多次元アルファ係数を用いた作文評価構造抽出²

平 直 樹¹

ABSTRACTING ESSAY EVALUATION STRUCTURE BY USING MULTIDIMENSIONAL ALPHA COEFFICIENT

Naoki TAIRA

When we abstract the structure of three-way essay evaluation data, the aggregated matrix tends to show more unidimensional feature than that of each rater if the equal weights are used. Multidimensional Alpha Coefficient (MAC) proposed by Yanai (1994) gives independent sets of scoring weights which vary dimensionality of the aggregated matrices. The present study tried to extract the evaluation structure from three-way essay data by using the method for preservation of the original multidimensionality. The essay data of Taira (1995) evaluated by seven raters were used. Some criteria, including the factor analysis with Procrustes rotation method, were set to choose the most suitable weights. The result showed that the weights for the sixth solution of MAC was regarded as the best. The aggregated matrix yielded three factors, consisting of one more dimension compared with the former analysis. The revised path diagram showed much clearer causal relationships. The emotional factors on writing and reading had an effect on the ability of story making, whereas Writing Habits influenced only upon the Loyalty to the Task Condition.

Key words : essay evaluation, structure, Multidimensional Alpha Coefficient (MAC), Multidimensional Intraclass Correlation Coefficient (MICC), Procrustes rotation method.

問 題

現代の教育において、作文評価は重要な研究課題の1つである。論述形式課題の評価の技術的問題は古くから認識され、今なお研究が続けられている。中でも重要なのは、採点者に関わる諸問題である。テスト理論の文脈では、複数の採点者が独立に評価を行った結果を総合することが望ましいとされる。採点の妥当

性・信頼性を確保することが目的である。その際、採点者は同じ属性を測定するための装置と見なされ、採点結果の不一致は測定誤差として扱われる。現在まで、そのような観点による評価の妥当性・信頼性に関する理論的、実証的研究は日本においても外国においても数多く見られる(例えば、平井・渡部, 1994; 渡部・曹, 1992; De Gruijter, 1984; Coffman & Kurfman, 1968)。

一方、作文課題の中で測定される属性は、1次元的なものには止まらないことも認識されており、作文の評価における採点者の評価構造に着目した研究も存在する(例えば、Remondino, 1959; Ackerman & Smith, 1988)。しかし、個々の採点者の評価をどのようにまとめて多次元的に記述するか、という問題について特に焦点を当てて検討した研究は見当たらないように思われる。

¹ 大学入試センター (The National Center for University Entrance Examinations)

² 本研究の内容は、1995年8月の行動計量学会第23回大会において発表したものの一部をまとめ直したものであり、1996年6月のAnnual Meeting of the Psychometric SocietyでTaira & Yanaiとして発表したものと同主旨のものである。

本研究は、柳井 (1994), Yanai & Taira (1996, 1997) の「多次元級内相関係数」, 及び, 「多次元アルファ係数」を応用し, 複数の採点者の評価構造を集約するための一方法を提示する。特に, 作文評価に際して個々の採点者に見られた多次元的な評価構造をできるだけ確保しつつ, 複数の採点者による評価を1つの構造に縮約することを目的とする。

1. 複数採点者による作文評価の構造

複数の採点者が全ての答案を評価する場合, 全ての採点者が与えた得点の単純和を基に答案に対する最終的な評価を決定するのが一般的である。例えば, 一般化可能性理論 (例えば, De Gruijter, 1980) は, 複数の採点者が平行測定装置であることを前提として信頼性を議論している。すなわち, 好みによって評価が著しく分かれてしまうような場合を除き, 信頼性の観点から言えば, 単純和による評価は有効な方略である。

ところが, 評価項目が複数の場合, 複数の採点者の採点結果から評価構造を捉えようとする場合には, 全ての採点者が与えた得点の単純和を評価項目ごとに行う方法が最良の方策とは言えない。平 (1995) では, 7名の採点者が153名の中学生・高校生の書いた作文を15の評価項目から採点した。具体的には, 「課題への忠実性³」が「描写力, 構成力, 設理解, 「基礎作文力⁴」が「文字, 基本的文法, 漢字, ことばの使用」, 「文章構成力」が「ストーリー構造化, ストーリー必然性, 完成度」, 「物語文のセンス」が「独創性, リズム・テンポ, 工夫, 内容, 文才」から評価された。また, 各項目に関して単純和を取った場合の相関行列を基に因子分析を行うと, 2因子を抽出して「基礎的言語能力」と「文章力」の尺度とすることが妥当とされた。

この結果が7名の採点者の評価の平均的構造を表わしていれば, 当初の想定と異なった結果であったとしても合理的と考えるべきだが, この場合は, そうとは言えない。TABLE 1 は7名の採点結果を固有値分解して得られた第1～第6固有値, 及び, 評価構造の1次元性の総合的な指標としての個人内アルファ係数の表である。評価構造の1次元性の強さに個人差がある一方で, 単純和を取った場合の相関行列の第1固有値が全採点者の平均値とかけ離れて大きく, 個々の採点結果のどれよりも1次元性が強くなっていることが分かる。

単純和を取ることで1次元性が強くなることは, 必

TABLE 1 採点者毎, および, 単純和を用いて得点を算出した場合の項目間相関行列の固有値 (第6固有値まで) とアルファ係数

採点者	固有値						アルファ係数
	1	2	3	4	5	6	
1	9.419	1.393	0.684	0.640	0.502	0.472	.953
2	8.868	1.273	0.843	0.684	0.634	0.524	.949
3	6.554	1.756	0.956	0.841	0.748	0.660	.893
4	7.237	1.751	1.111	1.060	0.784	0.551	.918
5	8.564	1.247	1.003	0.723	0.633	0.528	.942
6	6.390	1.944	1.621	0.962	0.756	0.695	.888
7	8.009	1.527	1.273	0.750	0.553	0.508	.934
平均	7.863	1.556	1.070	0.809	0.659	0.563	——
単純和	10.067	1.451	0.999	0.563	0.439	0.334	.960

然的な現象である。例えば, 全ての項目の信頼性係数が ρ で互いに平行である m 項目からなるテストの相関行列を考えた場合, 各項目間の相関係数は ρ であるので, その第1固有値は,

$$\lambda_1 = 1 + (m-1)\rho \quad (1)$$

となる。一方, 全ての項目に r 回繰り返し測定を行って和を取った場合, スピアマン＝ブラウンの公式 (池田, 1973) から, 信頼性係数は,

$$\rho_r = r\rho / (1 + (r-1)\rho) \quad (2)$$

となる。

$$\rho_r > \rho \quad (3)$$

であるから, それぞれ元の項目について r 回ずつの繰り返し測定のと和を取った m 項目からなるテストの相関行列の第1固有値は,

$$\lambda_{r1} = 1 + (m-1)\rho_r$$

となり,

$$\lambda_{r1} > \lambda_1 \quad (4)$$

である。この例から, 単純和を取ることで項目間相関が高くなり, 第1固有値が大きくなる様子が端的に見て取れる。 m 個の固有値が全て正であり, その和は m であるから, 第1固有値が大きくなると, 第2固有値以下が相対的に小さくなる。理論的には第2固有値以下は全て $1 - \rho_r$ である。多次元的な構造を持つ作文評価の場合, 評価構造に個人差がなければ, 有意意味な次元数に対応した数の固有値が応分に大きくなり, それ以外の誤差成分のみが小さくなるという想定も可能である。しかし, 実際には個人差はつきものであり, 上記の推論は非現実的である。

以上のことから, 多次元的構造を保ったまま, 複数

³ 平 (1995) では「出題要求への対応」

⁴ 平 (1995) では「作文のための基礎的能力」, 「基礎的言語能力」

の採点者の評価を縮約して構造を取り出そうと考えた場合、単純和に基づく分析をすることは、得策ではないことが分かる。

2. 多次元級内相関係数, 多次元アルファ係数

柳井 (1994), Yanai & Taira (1996, 1997) は, 級内相関係数, アルファ信頼性係数の概念を 3 相データ (three-way data) に拡張した。すなわち, 3 相データにおいても, 一般化固有値問題を解くことによって, 線型的に 2 相に集約して得られる行列に, 可能な最大の値から最小のものまでの級内相関係数, アルファ信頼性係数を持つような独立な重みベクトルが得られることを示した。例えば, FIGURE 1 のような $n \times r \times m$ のデータにおいては, 図の上側に示したような形で集約を考えた場合, 独立な $m \times 1$ の重みベクトルを持つ m 個の $n \times r$ 行列が得られることになる。柳井 (1994) においては, 得られた級内相関係数, アルファ信頼性係数はそれぞれ「多次元級内相関係数」「多次元アルファ係数」と呼ばれている。また, Yanai & Taira (1996, 1997) にしたがって, 本研究では, 以後それぞれを“MICC

:multidimensional intraclass correlation coefficients”, “MAC : multidimensional alpha coefficients” と略記する。

大きな方から i 番目の級内相関係数を r_i , アルファ信頼性係数を α_i と表わすと,

$$\alpha_i = mr_i / (1 + (m-1)r_i) \quad (5)$$

という通常の 2 相データで得られる関係は, MICC と MAC の場合にも保たれており, 対応する MICC と MAC においては, 重みベクトルも同一である (Yanai & Taira, 1996)。したがって, 重みベクトルや個々の重み係数を問題にする場合には, MICC と MAC とを区別する必要はない。

3. MICC, MAC を用いた作文評価の信頼性の検討と作文評価の構造の抽出

ところで, 柳井 (1994) の MICC, MAC による 3 相データの 2 相化の方法は, FIGURE 1 の上側のように「作文×採点者」という行列を得るためにも, 下側のように「作文×評価項目」という行列を作るためにも応用可能である。前者の場合, 個々の評価項目に掛け

FIGURE 1 作文評価データの 2 次元への圧縮の 2 つの方向

る重みを表わす独立な $m \times 1$ の重みベクトルが m 個得られ、後者の場合には個々の採点者に掛ける $r \times 1$ の重みが r 個得られることになる。両者は理論的には同等であるが、応用的には異なった意義がある。

まず、〔作文×採点者〕という行列を考えた場合には、採点の信頼性を考えるのが普通である。そこでは「最大のアルファ信頼性係数を与える重みが望ましい」という考え方が自然である。その際、大切なのは最大固有値に対応する第 I 解とその重みベクトルである。

しかし、〔作文×評価項目〕という行列においては、そのような一方向的な基準が唯一ではない。MICC, MAC は作文評価の構造を多次元的に捉える手段としても利用可能である。作文の評価構造を抽出する目的の場合、最大固有値に対応した最大の MICC, MAC を与える重みベクトルとは、評価項目の 1 次元性を最も強める解である。他方、それ以外の固有値に対応した MICC, MAC を与える重みは、評価項目の構造の多次元性の程度に応じた情報の縮約法と位置づけられる。すなわち、「データの構造を様々な形で集約する重みベクトルから、適度な多次元性を持つものを選択する」という考え方による作文評価構造の分析のアプローチが成立し得るのである。

本研究では、3 相の作文評価データに対する MICC, MAC の以上のような応用可能性のうち、後者に焦点を絞って報告する。すなわち、柳井 (1994) の多次元級内相関係数、多次元アルファ係数の方法を用いて、平 (1995) の作文評価データから多次元的な評価構造を抽出することを試みる。さらに、その結果を用いて、作文能力評価モデルの再検討を試みることにする。

方 法

本研究では、平 (1995) のデータに対し、以下のような手順で分析を行う。なお、FIGURE 2 は本研究における一連の手順をフローチャート風に示したものである。

1. 重みの選択

作文評価構造の分析を行うためには、まず、どの重みベクトルを用いて 3 相構造のデータを〔作文×評価項目〕に縮約すべきかを決定しなければならない。

本研究のデータにおいては、7 種類の MICC, MAC を与える、7 種類の独立な 7×1 の重みベクトルが得られることになる。作文評価構造の分析のためにどの解を用いるべきかについて、一義的な基準はない。したがって、本研究では以下の観点から総合的に判断して重みの選択を行う。

FIGURE 2 分析の手順

a. 重みの選択基準①—単純和との相関

第 I 解～第 VII 解の重みベクトル (TABLE 2 参照) を用いて、個々の評価観点の得点を計算する。その結果と単純和を用いて得られた得点との相関係数を解の選択基準の 1 つとする。すなわち、単純和に近い結果を与える解を選択することを考える。これは、単純和に基づく従来の分析方法との整合性を保つためである。ただし、多次元的構造の抽出という本研究の主旨に鑑みて、第 I 解は選択の対象から除外する。

b. 重みの選択基準②—因子分析のプロクラステス法

本研究では、それぞれの解に対応した重みで得られた評価構造がどの程度データを反映しているか、というチェックを行うことを考え、それを重み選択の基準の 1 つとする。評価構造のチェックのためには、プロクラステス法を用いた因子分析 (芝, 1979) を応用する。

プロクラステス法を用いる場合、最初に問題となるのは因子の回転の外部基準となるターゲットの定め方である。本研究では、前述のように、評価項目の作成

時に4因子的な仮説的構造を想定している。したがって、それを初期構造とするのは一案である。しかしながら、初期構造がデータの性質を充分反映している保証はない。したがって、さらに幾つかのプロセスを加え、ターゲット・マトリックスを決定していくこととする。

b-1. 因子数の予備決定

モデルに含むべき因子数は、初期構造とデータから得られる個々の採点者の評価次元数とを勘案して検討する。

個々の採点者の評価次元数(因子数)は、採点者ごとの〔作文×評価項目〕の行列の相関行列の固有値分解により、それぞれ、i) 1以上の固有値数、ii) 次の固有値との差の大きさ、iii) 累積寄与率の大きさ、といった基準を総合して決められる。これが全採点者で一致する保証はない。その結果と、初期の「4因子」という想定とを総合して、モデルに含むべき因子数を予備的に決定する。

その後、作文評価構造の分析に用いる重みベクトルから得られた固有値とを対比し、再度検討して最終的に決定する。

b-2. 項目フィットのチェック

初期構造をターゲットとする予備分析項目レベルで初期構造とのフィットが悪い項目を探すことが目的である。重みの選択を行うターゲットを定める際には、フィットが悪い項目を自由にパラメタにしてターゲット・マトリックスを修正する必要がある。

まず、MICC, MACの各解から得られた〔作文×評価項目〕の行列に対して、初期構造に合わせて4因子モデルの因子抽出を行い、初期構造をターゲットとした完全直交プロクラステス解を求める。次に、全ての解を総合して初期構造への項目のフィットを調べる。なお、プロクラステス法を用いた場合の因子比較の基準として、RMS指数、サリエント類似性指数等が提案されている(柳井他, 1990)が、結果が行列の形に表わされるためにコンパクトにならないという欠点がある。本研究では因子数はあらかじめ定めており、ターゲットとデータの因子の対応関係が明白であるので、より単純な方法で差し支えない。すなわち、項目jの残差を d_j とすると、

$$d_j = \sum_{k=1}^l \sum_{p=1}^m (t_{jkp} - a_{jkp})^2 \quad (6)$$

であり、 d_j の値が大きい項目のフィットが悪いと判断される。ただし、(6)式中の t_{jkp} は項目jの解k, 因子p

におけるターゲットの値、 a_{jkp} は因子負荷量、lは解の数、mは因子数である。ここでは、 $l=7$, $m=4$ となる。なお、 d_j の値の大小を絶対的に判断する基準は存在しないので、フィットの良し悪しは因子数を勘案しながら項目相互を比較して判断する。

また、tの値は因子負荷における共通性(解kの重みを用いて算出した項目jの場合、項目jを含む因子pに対しては $t_{jkp} = h_{jk}^2$, それ以外は0)を用いることとした。具体的には、第1因子に対しては、初期構造で「課題への忠実性」を測定とした3項目、第2因子には「基礎作文力」の4項目、第3因子には「文章構成力」の3項目、第4因子には「物語文のセンス」の5項目それぞれの共通性をターゲットの値とし、それ以外はすべて0とした。

b-3. ターゲット構造の修正—自由パラメタの設定

前節の分析結果から、ターゲットとなる因子パターンを新たに決定する。これを分析的評価項目の「修正ターゲット」と呼ぶ。まず、b-1の結果から修正ターゲットの因子数を定める。さらに、b-2の分析から自由パラメタとする項目を選択する。

b-4. 重みの選択

—修正ターゲットに対するフィット

b-2における分析と同様にMICC, MACの各解の重みベクトルを用いて得られた行列に対して、定められた因子数による因子抽出を行い、修正構造に基づいた不完全直交プロクラステス解を求める。そして、修正ターゲットの固定パラメタのフィットの良し悪しで、その重みがどれほどデータの構造を反映しているかという判定を行う。すなわち、第k解の重みベクトルで得られた行列の因子負荷量と修正ターゲットとの残差を e_k とすると、

$$e_k = \sum_{j=1}^r \sum_{p=1}^m \delta_j (t_{jkp} - a_{jkp})^2 \quad (7)$$

であり、 e_k の値が小さいほど、フィットがよいと判断される。ただし、(7)式中のrは項目数であり、ここでは、 $r=15$ となる。また、 δ_j は項目jの残差を式に算入するかどうかを決める係数であり、自由パラメタの場合には0、固定パラメタの場合には1となる。

なお、(6)式の場合と同様、 t_{jkp} には因子負荷の共通性を用いる。

c. 重みの選択と因子数の再検討

a, bの分析結果から、作文の評価構造を求めるための重みを決定する。その際、b-1で仮決定した因子数が、選択された解によって得られる行列の構造と経験的に一致するかどうかを確認する。そして、不一致

の場合には因子数を再検討し、b-3で作成した修正構造を再検討する。次に、新たな修正ターゲットに対して直交不完全プロクラステス法を行い、よりよい重みと最適な因子数が得られるまでこのプロセスを繰り返す。

2. 作文評価の構造の分析

選択された解の重みベクトルを用いた行列に対し、バリマックス法を用いた因子分析を行う。その結果をMICC, MAC を用いて得られた作文評価の構造とする。

3. 作文評価に影響する要因の分析

a. 尺度得点の計算

得られた作文評価の構造に基づき尺度を作成する。すなわち、項目ごとに算出された全採点者の単純和を用いて、新たに得られた構造にしたがった1次元尺度を作成し、尺度得点を計算する。以下、これらの尺度に基づいた分析を行う。

b. 作文評価モデルの再検討

平(1995)の作文評価モデルでは分析的評価観点の「文章力」尺度の部分が作文評価に関するパスダイアグラムに組み込まれていた。本研究では、それとの異同も考慮に入れ、上記の分析から得られた構造に置き換え、作文能力の評価に対する説明要因の再分析を行い、MICC, MAC の応用的意義について検討を加える。

結 果

1. MICC, MAC の値と7種類の重み係数

得られた7個のMICC, および, MAC とそれぞれの重みの値はTABLE 2に示すとおりである。それぞれの解について、採点者に掛かる重みベクトルは2乗和が1に規準化されている。

級内相関係数の値は.15～.65, アルファ信頼性係数

TABLE 2 MICC, MAC の値と採点者の重み

	第I解	第II解	第III解	第IV解	第V解	第VI解	第VII解
級内相関係数	0.65	0.44	0.41	0.39	0.23	0.21	0.15
アルファ係数	0.97	0.92	0.91	0.91	0.82	0.80	0.73
採点者	重み係数						
1	0.62	-0.64	0.36	-0.20	-0.20	-0.08	0.19
2	0.63	0.72	0.24	-0.34	-0.22	-0.40	0.26
3	0.13	-0.08	0.07	0.22	0.90	-0.45	0.38
4	0.17	-0.01	-0.08	-0.23	0.19	0.10	-0.70
5	0.35	-0.10	-0.84	0.12	-0.13	-0.06	0.29
6	-0.14	0.19	0.13	-0.14	0.08	0.78	0.30
7	0.19	0.13	0.29	0.84	-0.21	0.11	-0.30

は.73～.97の間に分布している。すなわち、評価観点にどのような重みを用いて各観点の合計を求めても、級内相関係数, アルファ信頼性係数はこれらの値の範囲を越えない。

全ての解において、採点者に掛かる重みに正負が入り交じっている。また、各解において、特定の採点者に対して特に大きな重みが掛かっており、そのパターンに解ごとの特徴が見られる。すなわち、第I解においては、採点者1と2に正の大きな重みが掛かり、この2名の採点結果をより重視した構造が得られたと考えられる。第II解では、採点者1に負、2に正の大きな重みが掛かり、残りの5つの解では、それぞれ1人ずつ正または負の大きな重みを持つ採点者が見られる。

2. 重みの選択

a. 因子数の仮決定

それぞれの採点者の評価次元数を固有値で判断したところ、2因子が適当と思われる者が3名、3因子が3名、4因子が1名であった(TABLE 1参照)。以上のことと、多次元的構造を取り出すという本研究の目的とを勘案し、3因子構造を前提として以後の分析を進める。

b. 単純和との相関

TABLE 3に各評価観点の単純和との相関係数のリストを示す。網かけの部分が、母相関係数0の帰無仮説に基づき検定を行った場合に5%水準で有意となる

TABLE 3 単純和(オリジナル)との相関係数

評価観点	第I解	第II解	第III解	第IV解	第V解	第VI解	第VII解
描写力	.936	-.157	.006	.061	.098	.375	.043
構成力	.938	-.056	-.048	.094	.085	.290	.051
設定理解	.941	.115	.027	.198	.101	.436	-.021
文字	.956	.163	-.117	.086	.569	.520	.193
基本的文法	.894	-.012	-.043	.021	.334	.393	-.146
漢字	.886	.038	-.050	.050	.247	.228	.151
ことばの使用	.899	.011	.213	.205	.163	.130	-.072
ストーリー構造化	.908	.131	-.080	.190	.222	.214	-.163
ストーリー必然性	.937	.065	-.120	.076	.228	.135	.018
完成度	.937	.178	-.271	.152	.332	.185	.026
独創性	.867	.023	-.108	.145	.145	.277	-.114
リズム・テンポ	.864	.203	.228	.033	.122	.264	-.247
工夫	.913	.125	.020	.047	.213	.254	-.031
内容	.911	.188	.010	.021	.049	.077	-.261
文才	.917	.333	-.235	.059	-.003	.225	-.171
正の相関の個数	15個	4個	2個	2個	8個	12個	1個
負の相関の個数	0個	0個	2個	0個	0個	0個	4個

正の相関係数であり、下線で示したのが負の相関係数である。第Ⅰ解が全ての項目でこの基準を満たし、値も大きい。が、「多次元的構造」を取り出すという本研究の目的には合致しない。次に単純和との相関関係がよいのは第Ⅵ解である。この基準からは第Ⅵ解が重みとして最適であり、次いで第Ⅴ解が適していると考えられる。

c. プロクラステス法による重みの選択

c-1. 項目フィットの分析と修正ターゲットの作成

TABLE 2 に示した7種類の重みから得た作文評価の行列に対し、初期構造をターゲットとする直交完全プロクラステス解を求め、(6)式による項目ごとの残差を算出した結果はTABLE 4 に示すとおりである。「文章構成力」の下位項目と想定されていた「ストーリー構造化」、「ストーリー必然性」、「完成度」のフィットが悪い。a 節の分析結果から、3 因子モデルで分析を進めていくので、これらの項目を全て自由パラメタの対象とする。また、その他の項目に関しては、1.03~1.56の値を示した「独創性」、「工夫」、「内容」、「基本的文法」、「描写力」、「漢字」の6項目(TABLE 4 中に網かけで示した部分) が比較的良好なフィットを示していると考えて固定パラメタとし、それ以外を自由パラメタとする。

c-2. 修正ターゲットに対するフィットを基準とする重みの選択

TABLE 4 初期構造をターゲットとした項目のフィット (残差)

評価観点	d
描写力	1.33
構成力	1.85
設定理解	1.96
文字	2.80
基本的文法	1.12
漢字	1.56
ことばの使用	2.27
ストーリー構造化	2.10
ストーリー必然性	2.60
完成度	2.78
独創性	1.03
リズム・テンポ	2.11
工夫	1.08
内容	1.12
文才	1.91

網かけ：固定項目

以上の過程で得られた修正ターゲットに対する残差を各解ごとに求めたところ、TABLE 5 が得られた。明らかに、第Ⅵ解が最適である。

以上の分析の結果、単純和との相関の基準でも、プロクラステス法を用いた因子分析の基準でも第Ⅵ解が重みとして最適であった。

TABLE 6 は第Ⅵ解の重みを用いて得られた項目間相関行列の固有値等の表である。予備的に決定した3 因子モデルをそのまま採用して問題はないと思われる。

2. 作文評価の構造の分析

第Ⅵ解の重み係数を用いて計算した作文評価結果の行列に対し、バリマックス法による因子分析を行った。因子の抽出は主因子法による。初期値にはSMCを用い、3 因子モデルで反復推定を行った。その後、バリマックス回転をして解を求めた。因子構造はTABLE 7 に示す。

明らかに、初期構造の「課題への忠実性」は第2 因子に、「基礎作文力」は第3 因子にそれぞれ対応している。また、「文章構成力」と「物語文のセンス」は第1 因子に対応している。本研究では、それを「物語作文力」と呼ぶ。なお、他の解の重みベクトルを用いた3 因子モデルの分析においては、TABLE 7 ほど明確な単純構造は得られなかった。

3. 作文評価に影響する要因の分析

前項までの分析で得られた作文評価の構造に基づき

TABLE 5 修正マトリックスに対する解のフィット (残差)

解	e
第Ⅰ解	1.02
第Ⅱ解	0.64
第Ⅲ解	1.73
第Ⅳ解	0.56
第Ⅴ解	1.55
第Ⅵ解	0.30
第Ⅶ解	1.05

網かけ：最小値

TABLE 6 第Ⅵ解の重みベクトルによる項目間相関行列の固有値 (第6 固有値まで)

固有値	5.14	1.89	1.61	1.09	1.00	0.81
次の固有値との差	3.25	0.28	0.53	0.09	0.19	0.16
寄与率	34.3%	12.6%	10.7%	7.3%	6.7%	5.4%
累積寄与率	34.3%	46.9%	57.7%	64.9%	71.6%	77.0%

TABLE 7 第VI解の重みを用いた場合の作文評価の構造 (3 因子モデル, パリマックス解)

評価観点	第1因子	第2因子	第3因子
描写力	.104	.837	.075
構成力	.237	.830	.185
設定理解	.213	.418	.293
文字	.058	.107	.330
基本的文法	.010	.087	.564
漢字	-.130	.055	.422
ことばの使用	.377	-.075	.570
ストーリー構造化	.711	.279	-.067
ストーリー必然性	.730	.332	-.043
完成度	.752	.364	-.115
独創性	.739	.077	-.032
リズム・テンポ	.584	.118	.300
工夫	.421	.145	.239
内容	.768	-.023	.018
文才	.683	.040	.242

尺度を作成した。すなわち、単純和の計算結果を用いて、「基礎作文力」、「課題への忠実性」、「物語作文力」に分類された項目ごとにそれぞれ1次元尺度を作成し、尺度得点を計算した。

次に、以上の3つの尺度をモデルに組み込み、新しく作文能力評価モデルを作成した。以前の分析で得られたモデルと対比しながら検討する。

FIGURE 3-1 は、平 (1995) の項目分類に基づく作文能力評価モデルである。ここでの「基礎作文力」はそのまま本研究の項目構成と一致する。一方、「文章力」

は、「課題への忠実性」と「物語作文力」の項目を合わせ「リズム・テンポ」を除いたものである。

FIGURE 3-2 に本研究から得られた作文評価の尺度を用いた新しいモデルを示す。

なお、いずれも太い線で示したパスが0.1%水準でパス係数が有意となったものであり、細い線で示したパスが5%水準で有意となったものである。係数の値が入っていないパスは、平 (1995) の FIGURE 5 と同じ値を持つ。

考 察

本研究における多次元級内相関係数、多次元アルファ係数を用いる作文評価データの分析方法は、1つの最適解を理論的に与えるものではない。したがって、得られた構造を経験的に検討して、妥当なものかどうかを判断することが肝心である。

FIGURE 3-1 と FIGURE 3-2 を比較すると、本研究の方法を用いた利点が浮かび上がってくる。まず、単純和を基にした分析によると、FIGURE 3-1 の「文章力」は「作文習慣」と「感想楽しい」という2つの要因からのパスを受けていた。それに対して、FIGURE 3-2 ではそれが「課題への忠実性」と「物語作文力」とに分かれた。この分類は当初想定した仮説的構造に対しても、より近い。そして、「課題への忠実性」は「作文習慣」からのパスを受け、「物語作文力」は「感想楽しい」と新たに「作文、読書好き」からのパスを受けている。すなわち、平 (1995) ではこの作文評価データにおける情緒的要因の重要性が強調されていたが、「文章力」として未分化にされていたもののうち、特に、本研究で

* : $p < .05$, ** : $p < .001$

FIGURE 3-1 単純和に基づくパスダイアグラム

FIGURE 3-2 第VI解の構造に基づくパスダイアグラム

言うところの「物語作文力」が情緒的要因の影響を強く受けたのであり、「課題への忠実性」はむしろ「基礎作文力」とともに、自主的に文章を書く習慣との関係が深いことが分かったのである。

以上の分析は、多次元級内相関係数、多次元アルファ係数を応用して初めて可能になった。したがって、作文評価データにおいては、多次元構造を保ったまま3相データを縮約して評価構造の分析を行うための1つの合理的な基準として、MICC, MACは有効であることが示された。

反面、本研究にはいくつかの課題が残されていることも否めない。

まず、重みベクトルの選択の方法の煩雑さである。本研究の場合、多次元アルファ係数を応用した初めてのケースであったために、重みベクトルの選択は試行錯誤的であった。本研究では、1度の手続きで適切な解と因子数が得られたが、判断を誤ると何度もやり直しをしなければならない。MICC, MACの考え方を多次元的構造を取り出すための方法として位置づけた場合、解の選択基準そのものはより主観的なものでよいかもしれない。しかし、その場合には、各解を用いて得られた構造自体の良し悪しを相互に評価する手続きが必要になるであろう。

本研究においては、評価を説明する因果モデルの中で新しい知見が得られたため、経験的に多次元アルファ係数の応用の有効性を示すことができた。しかしながら、このようなことが成立するデータの条件については、本研究から示唆することはできない。

さらに、本研究では、構造の分析にはMICC, MACの第VI解の重みを用いながら、作文評価に影響する要因の分析では単純和を用いた。それは、作文の出来そのものを評価する際に、一部の採点結果に負の重みを用いることを躊躇したためである。また、分析的評価に対して第VI解の重みを用いた場合、総合評価にはどのような重みを用いるべきか、という判断が難しくなるからである。この点に関しての妥当性については、議論の余地があるかもしれない。

いずれにしろ、今後、理論的な発展を待つとともに、作文評価データのみならず、多次元アルファ係数の考え方を応用した3相データの構造分析を経験的に蓄積して、この方法の有効性と限界を認識していくことが必要だと思われる。

引用文献

- Ackerman, T. A., & Smith, P. H. 1988 A comparison of the information provided by essay, multiple-choice, and free-response writing tests. *Applied Psychological Measurement*, 12, 117-128.
- Coffman, W. E., & Kurfman, D. A. 1968 A comparison of two methods of reading essay examinations. *American Educational Research Journal*, 5, 99-107.
- De Gruijter, D. N. M. 1980 The essay examination. In L. J. Th. van der Kamp, W. F. Langerak & D. N. M. De Gruijter(Eds), *Psychometrics*

- for *Educational Debates*. John Wiley & Sons Ltd. Pp.245—262.
- De Gruijter, D. N. M. 1984 Two simple models for rater effects. *Applied Psychological Measurement*, 8, 213—218.
- 平井洋子・渡部 洋 1994 小論文評点のカテゴリ化に関する測定論的考察 行動計量学, 41, 21—31.
- 池田 央 1973 心理学研究法 第8巻 テストII 東京大学出版会
- Remondino, C. 1959 A factorical analysis of the evaluation of scholastic compositions in the mother tongue. *British Journal of Educational Psychology*, 30, 242—251.
- 芝 祐順 1979 因子分析法 第2版 東京大学出版会
- 平 直樹 1995 物語作成課題に基づく作文能力評価の分析 教育心理学研究, 43, 134—144.
- 渡部 洋・曹 亦薇 1992 小論文評価における字の美しさの影響について 東京大学教育学部紀要, 32, 253—256.
- 柳井晴夫 1994 アルファ信頼性係数の多次元への拡張 行動計量学会第22回大会発表論文集, 140—143.
- 柳井晴夫・繁樹算男・前川真一・市川雅教 1990 因子分析 その理論と方法 朝倉書店
- Yanai, H., & Taira, N. 1996 Generalization of alpha coefficient and intraclass correlation coefficient to three dimensional data, *Program and Abstracts : 1996 Annual Meeting of the Psychometric Society*, 121.
- Yanai, H., & Taira, N. 1997 Generalization of alpha coefficient and intraclass correlation to three way data. Manuscript submitted for publication.

謝 辞

本研究においては、様々な点で大学入試センター研究開発部教授柳井晴夫先生にご教授いただきました。また、直交プロクラステス解の計算プログラムは、大学入試センター研究開発部助教授前川真一先生にいただいたものです。両先生の親切なご指導に感謝いたします。

(1997.5.16 受稿, 10.31 受理)