

多因子からなる小論文評価ルーブリック作成の試み

近藤裕子・石川勝彦

はじめに

本研究の目的は、山梨学院大学の初年次ライティング科目「基礎演習Ⅰ」において、1年生が執筆する小論文を評価するとともに、科目の具体的な到達目標として機能するライティング・ルーブリックを開発することである。

「基礎演習Ⅰ」は、1年次前期(15コマ)に必修の初年次ゼミとして位置づけられ、同時に、日本語ライティングの基礎を訓練する科目として運用されている。1つのゼミは15名～22名で構成され、ライティングの基礎を専用のワークブックを使用し、学んでいる。受講生は前期授業終了までに、小論文(テーマは自由、1200字程度)を1本提出することが求められているが、「基礎演習Ⅰ」はライティング科目であると同時に、初年次ゼミとして位置付けられているため、15コマ全てをライティングのトレーニングに費やすことはなく、おおよそ10コマを利用している。

「基礎演習Ⅰ」専用のライティング・ワークブックの構成は、「①小論文の特性(問い、主張、理由から成る)」、「②理由を示すとは(問い、主張、理由の繋がり)」、「③理由の示しかた(理由は客観的な根拠であること)」、「④小論文の構成(序論[問い・主張]、本論[理由・根拠]、結論[まとめ・限界]から成る)」、「⑤技法(注の付け方、剽窃の禁止)」の主に5つのパートから成っており、ワークブックの構成が示しているように、小論文の基本的な型・構成の習得と論理的な文章作成の訓練を主たる目的としている。

このワークブックで基礎を学び、最終的に書きあげた小論文は、石川・児島(2018)が開発したルーブリック(Table 1)を用いて評価を行っている。この石川・児島(2018)の開発したルーブリックは、「基礎演習Ⅰ」において、平均的な小論文(学内で実施している全受講者の参加する小論文コンテストにおける非受賞論文)のチェックリストとしては、多くの観点で、平均的な小論文もマイルストーンに到達できるレベル設定となっており、十分に機能している。しかし、平均的な小論文(コンテスト非受賞論文)と卓越した小論文(コンテスト受賞論文)の識別において一定の不十分さを抱えていた。そのため、これらを十分に識別するルーブリックを新たに開発するに至った(石川・児島、2018)(Table 2)。開発したルーブリックは、コンテスト受賞 vs 非受賞という、大きな質的差異に敏感なルーブリック項目を備えていることが確かめられた。しかしな

Table 1 平均的な小論文の評価に適したルーブリック (石川・児島、2018)

評価観点\評価レベル	観点	良 (3点)	可 (2点)	不可 (1点)
日本語表現・ルール	誤字脱字	誤字・脱字がない	誤字・脱字が3個以下である	誤字・脱字が4個以上ある
	文の簡潔さ	一文が簡潔にまとめられ、文章もわかりやすい	一文が長く、一部の文章にもわかりにくい箇所がある	一文が長く、文章全体もわかりにくい
	段落分け	全体的に段落分けが適切になされている	適切な段落分けが一部だけである	全体として段落分けが十分にされていない
	言い回し	正しい日本語の言い回しができている	日本語の言い回しに関するおかしな点が3個以下である	日本語の言い回しに関するおかしな点が4個以上ある
体裁	注の適切さ	全体として注が適切につけられており、図書・雑誌・インターネット記事等の資料情報についても適切な形式で書かれている	注が一部不適切につけられている、または図書・雑誌・インターネット記事等の資料情報について一部不適切な形式で書かれている	全体として注のつけかたが不適切もしくは十分につけられておらず、または図書・雑誌・インターネット記事等の資料情報についても全体的に不適切な形式で書かれている
構成	問い・主張・理由	問い・主張・理由がきちんと書かれている	問い・主張・理由が書かれているものの、わかりづらい	問い・主張・理由が書かれていない、あるいは非常にわかりづらく書いてある
	序論・本論・結論	序論・本論・結論が適切に書かれている	序論・本論・結論という形式では一部書かれていない	序論・本論・結論という形式でほとんど書かれていない
説得性	証拠の信頼性	理由をサポートする証拠(例: 数値で示されるデータ等)が信頼できるものである	理由をサポートする証拠(例: 数値で示されるデータ等)の一部が若干信頼できるか疑わしい	理由をサポートする証拠(例: 数値で示されるデータ等)が信頼できるか非常に疑わしい
	証拠の十分さ	理由をサポートする証拠について十分に調べてある	理由をサポートする証拠について少ししか調べていない	理由をサポートする証拠についてほとんど調べていない
	問いと主張の対応	問いと主張と理由がしっかりとつながっている	問いと主張と理由のつながりが弱い	問いと主張と理由のつながりがきちんとつながっていない

Table 2 平均的な小論文 (コンテスト非受賞論文) と卓越した小論文 (コンテスト受賞論文) に適したルーブリック (石川・児島、2018)

評価観点\評価レベル	観点	良 (各5点)	可 (各3点)	不可 (各1点)
構成	問い・主張・理由	問い・主張・理由がきちんと書かれている	問い・主張・理由が書かれているものの、わかりづらい	問い・主張・理由が書かれていない、あるいは非常にわかりづらく書いてある
	序論・本論・結論	序論・本論・結論が適切に書かれている	序論・本論・結論という形式では一部書かれていない	序論・本論・結論という形式でほとんど書かれていない
	段落分け	全体的に段落分けが適切になされている	適切な段落分けが一部だけである	全体として段落分けが十分にされていない
問い	問いの焦点化	問いが十分に焦点化されている	問いがある程度焦点化されているが、ややあいまい	問いがあいまいで焦点化できていない
	問いの魅力	問いが「答えを知りたい」と思わせるものである	問いがある程度「答えを知りたい」と思わせる	問いが「答えを知りたい」と思わない
論証	発見的な面白さ	論証の内容に発見的な面白さがある	論証の内容は堅実だが面白味みかける	論証の内容が平板でつまらない
	論証の説得力	論証には「なるほど」と思わせる説得力がある	論証は多少説得力がある	論証はあまり説得力がない
文章表現	文章の躍動感	新鮮で躍動感のある文章で書かれている	ある程度新鮮で躍動感を感じさせる文章で書かれている	メリハリのない文章で書かれている
	文の簡潔さ	一文が簡潔にまとめられ、文章もわかりやすい	一文が長く、一部の文章にもわかりにくい箇所がある	一文が長く、文章全体もわかりにくい
オリジナリティ	独自の視点・発想	書き手の独自の視点・発想が盛り込まれている	書き手の独自の視点・発想が、ある程度みられる	書き手の独自の視点・発想が、新しみが少ない

がら、教育の現場では、15コマの授業運用を通じて授業目標に到達するための形成的評価に最適化したルーブリック、つまり、評価の観点が過不足なく、また、レベル設定が適切であり、授業目標と対応したルーブリックが求められている。

そこで、本研究は「基礎演習Ⅰ」の小論文の到達目標に最適化したライティング・ルーブリックの開発を試みることにした。具体的には、実際に「基礎演習Ⅰ」の受講生が提出した小論文を参照しながら評価項目を構成する、という研究方法を採用した。つまり、現状の受講生のライティング能力に適合した観点を探索し、レベル設定を模索するとともに、提出された小論文に欠けている特性を析出することをねらいとする。

特に、本研究ではできるだけ多くの観点を案出し、網羅性の高い項目プールを作成することを試みる。Table 1・2にあるように、石川・児島(2018)のルーブリックは、授業での運用可能性を保証するため、観点を必要十分な数に制限している。さらに、1つの観点を1項目で測定する、という方針に基づき開発を行った。これは、評価者の負担が極めて軽くなるため運用可能性の観点からは望ましいことではある。しかし、小論文という複雑な構成物を1観点1項目で測定しようとする、1つの項目に幅広い内容を詰め込まざるを得なくなり、ボリュームが大きくなる、項目内容が抽象的になるといった傾向が確認された。逆に、1つの項目をコンパクトに収めようすると、内容を制限するがゆえに、主要なポイントを取りこぼすことにもつながる可能性もある。以上のことから、鈴木(2018)が指摘するように、原則的には1つの構成概念を過不足なく測定するためには、必要十分な項目数を費やすことが望ましいと考える。

そこで、新たなルーブリックにおいては、項目数を制限せず、1つの観点につき複数の項目を設定し、具体的かつシンプルな項目の整備を目指すこととする。

さらに、案出した項目を実際に採点に運用し、採点データを取得する。統計的な手法を用いて設置項目の自体を評価することが関心の中心である。得られた項目の適切性は項目反応理論(Item Response Theory: Lord & Nocick, 1952)を適用して評価する。項目反応理論は各項目の正答率から、識別力母数(a)、困難度(b)、受験者能力 θ を推定する手法である。適した項目の条件の1つは、困難度が高い項目ほど能力の高い受験者でなければ正答できないというという特性を備えていることである。さらに、識別力が適度に高いことも重要である(豊田, 2002)。識別力の低い項目は受験者の能力が大きく異なってもスコアがあまり変化しない項目であり、識別力の高い項目は受験者の能力が変化するに応じて敏感にスコアが変化する項目である。項目反応理論の利点は識別力を項目ごとに推定できる点である(加藤・山田・川端, 2015)。

なお、推定されたパラメーター情報は、授業が執筆者にもたらした学習上の影響を解釈するうえでも有用である。「基礎演習Ⅰ」を通じて学習した内容であるかどうか

によってスコアの高低がどのように変動するか観察することにより、「基礎演習Ⅰ」の授業内容をどのように見直していけばよいかという考察につなげたい。

方法

評価に用いた小論文

山梨県内の A 大学におけるライティング科目「基礎演習Ⅰ」において執筆された小論文156本を対象とした。うち42本は、受講生全員が参加した小論文コンテストのクラス代表に選定された小論文である。つまり、平均的な小論文とみなされた小論文が114本、卓越した小論文とみなされた小論文が42本であった。

観点の生成方法

日本語教育・ライティング指導を専門とする大学教員1名が、小論文を参照しながら観点を定めた。その際、以下の方法で観点を定めた。まず、基礎的・汎用的なライティング・ルーブリックとして、近藤・由井・春日(2019)に掲載されているルーブリック項目を参照し、「基礎演習Ⅰ」受講生が執筆・提出した小論文を実際に採点しながら項目を精査し、最終的に27項目を得た (Table 3)。

複数評価者データの生成

文化人類学を専門とする1名の大学教員に小論文の採点を依頼し、復習評価者データによる採点データを生成した。評価者はライティング科目にTAとして指導に入り、採点経験を有する。実働は約2年間である。依頼対象に小論文の電子データ及び採点表(Excel)を送付し、156本の小論文に対し、27個の観点(5件法:「1全く適切ではない」～「5非常に適切である」)から評価を行ってもらった。

分析の手順

分析では、複数評価者データを生成した上で、古典的テスト理論、項目反応理論を適用し、不適切な項目を検出して削除するという手続きを取った。このことを通して、採点者のバイアスに依存しない項目を抽出するとともに、困難度のバランスのとれた項目セットであり、かつ十分な識別力を備えた項目セットの構築を目指した。具体的には以下のステップで進めた。

1. 因子分析を行い、項目を因子にまとめる
2. 因子ごとに再度因子分析を行い、次元性を確認する
3. 因子毎に項目反応理論を適用し、項目の識別力及び困難度を推定する

4. 項目毎に項目特性曲線を描画する
5. 因子毎にテスト情報曲線を描画する
6. 識別力の不適切な項目を削除する

Table 3 案出したルーブリック項目

問いと答え	問いがある
	問いに対応した答えがある
	テーマにおいて適切な問いが設定されている
構成	序論・本論・結び（結論）から構成されている
	序論は、問題の背景・問題提起（問い）・答えからなっている
	本論では、答えを裏付けるに十分な根拠を提示している
	結び（結論）では、全体を総括し、答えを繰り返し述べている
説得性	信頼性のある資料を根拠に用いている
	3つ以上の資料を根拠に用いている
	自分の経験や見聞きしたことではなく、資料を用いて示している
	最新の資料データを使用している
	効果的に資料を用いている
	反証など、多視点からも言及している
引用技術	説明を尽くし、論理に飛躍がない
	自分の文章と資料の区別がある
	必要な引用をしている
	どこからどこまでが引用か明確である
	出典の示し方が適切である
文章技法	資料をどのように捉えたか、解釈が加わっている
	文体・語彙・表現など、小論文にふさわしい書き言葉で書かれている
	適切に段落を設けている
	表記が適切だ
	文と文のつながりがある
	句読点が適切に打たれている
論展開	一文の長さが適切だ
	オリジナリティがある
	思考の形跡があり、論展開も読み手が納得できるものである

結果

因子分析

27項目を対象に平行分析を行ったところ、対角 SMC が7 因子、MAP が4 因子を提案した。最小二乗法（プロマックス回転）による因子分析を行った結果の因子の解釈可能性およびスクリープロット上の固有値の落ち込みを考慮し、6 因子構造をベースに項目の整理を行った。いずれの因子にも .40以下の因子負荷量を示す項目を削除し、因子分析を繰り返したところ、23項目 6 因子が最も単純な構造かつ解釈可能性が高いと判断した (Table 4)。なお、削除した項目は「序論は、問題の背景・問題提起 (問い)・答えからなっている」「反証など、多視点からも言及している」「文と文のつながりがある」「出典の示し方が適切である」の4 項目だった。

第1 因子は「説明を尽くし、論理に飛躍がない」「本論では、答えを裏付けるに十分な根拠を提示している」「効果的に資料を用いている」など、論証の確かさに関する項目がまとまったため「説得性」と命名した。第2 因子には「最新の資料データを使用している」「3つ以上の資料を根拠に用いている」「信頼性のある資料を根拠に用いている」など根拠資料の信頼性を中心とした項目がまとまったため「根拠の信頼性」と命名した。第3 因子には「適切に段落を設けている」「自分の文章と資料の区別がある」「どこからどこまでが引用か明確である」など引用を中心とした形式面に関する項目がまとまったため「形式」と命名した。第4 因子には「テーマにおいて適切な問いが設定されている」「問いがある」「問いに対応した答えがある」など問いと、問いに対応した回答が完備されているかに関する項目がまとまったため「問い」と命名した。第5 因子には「句読点が適切に打たれている」「一文の長さが適切だ」など文章の技法に関する項目がまとまったため「技法」と命名した。第6 因子には「文体・語彙・表現など、小論文にふさわしい書き言葉で書かれている」「表記が適切だ」など表現の適切性に関する項目がまとまったため「表現の適切性」と命名した。

後の分析で、項目反応理論を用いて各項目の識別力および困難度を確認するが、項目反応理論では測定項目の因子の次元性が強く求められる。次元性が確保されていない項目群に項目反応理論を適用した場合、推定値や情報量が著しく不良となるからだ。そのため因子ごとに改めて因子分析を行った。分析の結果、第6 因子は因子負荷量がやや不良な値 ($\beta < .40$) であったため、分析から除外し、以降の分析は第1 因子から第5 因子を対象とすることとする。なお、第1 因子から第6 因子の個別の因子分析表は Appendix A に整理した。

Table 4 因子パターン

項目	F 1	F 2	F 3	F 4	F 5	F 6	共通性
第1因子 説得性							
説明を尽くし、論理に飛躍がない	1.00	-.08	-.02	.00	-.06	.05	.88
本論では、答えを裏付けるに十分な根拠を提示している	.98	-.03	-.08	.05	-.03	.13	.91
効果的に資料を用いている	.91	.06	-.05	.03	.03	-.04	.88
思考の形跡があり、論展開も読み手が納得できるものである	.90	-.10	.02	.10	-.05	.13	.84
自分の経験や見聞きしたことではなく、資料を用いて示している	.81	.33	-.08	-.12	-.09	-.05	.82
必要な引用をしている	.79	.11	-.04	.14	.06	-.14	.86
資料をどのように捉えたか、解釈が加わっている	.65	.05	.06	.01	.13	.06	.62
第2因子 根拠の信頼性							
最新の資料データを使用している	.01	.76	.01	.04	.07	-.05	.66
3つ以上の資料を根拠に用いている	.25	.69	.04	-.07	-.04	.04	.72
信頼性のある資料を根拠に用いている	.35	.58	.01	-.06	-.06	-.05	.62
序論・本論・結び（結論）から構成されている	-.16	.44	.36	.19	-.01	.17	.61
第3因子 形式							
適切に段落を設けている	-.21	.14	.62	.10	-.13	.19	.44
自分の文章と資料の区別がある	.38	.00	.58	-.10	-.02	-.16	.65
結び（結論）では、全体を総括し、答えを繰り返し述べている	.11	.00	.52	-.03	.09	.26	.48
どこからどこまでが引用か明確である	.48	-.05	.52	-.04	.11	-.24	.82
第4因子 問い							
テーマにおいて適切な問いが設定されている	.04	.01	-.04	.96	-.02	-.06	.89
問いがある	.02	-.03	.11	.91	-.12	-.04	.86
問いに対応した答えがある	.04	-.02	-.07	.80	.15	-.02	.67
オリジナリティがある	.35	.15	.03	.41	.05	.12	.69
第5因子 技法							
句読点が適切に打たれている	-.02	-.15	-.02	.10	.63	.06	.38
一文の長さが適切だ	-.02	.23	-.07	-.12	.56	.14	.37
第6因子 表現の適切性							
文体・語彙・表現など、小論文にふさわしい書き言葉で書かれている	.47	-.17	.12	-.15	-.01	.52	.43
表記が適切だ	-.12	.10	.05	-.01	.13	.41	.21
因子寄与	9.477	6.498	6.310	5.666	2.350	1.231	
α 係数	.96	.84	.77	.90	.45	.39	
ω 係数	.97	.87	.85	.93	.52	.42	

識別力・困難度の推定

本節では、項目反応理論 (Samejima, 1969) のサブモデルである段階反応モデル (推定は重みづけ最小二乗法) を用いて、各項目の識別力と困難度を確認する。

その前に粗点をベースとした場合に因子間でどのような差がみられるか確認するために、尺度得点を整理した (Figure 1)。因子間で統計的な差があるか確認したところ、因子間で平均値に差が検出された ($F(4,620) = 182.15, p < .000$)。大小関係を確認すると、根拠の信頼性・技法 > 形式 > 問い > 説得性となった ($ps < .000$)。尺度得点が高いほど

良い得点がつけられていることから、これは容易さを表していると解釈できる。つまり、根拠の信頼性、技法がもっとも容易な観点であり、形式、問い、説得性の順で困難度が高いと解釈できる。この結果は大まかなイメージとしては有用だが、より厳密には項目反応理論による推定を参照しなければならない。以下で項目反応理論による識別力、困難度を確認していく。

項目反応理論を用いて、各項目の識別力および困難度を推定した (Table 5)。識別力は受験者能力をはっきりと区別する程度を表し、困難度はその項目を通過する難しさを表す。なお第4因子は推定中にエラーを生じたため分析から除外することとした。

特徴的な点のみていく。識別力を見てもと0.70~5.33の間に分布している。豊田 (2002) は最低ラインを0.2であると示しているが、いずれもこの基準をクリアしている。一方「本論では、答えを裏付けるに十分な根拠を提示している」の識別力が5.33を示しており、やや高すぎる値である。識別力は受験者特性 θ と正答率の関数における傾きを表している。あまりに高い場合は、この項目で識別できる能力レンジが狭すぎると解釈できるため望ましくない。

続いて、項目特性曲線を確認しながら、5件法の選択肢が有効だったか、つまり5件法を用いて連続的にグレーディングしやすい項目だったかどうか確認する。項目特性関数を構成する曲線のピークの値 (項目特性関数の値) が0.4など相対的に小さい値にある場合、採点者が5件法で連続的に差をつけることが難しかったと解釈できる。例えば、容易すぎて多くの執筆者が5点をスコアリングされたり、困難すぎるがゆえに多くの執筆者が1点をスコアリングされるような場合である。あるいは項目の性質上5件法などの多段階のグレーディングが難しく、中間的な点数を与えることが難しい場合もこれに該当する。Figure 2 に、中間的な反応カテゴリについて、その項目特性関数の値が0.4未満の項目の項目特性曲線をピックアップした。第1因子では「資料をどのように捉えたか、解釈が加わっている」、第2因子はすべての項目、第3因子は「適切に段落が分けられている」、第5因子では「一文の長さが適切だ」が該当する可能性がある。これらの項目は、識別力の値の観点からは問題がないが、多くの執筆者が特定の値にスコアリングされる傾向があり、5件法が有効に機能していない可能性がある。

このことは困難度を観察して確認することができる。具体的には位置母数の間隔 ($b_1 \sim b_4$) に注目する。第1因子では位置母数の間隔がもっとも広いのは「思考の形跡があり、論展開も読み手が納得できるものである」で2.58という値であった。一方第1因子でもっとも位置母数の間隔が狭いのは「資料をどのように捉えたか、解釈が加わっている」で1.60という値だった。位置母数が狭いということは、反応カテゴリ間の困難度が類似しており、中間的なカテゴリに反応する確率が低いことを意味する。

こうした項目では、5件法の選択肢が機能しにくいと考えることができる。各因子においてももっとも位置母数間の間隔が狭い項目を挙げると、第2因子では「最新の資料データを使用している」で1.32、第3因子では「適切に段落が設けられている」で1.64、第5因子では2項目の位置母数の間隔が類似しており「句読点が適切に打たれている」で2.31、「一文の長さが適切だ」で2.35だった。

項目の困難度の違いを因子ごとにみてもみる。第2因子「根拠の信頼性」と第5因子「技法」は困難度がマイナスの値にふれており、容易性が高いことが伺える。項目ごとにみても、もっとも困難度が高いのは「思考の形跡があり、論展開も読み手が納得できるものである（第1因子）」と「結び（結論）では、全体を総括し、答えを繰り返し述べている（第2因子）」であった。そして「自分の文章と資料の区別がある（第1因子）」「本論では、答えを裏付けるに十分な根拠を提示している（第2因子）」と続いていた。第1因子「説得性」と第3因子「形式」の両因子から項目がピックアップされていた。記述に説得性を備えることと、小論文を適切な形式に整えることはともに高い困難度を有することがみてとれる。

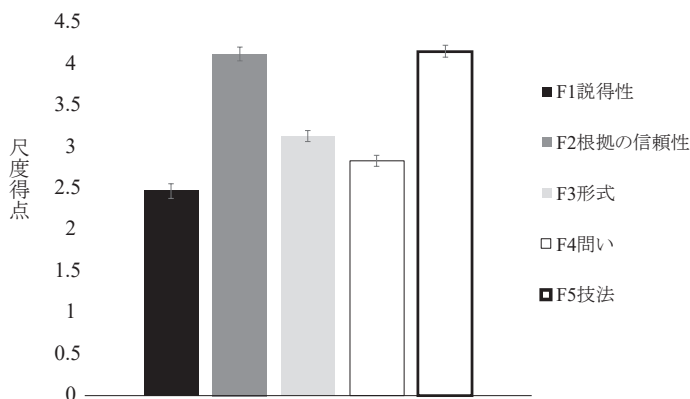


Figure 1 尺度得点

テスト情報曲線

各因子の項目群が執筆者の能力の識別にどのような強みを持っているかを確認するために、因子ごとにテスト情報曲線を描画した (Figure 2)。いずれの図にも受験者能力 $\theta = 0$ に縦破線を示した。第2因子「根拠の信頼性」および、第5因子「技法」は縦破線の左側、つまり受験者能力 θ が低いレンジに高い情報量を示している。特に、第2因子「根拠の信頼性」のテスト情報量は狭い範囲（およそ $-2.0 < \theta < 0.0$ ）に高い識別力を発揮していることが分かる。

他方、第1因子「説得性」、および、第3因子「形式」は縦破線の右側、つまり、

Table 5 項目の識別力と困難度

項目	a	b_1	b_2	b_3	b_4	平均値	SD
第1因子 説得性							
論理に飛躍がない	3.35	-0.52	0.20	0.91	1.59	2.37	1.22
本論では、答えを裏付けるに十分な根拠を提示している	5.33	-0.51	0.35	1.12	1.88	2.22	1.09
効果的に資料を用いている	3.70	-0.66	0.08	0.81	1.63	2.48	1.20
思考の形跡があり、論展開も読み手が納得できるものである	2.66	-0.50	0.44	1.32	2.08	2.15	1.05
自分の経験や見聞きしたことではなく、資料を用いて示している	1.96	-1.00	-0.24	0.50	1.27	2.85	1.29
必要な引用をしている	3.25	-0.82	-0.20	0.51	1.25	2.79	1.31
資料をどのように捉えたか、解釈が加わっている	1.36	-0.30	0.10	0.72	1.30	2.49	1.49
第2因子 根拠の信頼性							
最新の資料データを使用している	2.17	-1.32	-0.93	-0.76		4.33	1.34
3つ以上の資料を根拠に用いている	2.12	-1.36	-0.96	-0.58		4.10	1.46
信頼性のある資料を根拠に用いている	1.45	-1.21	-0.84	-0.34	0.40	3.58	1.46
序論・本論・結び（結論）から構成されている	1.16	-2.74	-2.16	-1.58	-0.84	4.55	0.90
第3因子 形式							
適切に段落を設けている	0.87	-2.69	-2.17	-1.83	-1.06	4.53	1.01
自分の文章と資料の区別がある	2.04	-0.71	0.40	1.02	2.06	2.31	1.11
結び（結論）では、全体を総括し、答えを繰り返し述べている	0.70	-3.07	-1.24	0.54	2.08	3.22	1.02
どこからどこまでが 用 用か明確である	2.03	-0.62	0.11	0.69	1.54	2.53	1.31
第5因子 技法							
句読点が適切に打たれている	0.86	-1.78	-1.33	-0.69	0.53	3.72	1.35
一文の長さが適切だ	0.86	-3.41	-2.70	-1.99	-1.06	4.61	0.82

Note. 第4因子は識別力・困難度の推定中にエラーが生じたため推定の対象から除外した。表中 a は識別力、 $b_1 \sim b_4$ は困難度を表す

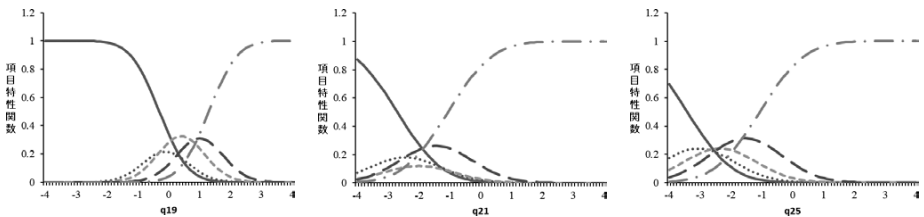


Figure 2 項目特性曲線

左図：第1因子「資料をどのように捉えたか、解釈が加わっている」；中図：第3因子「適切に段落が分けられている」；右図：第5因子「一文の長さが適切だ」

受験者能力 θ が高く、相対的に能力の高い執筆者たちの小論文の識別に対応した因子であることが見てとれる。特に、第3因子は、低い値とはいえ、 $\theta = -4.0$ 付近から $\theta = 3.0$ 付近まで広いレンジをカバーしており、もっとも広い範囲の執筆者の能力を評価できる因子であることが見てとれる。

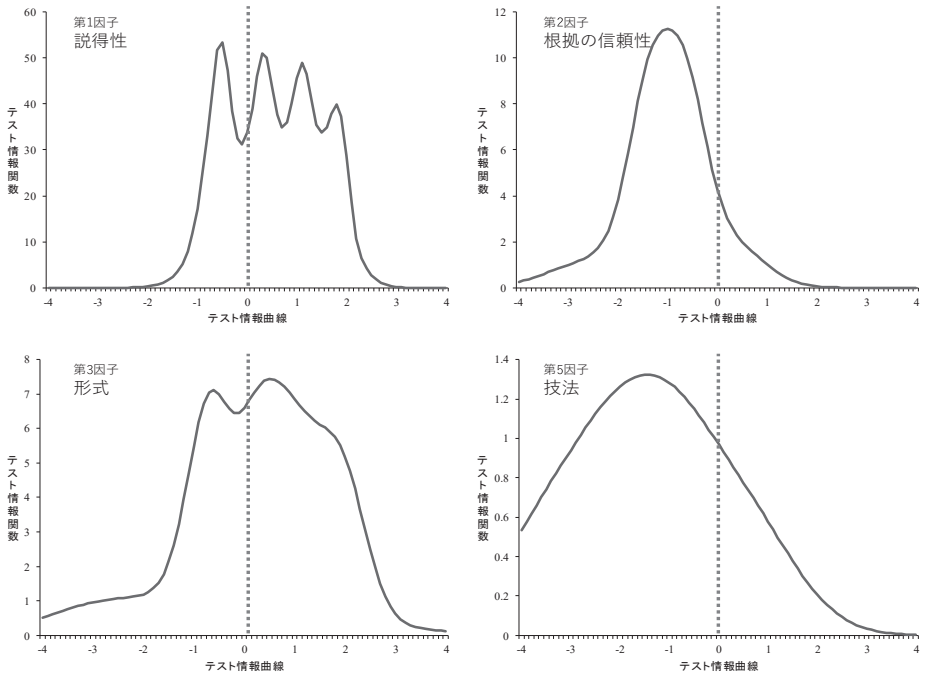


Figure 3 各因子のテスト情報曲線

まとめと考察

本研究は、初年次ライティング科目に対応したルーブリックの開発の試みである。

実際に執筆された小論文を読みながら観点を生成したところ、6つの構成概念について27項目が得られた。因子分析を行ったところ、「説得性」、「根拠の信頼性」、「形式」、「問い」、「技法」の6因子が得られた。統計的な根拠に基づき「説得性」、「根拠の信頼性」、「形式」、「技法」の4因子について項目反応理論によりパラメーター推定を行った。

テスト情報曲線から、根拠の信頼性、技法の難易度が低く、説得性、形式の難易度が高い傾向がみられた。こうした傾向は「基礎演習Ⅰ」の授業内容とどのような関係にあるだろうか。

基礎演習Ⅰにおいて学習している内容は主に「①小論文の特性（問い、主張、理由から成る）」「②理由を示すとは（問い、主張、理由の繋がり）」「③理由の示しかた（理由は客観的な根拠であること）」「④小論文の構成（序論〔問い・主張〕、本論〔理由・根拠〕、結論〔まとめ・限界〕から成る）」「⑤技法（注の付け方、剽窃の禁止）」である。

困難度が低かった「根拠の信頼性」については①②③を通じて繰り返しトレーニングしている。同様に困難度が低かった「技法」については、特にワークブックに学習項目としては立てられていない。ワークブックの運用の中で自然に適切な技法が習得されているか、大学入学以前に習得されている可能性がある。

困難度が高かった「説得性」、「形式」について見てみる。「説得性」の項目を見ると、①②③④によってトレーニングが試みられていると考えられる。一方で、特に「思考の形跡があり、論展開も読み手が納得できるものである」「説明を尽くし、論理に飛躍がない」「本論では、答えを裏付けるに十分な根拠を提示している」といった項目に十分キャッチアップした小論文は少数であるとの結果が得られた。学習に不十分な点が残されていることが伺える。「形式」については、「自分の文章と資料の区別がある」「結び（結論）では、全体を総括し、答えを繰り返し述べている」「どこからどこまでが引用か明確である」のいずれも困難度が高く評価された。「形式」に含まれる項目は⑤を通じて一定程度の学習を行っているが、不十分な点が残されていることが伺える。

授業の改善が必要と思われると同時に、これらの項目が求める水準が1年生前期にふさわしいものかどうか併せて検討しなければならないだろう。

本研究は、実際に受講生が執筆した小論文を参照しながら項目を案出したこと、1つの概念を無理やり1項目で測定するのではなく複数の項目で測定すること、の2つの方針により進めた。分析の結果、各観点の困難度がおおよそ見え、対応して、授業を通じて学習を支援することが容易な観点と、授業での学習を進めているにもかかわらず、習得が困難である観点が明らかになった。

今後の方向は、採点者をもう1人用意し、採点データを取得して知見の一般性を確認する必要がある。ループリック評価を含むパフォーマンス評価は、正答が一意に決まる多肢選択式のテストと異なり、採点表の運用自体は採点者の主観的な運用とならざるをえない。複数の評価者のスコアがあれば、評価者バイアスを考慮しながらパラメーターを推定でき、知見の一般性に寄与しうる。

引用文献

- 石川勝彦・児島功和 2018 識別力を重視したライティングループリック開発の試み—分散分析を用いた特異項目機能分析—。山梨学院大学経営情報学論集, 24, 45-58.
- 近藤裕子・由井恭子・春日美穂 2019 失敗から学ぶ大学生のレポート作成法。ひつじ書房.
- 加藤健太郎・山田剛史・川端一光 2014 Rによる項目反応理論。オーム社.
- Lord, F.M. & Nockick, M.R. 1952 Theory of Test Score. *Psychometric Monograph*, 7, Psychometric Society.
- Samejima, F. 1969 Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 17.

Appendix A 因子毎の因子負荷量

第1因子	F1	共通性
効果的に資料を用いている	.95	.90
本論では、答えを裏付けるに十分な根拠を提示している	.94	.89
説明を尽くし、論理に飛躍がない	.93	.86
必要な引用をしている	.91	.83
思考の形跡があり、論展開も読み手が納得できるものである	.90	.80
自分の経験や見聞きしたことではなく、資料を用いて示している	.86	.74
資料をどのように捉えたか、解釈が加わっている	.76	.57
因子寄与	5.59	
α 係数	.96	
ω 係数	.96	

第3因子	F1	共通性
どこからどこまでが引用か明確である	.84	.71
自分の文章と資料の区別がある	.82	.68
結び（結論）では、全体を総括し、答えを繰り返して述べている	.58	.34
適切に段落を設けている	.48	.23
因子寄与	1.95	
α 係数	.77	
ω 係数	.80	

第5因子	F1	共通性
一文の長さが適切だ	.70	.49
句読点が適切に打たれている	.46	.21
因子寄与	.70	
α 係数	.45	
ω 係数	.45	

第2因子	F1	共通性
3つ以上の資料を根拠に用いている	.85	.72
最新の資料データを使用している	.80	.63
信頼性のある資料を根拠に用いている	.77	.59
序論・本論・結び（結論）から構成されている	.65	.43
因子寄与	2.37	
α 係数	.84	
ω 係数	.86	

第4因子	F1	共通性
テーマにおいて適切な問いが設定されている	.95	.91
問いがある	.92	.84
問いに対応した答えがある	.80	.64
オリジナリティがある	.71	.50
因子寄与	2.89	
α 係数	.90	
ω 係数	.91	

第6因子	F1	共通性
表記が適切だ	.64	.41
文体・語彙・表現など、小論文にふさわしい書き言葉で書かれている	.38	.15
因子寄与	.55	
α 係数	.39	
ω 係数	.44	