移動平均で得られた曲線で色を塗り分けることで，正答および各誤答の選択状況と総得点との関連性を可視化した．それによって，各設問の難易度，識別力，誤答の選択比率などといった基本的な特性が直観的に示されるとともに，成績階層による思考過程の違いを推測することが可能となった．

今後の課題として，２問もしくはそれ以上の設問について，解答状況の相関を可視化することがあげられる．特に，既存の理論では解析が困難な，大問形式の問題への適用を目指したい．

謝辞

実験データを提供いただいた CG-ARTS 協会 宮井あゆみ氏に感謝する．

参考文献

[1] 豊田秀樹，項目反応理論［入門編］，朝倉書店，2002.

[2] 高橋正視，項目反応理論入門 －新しい絶対評価－，イデア出版局，2002.