

# ルーブリックに基づくレポート自動採点システム

山本 恵<sup>1)</sup>, 梅村 信夫<sup>2)</sup>, 河野 浩之<sup>3)</sup>

1) 名古屋外国語大学 現代国際学部

2) 名古屋学芸大学 メディア造形学部

3) 南山大学 理工学部

yamamoto@nufs.ac.jp

## Automated Report Scoring System Based on the Rubric

Megumi Yamamoto<sup>1)</sup>, Nobuo Umemura<sup>2)</sup>, Hiroyuki Kawano<sup>3)</sup>

1) School of Contemporary International Studies, Nagoya Univ. of Foreign Studies

2) School of Media and Design, Nagoya Univ. of Arts and Sciences

3) Faculty of Science and Engineering, Nanzan Univ.

### 概要

レポート評価では、採点者（評価者）による採点結果のばらつき、同一採点者内での採点の偏り、過度の作業負担などが問題視されている。改善に向け、レポート評価のルーブリック（記述的・客観的な採点指標）を基盤とした、レポート自動採点システムを提案する。システムでは、教員の手動採点用として策定したルーブリックの5つの評価観点を細分化し、自動採点用ルーブリックとも言うべき評価項目を設定する。また評価項目から自動採点可能な項目（自動採点結果の信頼性が高いもの）を抽出し、それらの計算結果を説明変数とする予測モデルに基づいて評価観点や成績レベルを推測する。評価実験を通して、①評価観点「文書スタイル」に帰属する項目と観点の手動採点結果間、②5つの評価観点間すべてについて相関を認めた。①・②から、「文書スタイル」の諸項目を精査することによって、レポート自動採点の実現および予測精度の向上が可能と考える。

## 1 はじめに

著者らは、学生のレポート作成能力育成および教員の採点を支援する LMS (Learning Management System) 上の自動採点システムの開発を試みている。

レポート評価では、採点者（評価者）による採点結果のばらつき、同一採点者内での採点の偏り、採点者の時間的負担など、様々な問題がある。多くの採点者はチェックリストや採点時の評価指標（いわゆるルーブリック）を定めて評価の厳正化を保つ努力をしている。したがって自動採点システムを構築するにあたり、こうしたルーブリックを基盤に採点のアルゴリズムを設計することは自然である。

折しも昨今の大学教育では、アクティブラーニング（能動的学習）の必要性やそれらの授業での評価の厳正化について議論されている。このような授業での学修成果を従来のテスト形式で評価

することは困難であり、レポートや論文、プレゼンテーションなど、いわゆるパフォーマンス評価が採用される。特にレポートは、学習者の知識はもとより、思考力や問題解決能力など多くの習熟度を測ることができ、今後ますます有用な評価方法として利用されるであろう。したがって学生のレポート作成指導、教員の採点の厳正化、負担軽減が希求の課題である。

これらの問題を解決するための自動採点システムの開発にあたり、本稿では、ルーブリックの提案、ならびに、自動採点対象となる評価観点・評価項目と評価のアルゴリズムを検討する。また、自動採点対象項目の採点結果から、他の評価観点を推測可能かどうか検討し、システムの基盤となる部分の確立をめざす。

## 2 先行研究

本章では、自動採点システムの先行研究の概観および、代表的なシステムの採点方法について

述べる。e-rater や Jess の特徴と問題点を述べた後、ルーブリックの必要性について述べる。

## 2.1 自動採点システム

自由記述文の試験には、小論文（エッセイ）タイプと短答式タイプがある。ここでは著者らが対象とする小論文の自動採点システムに着目する。

自動エッセイ評価システムの研究は 1960 年代の Page（アメリカ）に始まったとされる[1]。その後、自然言語処理や情報検索技術の進歩、コーパスの利用により、様々な手法によるアプローチが見られ、現在も研究が進んでいる。アメリカではすでに商用システムとして運用段階にはいるものが複数ある。なかでも e-rater、PEG、IEA、IntelliMetric、BETSY、CRASE が代表的なシステムである[2]。e-rater、PEG、IEA、Jess は連続量で総合スコアを計算するのに対し、IntelliMetric、BETSY は、はじめからスコアカテゴリーに分類する手法をとる。ここではアメリカで長きに渡り大規模な試験の採点に利用されている e-rater や IntelliMetric と、日本のセンター試験での利用を目的に開発された Jess について述べる。

e-rater は英語の小論文を対象にした最も代表的なシステムで、ETS（Educational Testing Service）が提供しており、アメリカの GMAT（Graduate Management Admission Test）の小論文や TOEFL 試験などで広く利用されている[3]。構造／組織／内容に関する 12 の変量、例えば、文法エラーの割合、スタイルのエラーの割合、談話（discourse）ユニットの数、語彙の困難度、平均の単語長、単語の総数などを求め、それらに係る重み付けを経験則によって定め、重回帰モデルによりスコアリングを行っている[4]。重み付けは、談話（discourse）ユニットの数が 0.21 と最も高く、次いで単語の総数が 0.20 としている。人間の専門的な評価者の採点結果と比べると 97%の精度である[5]。

IntelliMetric は、知識工学的なアプローチで採点する。予め採点が終わっている模範解答を学習し、各採点ポイントのデータから、人間の採点者の採点ルールを推定する。すなわち、各エッセイがどのスコアに相応しいか分類する。一貫性／内容／構成など 5 つの評価スコア観点を各々 1～6 点、および総合点を 6 点満点で評価する [4]。

Jess は、日本語小論文の自動評価システムとして、日本で初めて、石岡らによって開発された。評

価基準は e-rater の構造、組織、内容をそのまま踏襲しており、修辞、論理構成、内容という観点で評価している。10 点満点で総合点を算出する。3 つの観点の重み付けは 5:2:3 であるが、ユーザーが変更可能である[6]。

## 2.2 ルーブリックの必要性

石岡は、自動採点システムに望まれる要件の議論で、e-rater V.2.0 および Jess は論題によらず評価モデルは一定で、評価基準表に従った採点を行っている、として妥当性を述べている[7]。ここでいう評価基準表は、いわゆるルーブリックととらえることができるが、アメリカの経営大学院への入学試験である GMAT の採点基準を踏襲して設定している。

本研究では大学の授業で一般的に提出されるレポートの採点を対象としており、教育現場で実践的に活用できるシステムをめざしているため、既存のレポート評価のルーブリックを綿密に分析・作成し、これに基づいて自動採点システムを設計すべきであると考えます。

## 3 ルーブリック評価と自動採点

本章では、先行研究を調査・比較し、レポート評価のためのルーブリックを提案する。次に、具体的な提案ルーブリックの各評価項目に対し、文体の統一性や誤字・脱字など、自動採点を行うことが可能な項目を示す。

### 3.1 ルーブリックに関する先行研究

ルーブリックの動向に関する論文[8]によると、米国では全米カレッジ・大学協会（以下 AAC&U）がバリュープロジェクトを立ち上げ、機関を超えて活用可能なルーブリックを、1 年以上かけて作成している。これは探求と分析力、批判的思考力など基本学習成果として 16 領域をあげ、各々のルーブリックを作成したものである。

日本では、学習指導要領改訂により初等中等教育でルーブリックによる成績評価が進んでいるが、高等教育では現在も発展途上と言える。近年、大学など多くの研究機関で研修会が盛んに行われ[9]、成績評価への導入が推奨されている。授業科目によって評価内容が異なる点是否めないものの、可能な範囲で汎用性を保てるルーブリックの作成をめざし、レポートや論文、自由記述に関する汎用性の高いルーブリックを先行研究から洗い出し、特徴を調べた。

様々な大学で運用されているルーブリックのうち、レポート評価に関するもので、大学など組織レベルで作成・活用されているものとして米国のAAC&Uの文章コミュニケーションVALUEルーブリック（Written Communication VALUE Rubric[10]）および、関西国際大学のCOMMONルーブリック（ライティング）[11]があげられる。

### 3.2 採点指標となるルーブリックの提案

ルーブリックは、構造（分析的か一般的か）、スコープ（課題を絞り込むか一般的か）、スパン（長期か短期か）により、様々なタイプが存在する[12]。本研究では、以下の条件のルーブリックを作成する：

- ・構造：評価の観点と評価レベルをマトリックス型で複数設定する分析的な構造
- ・スコープ：学問分野や授業科目は特化しないが、課題がレポート方式に限られるという点で一部限定的
- ・スパン：短期間でスナップショット的に使われる採点用ルーブリック。

表1にレポート評価のための手動採点用ルーブリックおよび自動採点用ルーブリックを提案す

る。評価観点は大綱的であるため、細分化して評価項目を設定し、コンピュータによる自動採点のためのルーブリックを作成した。提案ルーブリックは3.1節で述べた文章コミュニケーションVALUEルーブリックおよび、COMMONルーブリック（ライティング）を参考に作成した。なお前者については、AAC&Uのウェブサイトで開催されている原文と松下の訳[12]を参考にしている。

評価項目については、ルーブリックに関する論文やルーブリックバンクなどWeb上で公開されているレポート採点用ルーブリックで多く見られる項目や表現を参考にした。教員の手動採点では、5つの評価観点：課題の理解度と解答（記述）内容の妥当性（Content）、論理的な展開（Structure）、資料と根拠（エビデンス）の妥当性（Evidence）、文章作法の遵守と適切な推敲（Style）、読みやすさ・表現の巧みさ（Skill）、のそれぞれについて、A+、A、B、C、Dの5段階の成績レベルとして記述されている内容に当てはまる位置を確認する。さらに、各レベルの中でどちらの点数になるか決め配点する。例えば、論理的な展開についてBレベルであれば、4点か5点かを定めることになる。

表1 レポート評価のためのルーブリックおよび自動採点用ルーブリック

手動採点時ルーブリック					自動採点時ルーブリック		
評価観点	達成レベルと配点					評価項目	自動採点可能
	D(0-1)	C(2-3)	B(4-5)	A(6-7)	A+(8-9)		
[Content]							
課題の理解度と解答（記述）内容の妥当性	解答内容が、課題とは無関係である。	課題を理解解答しているが、誤りがある。	課題を理解解答しているが、記述が不足している。	課題を理解し的確な解答であるが、改善の余地がある。	的確な解答である。関連用語を適正に用いている。改善の必要はない。	1 論題と記述の合致度 2 主要な関連語の存在 3 出題意図の理解度 4 内容の総合評価 5 学修内容の理解度	○ ○ × × ×
[Structure]							
論理的な展開	記述内容にまとまりがない。	理論の展開に矛盾がある。	順序立てて理論を展開しているが、改善すべき点が複数ある。	順序立てて理論を展開しているが、説得力がない。	順序立てて理論を展開している。意見・主張があり、説得力がある。	6 論理性の水準 7 意見・主張の妥当性 8 事実と意見の区分け 9 説得力	× × × ×
[Evidence]							
資料と根拠（エビデンス）の妥当性	資料を全く参照していない。根拠を示していない。	資料を参照していないが、根拠を示そうとしている。	参照しようとしている資料は相応しくない、または信頼性がない。	信頼でき、関連性のある資料を参照しているが、引用・参照の仕方に誤りがある。	当該の学問分野に相応しく信頼でき関連性のある資料を、うまく使いこなしている。	10 参照資料の質水準 11 参照資料の関連性 12 論拠資料の妥当性 13 図表への説明付加 14 引用量の妥当性	× × × × △
[Style]							
文章作法の遵守と適切な推敲	複数にわたってルールを守っていない。文章が全く推敲されていない。	ルールを守っていない、誤字・脱字、文体の誤りなどが複数ある。	大よそのルールを守っているが、訂正すべき点が複数ある。	訂正すべき点はないが、改善の余地がある。	よく推敲してある。誤りが全くない。	15 文体の統一性 16 誤字・脱字の排除 17 構文の妥当性 18 主述関係の妥当性 19 句読点の妥当性 20 冗長さの排除 21 表記ゆれの排除	○ ○ ○ ○ ○ ○
[Skill]							
読みやすさ・表現の巧みさ	文章が読み辛い。明らかに文章スキルがない。	文章が長すぎるなど、複数の改善すべき点がある。	文章が概ねまとまっているが、改善すべき点がある。	文章が読みやすい。語彙が豊富である。	読み手に明確に意味を伝えることができ、読みやすい。語彙が豊富である。	22 漢字の使用率 23 文長の妥当性 24 語彙の豊富さ 25 語彙の水準	○ ○ ○ ○

したがって観点ごとに、0-9 の範囲で評価値が決まることになる。

### 3.3 ルーブリックにおける自動採点可能項目の抽出

自動採点可能かどうかを判断し、自動採点時ルーブリックの「自動採点可能」欄に、○/△/×で示す。評価観点のうち Content、Structure、Evidence (以下 CSE) は教員の価値判断に強く依存するため、自動採点することが困難である(×や△で示す項目)。他方、Style、Skill (以下 SS) の多くの項目は、自動採点の対象となり得ると判断する(○で示す項目)。

## 4 ルーブリックに基づく自動採点システム

本章では、計算機内部における学生レポートを処理する過程をフローチャートで示した後に、前章で提案した自動採点用ルーブリックの各評価項目の計算方法について説明する。

### 4.1 自動採点システムのアーキテクチャ

図 1 に自動採点の流れを示す。

表 1 のルーブリックで提示した評価項目のうち、自動採点可能欄に○を付した項目のみ自動採点を行う。

LMS 上で入力された学生レポートのテキストデータは、文章校正ツールにより基本的なチェックを行い、エラーの内容と数を算出する。また形態素解析器に付属の辞書を参照しながら、形態素に分類し、文節にまとめ上げ、品詞や文節位置など、評価項目の計算に必要な情報を付加する。

次に、ルーブリックで提示した評価項目ごとの計算アルゴリズムに必要な変数となる基本統計量を計算する。構文・係り受け処理では、構文の妥当性を計算するための変数を算出する(次節 4.2 参照)。これらの変数から評価項目ごとの特徴量を算出する。特徴量とは評価項目のスコアリングのもととなる値である。例えば、「文体の統一性」であれば「非統一の割合」を意味する(表 2 の No.15 参照)。計算の過程で評価項目によって、語彙水準表や独自辞書を随時利用する。独自辞書については、大学生レベルの書き言葉言語コーパスを現在開発中である。算出された特徴量と、あらかじめ設定した特徴量の基準値(望ましい値)との距離を求め、特徴量とする。

スコアリングでは、特徴量の意味するところに即して値が大きいほど高いスコアになるように補正する。例えば、語彙の豊富さは特徴量が大きいほど配点が高くなるが、誤字・脱字では特徴量が大きいほど配点が低くなる。さらにクラス全体の平均、最大値、最小値をもとに補正し、9 点満点で配点して表示する。

評価観点/成績レベル推定では、教員の手動採点結果を目的変数、自動採点結果を説明変数として線形重回帰分析を行い、重回帰式から評価観点の予測値を求

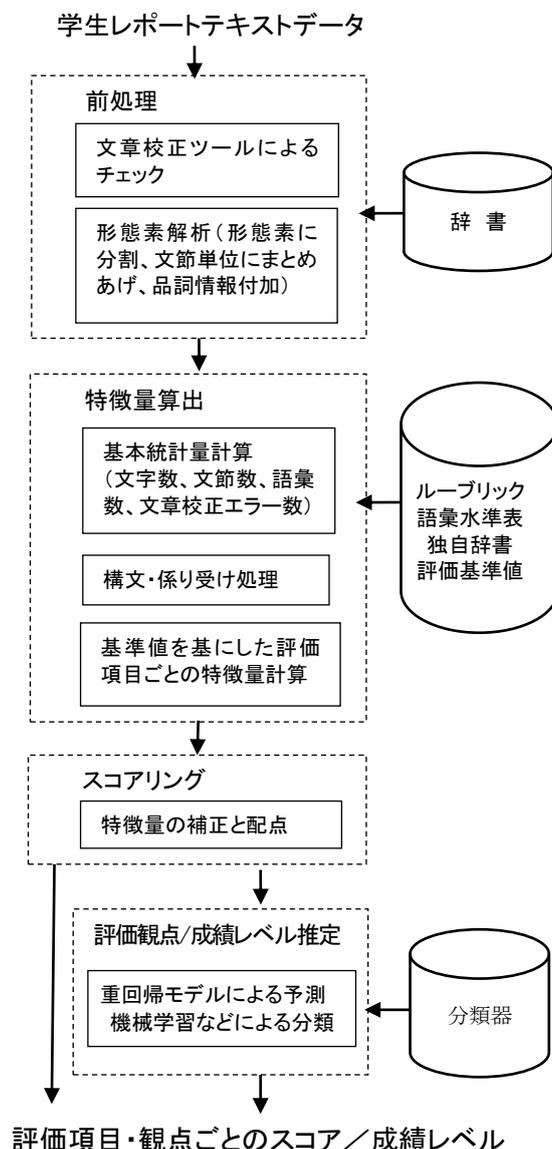


図 1 自動採点の流れ

める。また、過去のレポート採点結果をもとにサポートベクタマシンなどで分類器を作成しておき、SSのスコアリング結果に対して、成績レベル(A+, A, B, C, Dの5段階)の分類を行い表示する。

### 4.2 評価項目の採点に用いるアルゴリズム

自動採点可能項目として判断した評価項目のうち、No.15-25の自動採点計算式は表 2 のとおりである。

「15 文体の統一性」では、文体の非統一箇所の個数を、「16 誤字・脱字の排除」では誤字脱字の箇所を、何れも Microsoft® Office Word 2016 の文章校正ツールを利用して求め、文節数全体の数で割った値を誤りの割合とする。「17 構文の妥当性」では、構文解析ツール CaBoCha が出力する係り受け情報から、係り受け関係にある文節間の距離を算出し、全文節数で除することで、各レポートの平均距離を算出する。距離の妥当性については定まった基準が見当たらないため、新聞記事(2900 文字)から求めた平均距離 1.988 を適正基準と

定め、この指標からの偏差を特徴量とする。「18 主述関係の妥当性」では、係助詞（主語に接尾語）の数を文の数で除した値を算出する。主語と述語が1対1であれば読みやすい文章であり、複文やねじれの可能性が低くなると考え、1.0を基準値とする。「19 句読点の妥当性」では、句読点間の平均文字数を求める。具体的には、レポートの全文字数を句読点の数で除した値により判断する。これらの適正範囲は文章校正ツールに共通で示されているものである。「20 冗長さの排除」では、「の」の繰り返しなど、文章校正ツールが検出する誤りの個数を冗長さの数とする。またレポート中に同じ動詞を5回以上使用している場合（例えば「思います」が頻出するなど）も冗長さとする。RMeCabでタム文書行列を作成し求める。以上の冗長さを示す個数が、全文節数に占める割合を特徴量とする。

表 2 自動採点項目の評価内容

No.	評価項目	評価内容	計算式	適正基準
15	文体の統一性	「である調」または「ですます調」での統一の可否	非統一の割合=不統一の数 <sup>※1</sup> ／文節総数 <sup>※2</sup>	0.0
16	誤字・脱字の排除	誤字・脱字の有無	誤字・脱字出現率=誤字・脱字の数 <sup>※1</sup> ／文節総数 <sup>※2</sup>	0.0
17	構文の妥当性	係り受けに関する文節間の平均距離	係り受けの平均距離=1文書(レポート)の全係り受け数／文節総数 <sup>※2</sup>	1.988 <sup>※6</sup>
18	主述関係の妥当性	1文における係助詞の数	係助詞の平均数=係助詞の数／文の数	1.0
19	句読点の妥当性	句読点間平均文字数	句読点間平均文字数=全文字数／句読点の数	35文字 <sup>※6</sup>
20	冗長さの排除	ことばの重複や繰り返しの有無	冗長さ出現率=(重複語の数 <sup>※1</sup> +語の繰り返しの数 <sup>※5</sup> )／文節総数 <sup>※2</sup>	0.0
21	表記ゆれの排除	表記ゆれの有無	表記ゆれ出現率=表記ゆれ出現数 <sup>※1</sup> ／文節総数 <sup>※2</sup>	0.0
22	漢字の使用率	漢字の使用率	漢字の使用率=漢字の数／全文字数	35% <sup>※6</sup>
23	文長の妥当性	文の平均の長さ	文の平均文字数=全文字数／文の数	35文字 <sup>※6</sup>
24	語彙の豊富さ	トークン比	トークン比 <sup>※3</sup> =異なり語数／述べ語数	値が大きいほど豊富
25	語彙の水準	主要語彙の平均水準	主要語彙の平均水準=全名詞、形容詞、動詞の語彙水準 <sup>※4</sup> *当該語彙の重みの平均	値が大きいほど高レベル

※1 Microsoft® Office Word 2016 の文章校正ツールを利用  
 ※2 1文書(レポート)の全文節数(CaBoChaにより算出)  
 ※3 TeMP[13]により算出  
 ※4 語彙水準表[14]をもとに算出  
 ※5 同じ動詞がレポート中に5件以上繰り返されている場合  
 ※6 適正範囲の中央値

「21 表記ゆれの排除」では、文章校正ツールにより表記ゆれと示される個数を、レポートの全文節数で除すことにより誤りの割合を求め、特徴量とする。「22 漢字の使用率」では、文章校正ツールにより表示される漢字の使用率と基準値との距離を特徴量とする。「23 文長の妥当性」では、レポートの全文字数を文の数で除した値を各レポートの平均文長とみなし、この値と基準値との距離を、特徴量とする。「24 語彙の豊富さ」は、トークン比(異語率)で判断する。トークン比は、異なり語数を述べ語数で除した値で、筆者らが開発したLMS上のテキストマイニングツール TeMP[13]により算出する。「25 語彙の水準」は、砂川ら[14]が研究成果として提供している日本語教育語彙表を用い、レポートから抽出した形態素に対して6段階の評定値を割り当てる。砂川らは、語彙を「初級前半」～「上級後半」6段階のレベルに分け、各語彙にレベルを付記した一覧表を作成している。そこで、レポートに用いられている語彙のうち名詞・動詞・形容詞の索引語(単語)文書行列を TeMP で作成し、上記語彙表に存在する語彙すべてのレベルを割り当て、平均を求める

## 5. ルーブリックに基づく自動採点項目の評価実験

本章では、学生が提出した小レポートを用いて、提案ルーブリックの自動採点可能項目を計算し、教員の手動採点結果と比較することで、是非を確認する。また評価観点間の相関を求め、自動採点可能項目と他の項目との相関を確認する。

### 5.1 分析対象レポートの特性

情報リテラシー科目(表計算ソフトを使ったデータ解析をテーマとする)を受講する43名の大学1年次生が提出した小レポートを、評価実験に用いた。レポートの平均文字数は427.6、標準偏差216.0であった。

なお、今回の実験ではシステム全体の構築の準備段階として、組み込むべき自動採点項目を明らかにする目的であるため、ローカルコンピュータ上で採点を行った。

### 5.2 教員採点と自動採点の相関

教員はルーブリックにしたがい、手作業で評価観点の採点を行った。一方で、No.15～25のSS項目群の自動採点を行い、教員採点結果との相関を求めた(表3)。各評価項目は正規分布とならなかったため、スピアマンの順位相関で求めた。

表3に示すとおり、Styleのほとんどの項目について、ある程度の相関が認められる。ただし、相関係数が低いNo.20・21については、評価のアルゴリズムや項目の在り方を再検討する必要がある。一方、Skill項目群ではNo.22のみ、教員評価との相関が認められた。

また、教員の採点結果を目的変数、自動採点結果の項目群を説明変数として線形重回帰分析を行った。さらに、教員の採点結果（実測値）と重回帰式から算出した予測値との相関を求めたところ、Style 項目群については 0.602 と高い相関を認めたが、Skill 項目群については 0.21 と低い値であった。Style の各評価項目の精度を上げることで、Style 評価観点の予測モデルによる採点が可能であると言える。

表 3 教員の採点と自動採点の相関 (n = 43)

Style 項目群		Skill 項目群	
No.	相関係数	No.	相関係数
15	0.329*	22	-0.322*
16	0.335*	23	0.017
17	0.275	24	0.121
18	0.248	25	-0.144
19	0.375*	予測値	0.21*
20	-0.051	$p < 0.05, p < 0.001.$ (表内の係数値はスピアマンの順位相関に基づく。※は相関が認められることを示す)	
21	0.190		
予測値	0.602*		

### 5.3 CSE-SS 間の相関

教員の採点結果のみで調べたところ、CSE 間、SS 間、さらに CSE と SS 間で 5%水準で有意な相関を認めた(表 4)。特に Style 項目は CSE のうち Content との相関が最も高い。このことは、自動採点で求められる SS から、CSE の評価点に関するある程度の情報(例えば成績のレベルなど)を推測できることを示す。

表 4 教員の評価観点間の相関 (n = 43)

	Content	Structure	Evidence	Style	Skill
Content	—	0.64	0.59	0.53	0.50
Structure	0.64	—	0.48	0.48	0.53
Evidence	0.59	0.48	—	0.32	0.51
Style	0.53	0.48	0.32	—	0.75
Skill	0.50	0.53	0.51	0.75	—

## 6. むすび

レポート採点用の詳細な評価項目をもつルーブリックを提案し、自動採点が可能な評価項目と評価観点を明らかにした。特に Style 項目群のほとんどが、教員による採点結果との相関を認め、自動採点可能であることがわかった。また、5 つの評価観点間すべてについて相関を認めたことから、SS の自動採点結果により、CSE の予測が可能であることがわかった。ただし、Skill 項目群は、アルゴリズムの再検討もしくは、評価項目を追加して精度を高める必要がある。今後は SS の自動採点の精度を高め、その採点結果から CSE の各

評価を推測するとともに、成績の総合レベルを求め、採点資料として提示する。また、LMS への自動採点システムの導入を進める。

## 参考文献

- [1] Semire Dikli, An Overview of Automated Scoring of Essays, The Journal of Technology, Learning and Assessment, Vol 4, No 3, p.5, 2006.
- [2] 石岡恒憲, コンピュータ上で実施する記述式試験について, 情報処理学会研究報告, Vol.2016-CE-133 No.19, pp.1-2, 2016.
- [3] “About the e-rater Scoring Engine”, <https://www.ets.org/erater/about/>, (参照 2016-10-20).
- [4] 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, 第 31 巻, 第 2 号, pp.67-87, 2004.
- [5] Yigal Attali, Jill Burstein, Automated Essay Scoring With e-rater® V.2, The Journal of Technology Learning and Assessment, Vol 4, No 3, p.22, 2006.
- [6] 石岡恒憲, 亀田雅之, コンピュータによる日本語小論文の自動採点システム, 電子情報通信学会技術研究報告. Vol.102, pp.43-48, 2002.
- [7] 石岡恒憲, 小論文およびエッセイの自動評価採点における研究動向, 人工知能学会誌, Vol.23, pp.17-24, 2008.
- [8] 吉田武大, アメリカにおけるバリュールーブリックの動向, 教育総合研究叢書, vol.4, pp.1-12, 2011.
- [9] 佐藤真, 香田 健治, ルーブリックの開発に関するモデレーション研修の比較検討: 総合的学習におけるレポート評価を通して, 教育学論究 Vol.6, pp. 61-68, 2014.
- [10] Association of American Colleges & Universities, VALUE Rubric Development Project, <https://www.aacu.org/value/rubrics/written-communication>, (参照 2016-10-20).
- [11] コモンルーブリック (ライティング) 2012, <http://renkei.kuins.ac.jp/pdf/3writing.pdf>, (参照 2016-10-20).
- [12] 松下佳代, パフォーマンス評価による学習の質の評価: 学習評価の構図の分析にもとづいて, 京都大学高等教育研究, vol.18, pp.75-114, 2012.
- [13] 山本恵, 梅村信夫, 「語彙の豊富さ」からみた学生レポートの分析・評価, 日本ムードル協会全国大会発表論文集, pp.6-8, 2016.
- [14] 砂川有里子, 学習辞書編集支援データベース作成について—『学習辞書科研』プロジェクトの紹介—, 日本語教育連絡会議論文集, pp.164-169, 2012.