

「読みの深さ」を評価するルーブリックの開発過程に関する一考察

—「一般化可能性理論」による信頼性の検討を通して—

神部 智 ・ 山元 隆春¹

A study on the development process of rubrics for assessment of the qualitative level of "reading"
—Through the examination of reliability by the generalizability theory—

Tomo Kanbe · Takaharu Yamamoto

Abstract : In order to evaluate the qualitative level of students' "reading", performance assessment and rubrics are said to be effective. However, it has been pointed out that unrelated factors such as graders and evaluation criterion affect the score value in such performance assessment, and it is necessary to solve this problem. Therefore, authors focused on generalizability theory and proposed rubrics making process to numerically visualize the factors. We also examined the reliability of the rubric making process through applying the generalizability theory empirically by comparing the quality of the evaluation activities with controlled conditions. In addition, we developed the rubric.

Key words : rubrics, generalizability theory, reliability, "The spirit of capitalism in the history of football"

キーワード : ルーブリック、一般化可能性理論、信頼性、「サッカーにおける『資本主義の精神』」

1. 問題の所在

中央教育審議会(2014)は「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について(答申)」において、「高等学校段階の基礎学力を評価する新テストの導入なども踏まえると、『知識・技能』を単独で評価するのではなく、『知識・技能』と『思考力・判断力・表現力』を総合的に評価するものにしていくことが必要である。」(p.14)と報告した。高等学校国語科においては、間瀬(2015)が「高次の能力について、高等学校の国語科とりわけ『読むこと』の領域を中心とした学習指導においては、どのように問い、習得させ、評価することができるのか。日々の授業における取り組みとして、単元学習の発想を取り入れた複合的な学習にパフォーマンス評価を取り入れることがまず想定される。」(p.219)と述べ、一問一答式のテストでは測定しきれない高次読解力を評価するひとつの方向性を示した。「読むこと」領域におけるパフォーマンス評価

としてまず思い浮かべられるのは PISA(Program for International Student Assessment)の読解力調査問題や全国学力学習状況調査B問題だが、多数の評価問題を用意して標準化されたペーパーテストを作成するのは大規模な学力調査の枠組みであればこそ実現可能であり、こうした評価の方法を学校現場で採用することは難しい。一般に、各学校で実践されるパフォーマンス評価とは、小論文・ポスター・口頭発表といった総合的な学習課題の成果物を複数採点者が複数観点から評価するようなものであり(分寺,2016;宇佐美,2010)、そうした評価の方法が形成的評価として有効であることも明らかにされている(石井,2015;遠藤,2015)。実際高等学校国語科においても、学習指導要領やICEモデル等を理論的基盤として学習者の読解の質を評価する観点別評価表が作成されてきた(田中,2016;佐藤・岡田,2010;二田,2009)。しかし、こうしたパフォーマンステスト形式の高次能力評価には採点者の個人差・水準理解度といった様々な要因が影響することが指摘されており(平井,2007;山下・大野・尾関,1990;渡部・平・

¹ 広島大学大学院教育学研究科

井上,1989)、妥当性(狙った構成概念を実際に評価できているか)と信頼性(いつ誰が評価しても安定して学習者の能力を評価できているか)に関する測定論上の課題を克服しなければ、評価データを授業改善や成績評定のために活用しにくいという問題があった。

2. 「一般化可能性理論」とは何か

このような問題の解決に資するため、稿者は統計的手法のひとつである「一般化可能性理論」に注目する。

「一般化可能性理論」とは、「信頼性係数の推定を古典的テスト理論ではなく、分散分析の枠組みを利用して行う(Cronbach,Nageswei,&Gleser,1963)。この方法は、評定者や項目など評価課題の諸側面ごとに誤差を分離して、より洗練された形で信頼性係数を測定することにより、古典的テスト理論に基づく考え方では解決できないパフォーマンス評価に伴う問題点を克服可能である(池田,1994;Shavelson&Webb,1991)。」(松下・小野・高橋,2013,p.112)と説明される。この手法は、採点値の分散成分を相ごとに推定して一般化可能性係数を導く一般化可能性研究(Generalizability study)と、各分散成分の推定値から評価観点や採点者の数をどのように改善すればよいかの示唆を得る決定研究(Decision study)から成り立つ。この手法の特徴は、学習者のパフォーマンスの質の採点値に影響を及ぼしている採点者や評価観点などの要因を考慮することができるため、得点平均値や採点一致率を計算した場合よりもルーブリックの具体的な改善点についての示唆を得やすいという点である(註1)。実際、この一般化可能性理論を適用することによってより質の高い評価のあり方について考察を加えている先行研究は各領域に存在する(森重・渡部,2017;松下ら,2013;山口・清水,2009;石原・山田,2008;山西,2005;佐々木・村木,2005)。しかし国語科「読むこと」の評価領域においては、まだこのような手法で評価の質の向上が検討されたことはない。よって稿者は、この一般化可能性理論という統計的手法を適用することにより、高等学校国語科授業における高次読解力の評価の質(総合的な学習課題における学習者の成果物を複数採点者が複数観点から評価するルーブリックの質)をより高めていくことができるのではないかと考えた。

3. 研究の目的

高等学校国語科において、一般化可能性理論を適用して読解の質を複数観点から分析的に評価するルーブリック(以後分析的ルーブリックと表記する)を開発

し、ルーブリック開発過程における一般化可能性理論適用の有効性について仮説検証することを目的とする。

4. 研究仮説と検証方法

(1) 研究の仮説

高等学校国語科「読むこと」の領域においても、一般化可能性理論を適用して分散成分の推定値と一般化可能性係数を測定し、評価観点・評価水準・採点環境の改善点等に関する示唆を得るという過程を踏まえることで、学習者の読解の総合的な質を評価する分析的ルーブリックの質を向上させることができるだろう。

(2) 研究仮説の検証方法

まず稿者自身が「サッカーにおける『資本主義の精神』」の授業を実践し、読解の質を表出させる記述式の学習課題に学習者を取り組ませる。次に、その記述物の質を評価するルーブリックの観点と水準を妥当な理論的基盤等に依りつつ作成する。続いて、複数採点者がそのルーブリックに基づいて記述物をそれぞれに採点する。各採点結果を集めたのち、稿者が一般化可能性理論を適用して分散成分の推定値と一般化可能性係数を測定し、評価観点・評価水準・採点環境の改善点等に関する示唆を得る。各示唆をうけ、採点値について話し合う場の確保・評価水準の調整等の対応をとり、学習者の記述物の質の採点値に影響を及ぼしている採点者や評価観点等の要因等をより取り除くことができるようなルーブリック改案を作成する。その後再び分散成分の推定値と一般化可能性係数を測定し、分析的ルーブリックの質に変化があるかについて検証する。

5. 授業実践の概要

(1) 読解の質を表出させる学習課題

教材は「サッカーにおける『資本主義の精神』」という説明的文章である。実地研究期間は2017年6月1日から同年6月29日までであり、二週にわたる参与観察ののち、全五時間の授業を稿者自身が行った。授業実践の対象は広島県内の公立高等学校三年生計64名(A学級計33名、B学級計31名)である。単元の中には、教材研究に基づき「筆者の主張に対して反対意見を述べた指導者の意見文を踏まえた上で、自分の意見(賛成・反対・どちらでもない)を、理由とともにできるだけ具体的に考え、説得的に記述する。」という学習課題を設け、第五時の授業でその意見文記述活動に実際取り組ませた(註2)。なお、あらかじめ指導者の意見文を踏まえさせた理由は、学習者に意見文を書く際の手がかりを与え、かつ社会的な思考の

水準から意見を述べる文章のモデルを示すためである。第五時に至るまでの授業においては、オフサイド・ルールに関する読解、各段落に題名を付けて原著の各節名と比較する、筆者の主張は何か考える、筆者はサッカーが好きかそれともバスケットボールが好きかもしくはスポーツはそれほど好きではないか考える、といった各学習課題に段階的に取り組ませておいた。

(2) 妥当性のある分析的ルーブリック

続いて、読解の総合的な質を評価する分析的ルーブリック

の原案を作成した(表1)。山口(2015)が「信頼性を犠牲にしても、妥当性を重視しなければならないのが、統計学に限らず、評価をする際にも重要な点である。」(p.14)と述べるように、まずはこのルーブリック妥当性について説明を加えておく必要がある。

今回、学習者の読解の総合的な質を評価する観点として、「キーワードの理解」「テキストの意味内容についての具体的推論」「書き手が提示するテキスト世界における問題や書き手の世界認識の方法などに対す

表1 学習者の読解の総合的な質を評価する分析的ルーブリック(原案) (註3)

理論的基盤		記述物において表出する読解の質の諸側面		
		「②本文を問う」	「③テキスト世界を問う」	「⑤テキスト世界と現実世界の関係を問う」
評価観点		キーワードの理解	テキストの意味内容についての具体的推論	書き手が提示するテキスト世界における問題や書き手の世界認識の方法などに対する意見表明
評価水準	0	無解答	無解答、もしくは具体例を挙げようとしている記述が全くない	無解答、もしくは意見を表明していない
	1	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解していることが確認できない	「人生における区切りとリズムがなくなること」が具体的にどのような事態をもたらすのか、本文や指導者の意見文に書かれたままのことを想像している	「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えずに、個人的印象を根拠として意見文を記述しており、書き手と同様の社会的な思考の水準から意見を表明することができていない
		【該当する記述例】 一日の終わりがなくなり、眠る時間が少なくなってしまうのは嫌だ	【該当する記述例】 サッカー、病気の克服、サイボーグ化による生産効率化・仕事量の軽減、年末年始、誕生日、	【該当する記述例】 生命や生活から区切りがなくなれば人間らしさが私から失われていく気がするので、区切りの存在を受け容れたいと思う
	2	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解しているとわかるが、それら事象と「終わり」という言葉を結びつけて記述していない	「人生における区切りとリズムがなくなること」が具体的にどのような事態をもたらすのか、本文や指導者の意見文に明示的に書かれていないことを想像して、説明している	「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えた上で、それを根拠として意見文を記述しており、書き手と同様の社会的な思考の水準を踏まえて意見を表明することができている
		【該当する記述例】 毎日楽しいことばかりあると頑張ろうと思えなくなるかもしれない。それに、越えた先の楽しさはより小さくなってしまっているのではないと思う	【該当する記述例】 不老不死、サイボーグ化によって起こりうる未来の想像、区切りやリズムの重要性に対する自身の経験に基づく主観的判断	【該当する記述例】 延命治療技術を発展させると、確かに寿命は伸ばすことができるが、人間性を保つことは難しいだろう。結果として人類は人間性を失った機械的生活を送るだけになってしまい、きっと好ましくない
	3	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解しているとわかり、かつそれら事象と「終わり」という言葉を結びつけて記述している	「人生における区切りとリズムがなくなること」がもたらす具体的な事態ではなく、「終わりの反復と意味の消耗」や「反復によるエクスタシーの増大」がもたらす具体的な事態までを想像して、説明している	「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えた上で、「終わりとは～」「人生とは～」「生命とは～」といった書き手の簡潔な警句を模倣したり、書き手の世界認識の方法の特徴を評価したりするなどして、書き手と同等の思考の水準から意見を表明しようとしている
【該当する記述例】 つまり、終わりに向かう過程でつらいことがあればあるほどに、後からより大きな喜びや幸せのようなものを感じることができるのだと思う		【該当する記述例】 人生に彩りを与えるものとしての終わりや、自然現象や芸術品を美しいものたらしめるものとしての終わりの消失	【該当する記述例】 終わりとは人類が自身の歴史を反省するために必要不可欠なものなのだ。また筆者と指導者の間で「終わり」という言葉を解釈する前提としての人生観や科学観に異なる点がありその点で困難がある	

る意見表明」を設定した。観点設定に際して理論的基盤としたのは、PISA の読解力調査やヴァン・ダイクとキンチュの文章理解モデル等に基づき間瀬(2015)が提案する「読解モデル」(pp.225-226)である。この「読解モデル」のうち、課題において表出する「②本文を問う…本文に明示的な情報・内容を理解する力を問う」「③テキスト世界を問う…非明示的な内容を推論・解釈し、テキスト世界を理解する力を問う」「⑤テキスト世界と現実世界の間を問う…テキスト世界における問題と読み手が存在する現実における問題を関連させて考える力を問う」という読解中の認知活動の諸側面の質を、観点を分けて分析的に評価することによって学習者の読解の総合的な質を評価することができるかと稿者は考えた。そこで、「読解モデル」の各観点を教材と課題に適合するように具体的に書き換え、上記3観点として設定した。さらに、学習者の実際の記述物62枚をサンプルとして、それ以外に考えられる観点はるか、各観点の評価水準や記述例をどう設定すればよいか、などについて話し合い検討する場を複数回設けた。この検討活動に携わったのは、高等学校国語科の教員を志望する大学院生(博士課程在籍)1名、大学院生(専門職修士課程在籍=稿者)1名、学部生1名の計3名である。3名は高等学校国語科の教員を志望し、かつ国語科教育学の評価領域について専門的に研究しているため、学習者の記述物を評価する力を十分もっている人材だと考えられた。この検討活動で合意を形成した後、稿者が各評価水準と記述例とを書き込んだ。こうした作成手順を踏まえ、学習者の読解の総合的な質を評価するにあたり妥当性がより高いと思われる分析的ルーブリックを完成させた。

6. 一般化可能性理論の適用と調整

(1) 一般化可能性研究の結果と考察

授業実践後、学習者の記述物(有効回答数=62)を

回収して採点活動を行った。採点を担当したのは、大学院生(博士課程在籍)1名、大学院生(専門職修士課程在籍=稿者)1名、大学院生(修士課程在籍)3名、学部生1名の計6名である(うち3名はルーブリック作成活動にも参加)。6名は高等学校国語科の教員を志望し、かつ全員国語科教育学分野について専門的に研究しているため、学習者の記述物を評価する力を十分もっている人材だと考えられた。各採点者は、配付されたルーブリック(表1)に基づき、相互に相談することなく採点活動を一時間程度かけて完了させた。ここで得られた採点データは、62名の学習者(persons, p)が書いた記述物を3つの評価観点(tasks, t)から6名の採点者(raters, r)が評価する三相データである。この採点値をもとに、SPSS Statistics24に組み込まれているVARCOMPを使用して分散成分の推定を行った。分析デザインは評価観点と採点者を変量モデルとした二相完全クロス計画である。分析の結果、各要因の分散成分の推定値、全分散に対する各分散成分の割合、一般化可能性係数の数値を得た(表2)。

一般化可能性係数は0.64であった。信頼性係数は慣習的に0.80以上で十分とみなされるので、評価の信頼性は不十分であった。ゆえに、ルーブリック原案による分析的評価では、学習者の読解の総合的な質の真値を評価できていない蓋然性が高いと考えられる。よって、得られた評価データは総括的評価として評定やアカウンタビリティに活かしにくいだけでなく、形成的評価として授業改善にも活かしにくいものだと解釈する。続いて、各分散成分の推定値等から一般化可能性係数の低さの原因と改善点について考察していく。

第一に、学習者の主効果を示す要因 p の全分散に対する割合が21.29%であることに注目する。この数値はあまり高いものではなく、学習者の能力による得点のばらつき少なかったことを示す。このようなことから、ルーブリック原案は学習者の能力の個人差を十分とらえることができているとは言い難いと解釈できる。

表2 各分散成分の推定値と一般化可能性係数(一般化可能性理論適用前)

変動要因	分散成分の推定値	全分散に対する割合(%)
学習者 (p)	0.109	21.29%
評価観点 (t)	0.003	0.58%
採点者 (r)	0.017	3.32%
学習者×評価観点 (pt)	0.129	25.20%
学習者×採点者 (pr)	0.035	6.84%
評価観点×採点者 (tr)	0.016	3.13%
学習者×評価観点×採点者 (ptr)	0.203	39.65%
一般化可能性係数	0.64	

よって、学習者個々人の能力差をより正確に捉えられるように評価水準をより細分化したり拡張したりする必要があると示唆された。第二に、一次交互作用要因である *pt* の全分散に対する割合が 25.20% であることに注目する。この数値は比較的高いものであり、学習者が各観点によって得手不得手があったことを示す。このようなことから、複数観点からの分析的評価が機能していることを示すが、一方で観点によって学習者の能力差が大きくあられすぎているとも解釈できる。よって、各評価観点の水準設定をもう少し均一化する必要があると示唆された。第三に、一時交互作用要因 *tr* の全分散に対する割合が 3.13% であるのに対し、一次交互作用要因 *pr* の全分散に対する割合が 6.84% であることに注目する。前者の数値は高いものとはいえず、採点者それぞれの評価観点の捉え方には一貫性があったことを示す。それに対し後者の数値は比較的高いものであり、各採点者によって学習者の順位に変動が生じやすかったことを示す。このようなことから、採点者個々人は各評価観点や水準の内容を比較的良好に捉え一貫した採点を行っているが、各採点者間でそれら理解の仕方にはずれがあったと解釈できる。つまり、ループリックの観点や水準の設定よりも、むしろその評価言のわかりやすさに問題があったということである。よって、採点値に差が出た記述物をサンプルとして話し合うなどしてループリックの評価言をより充実させ、観点や水準の理解を相互に深める必要があると示唆された。第四に、二次交互作用要因 *ptr* の全分散に対する割合が 39.65% であることに注目する。これは今回の分散成分の推定値の割合のうち最も高く、それぞれの主効果やお互いの交互作用で説明できない要因が大きかったことを示す。このようなことから、採点活動を実施した際に、測定誤差を生むような何らかの要因が介在していたと解釈することもできる。ループ

リックの評価言をより誤解を生みにくいものに洗練するだけではなく、採点活動を実施する環境を改善するなど、不必要なばらつきを減らす努力をより包括的にする必要があると示唆された。以上述べてきたように、学習者の読解の総合的な質の真値を評価するためには評価水準・評価言・採点者・採点環境という様々な要因に改善を加える必要があると示唆された。

(2) 決定研究の結果と考察

続いて上記一般化可能性研究 (Generalizability study) で得られた分散成分の推定値を利用して、評価観点数と採点者数を変化させたときの一般化可能性係数の変化をシミュレーションした。片方の数値はループリック原案による評価活動時と同じ条件に設定した (評価観点数: 3 もしくは採点者数: 6) 上で、評価観点数と採点者数の変化量を 1 から 10 までの間に設定した。この決定研究 (Decision study) の分析結果を図 1 に示す。なお当然ながら、評価観点数を 3 観点でシミュレーションした時と、採点者数を 6 人でシミュレーションした時の一般化可能性係数はどちらも表 2 下部で示した数値と同じ値である 0.64 を示す。

仮に採点者が 1 人であった場合、一般化可能性係数は 0.43 にすぎないと算出された。先行研究において、ループリックを用いるだけで信頼性が保証されるとは限らないとされているが (松下ら, 2013; 斎藤, 2016)、このシミュレーション結果から同様のことが考察できる。また、採点者数に関するシミュレーショングラフと評価観点数に関するシミュレーショングラフでは、後者が傾きが大きいと算出された。このようなことから、採点者数を増やすよりもむしろ評価観点数を増やしたほうが上記ループリックの信頼性を高めるために効果的だと考えられた。ただ、目標とする一般化可能性係数の数値 (0.80 以上) に到達するためには評価観点を 5 つ以上も追加しなければならないと算出されて

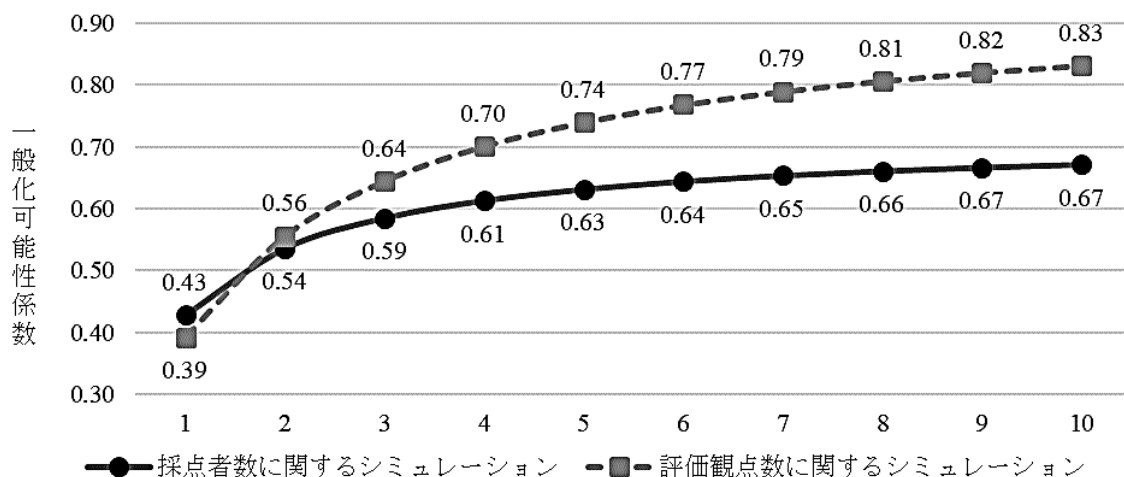


図 1 一般化可能性係数のシミュレーション結果

おり、ルーブリックを改善して評価の質を保証するためには採点者数や評価観点数を単純に増加させるだけでは難しく、評価水準・評価言・採点者・採点環境の質すべてを向上させる必要があると示唆された。一般化可能性理論を適用することによって得たこれら分析結果と示唆等をうけ、新たに付け加えられそうな評価観点の模索、評価言や記述例の適切化、評価水準の細分化と均一化等を話題の中心としたルーブリック調整活動の場を組織することで、分析的ルーブリックの質を高めることと試みることにした。なお、採点者数を増加させることについては、松下ら(2013)が「評価の実行可能性(feasibility)を低下させ、現実的ではない」(p.114)と述べていたことを踏まえ、見送ることとした。学校現場において、学習者の記述物を評価するにふさわしい人材を多数確保するのは難しいと思われるため、本研究ではあくまで限られた人材資源の中で分析的ルーブリックの質を高める方法について検討していく。

(3) ルーブリックの調整

一般化可能性理論を適用することで得た示唆をもとに、ルーブリックを調整した(表3)。このルーブリック調整活動に参加したのは、採点活動にも参加した高等学校国語科の教員を志望する大学院生(博士課程在籍)1名、大学院生(専門職修士課程在籍=稿者)1名、学部生1名の計3名である。採点者は、各自が採点した記述物62枚をサンプルとして、ルーブリックの改善点について話し合った。この際話題の中心となったのは、なぜこの学習者の記述物の採点値がこなにも採点者によって異なっているのか、新たに付け加えられそうな評価観点はあるか、評価言や記述例をよりわかりやすくするにはどうすればよいか、評価水準をより細分化するにはどうすればよいかといった、上記分析によっても示唆されていた論点であった。

次にルーブリックの主な調整点について説明する。第一に、読解における情意的側面を評価する観点として「主体的な意見表明」を付け加えた。各採点者が採点後に抱いた実感と、教育評価研究領域における石井(2015)の「思考する意欲を育てる上では、自分なりにこだわりをもって考えたことが評価される余地を残しておくことが重要です」(p.64)という指摘、そして国語科研究領域における山元(2014)の「『PISA型読解力』の基本は『喜びを味わう読み』=『読むことへの取り組み(reading engagement)』である」(p.64)という指摘や、松尾(2016)の「リテラシー概念は、最低限の読み書き能力から高次の情報処理能力へと拡張され、さらに、情意を含む人間の全体的能力を視野に入れるコンピテンシーへの展開がみられた」(p.163)という指摘等から、読解活動における学習者の内的思考過

程の総合的な質を捉えるためには認知的側面だけでなく情意的側面の質にも注目する必要があると考えたため、学習者の「主体性」を記述量という側面から評価するこの観点を新たに設けた。なおこの観点は、課題の難度と学習者の実態を踏まえ、かつ他の観点と水準設定がより均一になるように注意し、稿者が適切な水準を設定した。第二に、採点者間で採点値が比較的ばらついていた観点「書き手が提示するテキスト世界における問題や書き手の世界認識の方法などに対する意見表明」について共通理解を深め、各水準の評価言と記述例を全採点者が納得できるような表現に改めた。第三に、「全学習者の記述物を各観点から一気に採点しない」「集中できる環境を整えてから行う」といった採点に関わる注意事項の指示を徹底することとした。

7. 研究仮説の検証結果と考察

(1) 一般化可能性係数の値の比較

表3で示したルーブリック改案を用いて再度学習者の記述物(有効回答数=62)を採点し、前節第一項と同様の手順で各分散成分の推定値、全要因に対する各要因の分散成分の割合、一般化可能性係数の数値を得た(表4)。採点を担当したのは、前回の採点活動に参加していた6名と全く同じ人物である。なお、評価活動そのものの質を比較するための統制された条件を整えるため、二ヶ月以上の期間を空けて今回の採点活動を実施している。各採点者6名は、「全学習者の記述物を各観点から一気に採点するのではなく、観点をひとつに絞ってから全学習者の記述物を採点するようにしてください」「集中できる環境を整えてから行う。また、採点活動を行う際途中で30分以上の休憩をはさまないでください」という指示のもと、稿者が配付したルーブリック(表3)に基づき、相互に相談することなく採点活動を1時間程度かけて完了させた。

まず表2と表4に示した一般化可能性係数の値を比較し、数値の変化について考察を加える。ルーブリック改案による評価の一般化可能性係数は0.78であった。信頼性係数は慣習的に0.80以上で十分とみなされるので、今回の評価の信頼性は十分でない。ただ、本研究で採点した記述物62枚は全て同一高等学校に在籍する学習者によるものがあるため、学習者間の能力差を示す要因 p の値、そして p の値に影響を受ける一般化可能性係数もあまり高い数値を示さないと考えられた。同様の考察結果を導く先行研究はいくつかある。例えば佐々木ら(2006)は、東北大学の学生20名の口頭発表を3名で評価したところ一般化可能性係数が0.6未満だったと報告している。また松下ら(2013)も、

表3 学習者の読解の総合的な質を評価する分析的ルーブリック（改案）

評価観点	キーワードの理解	テキストの意味内容についての具体的推論	書き手が提示するテキスト世界における問題や書き手の世界認識の方法などに対する意見表明	主体的な意見表明	
評価水準	0	無解答	無解答、もしくは具体例を全く挙げようとしていない	無解答、もしくは全く意見を表明していない	主体的に意見を表明し、十分に説得力ある量の記述を行っていない（解答欄4行未満）
	1	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解していることが確認できない	「人生における区切りとリズムがなくなること」が具体的にどのような事態をもたらすのか、本文や指導者の意見文にほぼ書かれた通りのことを想像している	テキスト世界における「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えずに、個人的印象を根拠として意見文を記述しており、書き手の提示する社会学的な思考の水準から意見を表明することができていない	主体的に意見を表明して、十分説得力ある量の記述を行った（「なぜなら」の行を含む4行～9行の解答欄が完全に埋められている）
		【該当する記述例】 一日の終わりがなくなり、眠る時間が少なくなってしまうのは嫌だ／ <u>科学技術が進歩して便利になるのはいい</u>	【該当する記述例】 サッカー、サイボーグ化による生産効率化・仕事量の軽減、年末年始、誕生日、身内の病氣や死、医療技術	【該当する記述例】 生命や生活から区切りがなくなれば人間らしさというものが私から失われてしまう気がするので、区切りの存在を受け容れたい／ <u>毎日終わりをめざして仕事を頑張っている人もいると思うから、反対です</u>	
	2	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解しているとわかるが、それら事象と「終わり」という言葉を結びつけて記述していない	「人生における区切りとリズムがなくなること」が具体的にどのような事態をもたらすのか、本文や指導者の意見文に明示的に書かれていないことを想像して、説明している	テキスト世界における「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えた上で、それを根拠として意見文を記述しており、書き手の提示する社会学的な思考の水準を踏まえて意見を表明することができている	より主体的に意見を表明して、より説得力ある量の記述を行った（「なぜなら」の行を含む以降の9行以上の解答欄が完全に埋められている）
		【該当する記述例】 毎日楽しいことばかりがあると頑張ろうと思えなくなるかもしれない。それに、越えた先の楽しさはより小さくなってしまふ。／ <u>一日や一年が終わる瞬間に向け生きているから、最後が楽しい。</u>	【該当する記述例】 不老不死、科学技術発展とサイボーグ化によって起こる未来世界の想像、 <u>AI、人間と文化や歴史、テレビゲームのゲームオーバー、始業のチャイム、受験勉強</u>	【該当する記述例】 延命治療技術を発展させると、確かに寿命は伸ばすことができるが、人間性を保つことは難しいだろう。結果として人類は人間性を失った機械的生活を送るだけになってしまうので、好ましくない／ <u>終わりとしての意味を消耗させようと努力することで、現代社会の科学技術をより発展させることができる。</u>	
	3	「終わり」というキーワードの特徴である「反復と意味の消耗」または「反復によるエクスタシーの増大」という事象を理解しているとわかり、かつそれら事象と「終わり」という言葉を結びつけて記述している	「人生における区切りとリズムがなくなること」がもたらす具体的な事態ではなく、「終わりの反復と意味の消耗」や「反復によるエクスタシーの増大」がもたらす具体的な事態までを想像して、説明している	テキスト世界における「人生における区切りとリズムがなくなる」「終わりの反復と意味の消耗」「エクスタシーの増大」という諸問題が人間社会全体に与える影響を考えた上で、書き手の世界認識にはそぐわない反証事例を警句や故事成語を用いて説明したり、書き手の世界認識の方法の特徴を評価したりするなどして、書き手を批評するような思考の水準から意見を表明している	
【該当する記述例】 つまり、終わりに向かう過程でつらいことがあればあるほど、後からより大きな喜びや幸せを感じることができるのだ／ <u>終わりがあるからこそ、それまでの過程を美しく感じ、大切にできる。</u>	【該当する記述例】 人生により大きな彩りを与えるものとしての終わりを象徴する具体的事例（単なる主観的判断は含まない）や、自然現象や芸術作品を美しいものたらしめるものとしての終わりの性質	【該当する記述例】 <u>人類が科学技術の発展を実現できたのは、終わりに伴う失敗経験と反省があったからだ。人類は、「終わりとしての意味」をもっていたから今がある。</u> ／ <u>終わりとは人類が自身の歴史を反省するために必要不可欠なものなのだ。また筆者と指導者の間で「終わり」という言葉を解釈する前提としての人生観や科学観に異なるところがありその点で困難がある。</u>			

（下線部はルーブリック原案に調整を加えた主要な部分を示す）

表 4 各分散成分の推定値と一般化可能性係数（一般化可能性理論適用後）

変動要因	分散成分の推定値	全分散に対する割合 (%)
学習者 (p)	0.130	27.03%
評価観点 (d)	0.012	0.03%
採点者 (r)	0.000	0.00%
学習者×評価観点 (pd)	0.188	39.09%
学習者×採点者 (pr)	0.000	0.00%
評価観点×採点者 (tr)	0.004	0.01%
学習者×評価観点×採点者 (ptr)	0.147	30.56%
一般化可能性係数	0.78	

新潟大学の学生 64 名のレポートを 3 名で評価したところ一般化可能性係数が 0.62 であったと報告し、その上で「今回の係数の値はあながち低いものとはいえ、先行研究とほぼ同等の結果を得られたものとみることでもできる」(p.114)と述べている。このように、対象学習者が同じような学力層に位置することを考慮するならば、0.78 という数値をある程度信頼できるものとみなすこともできるだろう。いずれにせよ、一般化可能性理論を適用してループリックを調整したことで一般化可能性係数が 0.64 から 0.78 に向上したことは事実である。ループリック原案 (表 1) は、一般化可能性理論を適用した調整によって、より信頼性が高いループリック改案 (表 3) へと改善されている。よって稿者は、一般化可能性係数の値の変化について考察した結果から、研究仮説の正しさが支持されたと考える。

(2) 各分散成分の推定値の比較

続いて、表 2 と表 4 に示した各分散成分の推定値を比較し、数値の変化について考察を加える。

第一に、学習者の主効果を示す要因 p の全分散に対する割合が 21.29% から 27.03% に増加していることに注目する。これは、ループリックの評価言と記述例をわかりやすくしたことにより第 1 水準や第 3 水準に該当する記述物が増え、採点値の中央化傾向が減り、学習者の読解の質の差を捉えることができたことの証拠だと考えることができる。また、採点者の主効果を示す要因 r と一次交互作用要因 pr の全分散に対する割合がそれぞれ 3.32% から 0.00% と 6.84% から 0.00% とに減少していることに注目する。これは、ループリックをわかりやすく改善したことや話し合い活動を設けたことにより、各採点者が評価観点と水準について共通理解を深め、採点者の主観や個人差という要因による影響を小さくすることができたことの証拠だと考えることができる。続いて、二次交互作用要因 ptr の全分散に対する割合が 39.65% から 30.56% に減少してい

ることに注目する。これは、採点環境に関する指示の徹底したことや、評価言と記述例をわかりやすく書き換えたことなどにより、学習者の読解の質を評価する上で不必要な要因を取り除くことができたことの証拠だと考えることができる。なお、一次交互作用要因 pt は分散成分の推定値とその割合ともに増加しているが、これはむしろ「主体的な意見表明」という観点を加えたことにより学習者の総合的な読みの質をより多面的に捉えることができた証拠だとして肯定的に捉えることができるだろう。以上、分散成分の推定値とその全分散に対する割合の値の変化についての考察結果からも、研究仮説の正しさが支持されたと稿者は考える。

8. 成果と課題

本研究の成果と課題についてまとめる。第一の成果は、より信頼性の高い学習者の読解の質を評価する分析的ループリックを開発する過程で、一般化可能性理論を適用することの有効性を示したということである。もちろん、評価観点や水準に関する話し合いの場を設けたり学習者の記述物をサンプルとして採点値を擦り合わせたりすることでループリックの質をより高めていくというこの基本的な方法自体は、以前から行われてきたものと変わらない。ただこの一般化可能性理論を用いた方法は、そうしたループリック開発の過程を可視化することができるという点でより優れていると考えられる。分散成分の推定値等を求める一般化可能性研究 (Generalizability study) と変数の増減をシミュレーションする決定研究 (Decision study) を組み合わせることで、ループリック改善活動において「次にどの部分を変えるべきなのか」「どの程度調整を加えればいいのか」といったことを視覚的に捉えやすくなる。ゆえに、一般化可能性理論を用いた方法を採用することによって、質の高い分析的ループリックをよ

りの確にかつより少ない時間で開発していくことができると稿者は考える。ただし、本研究ではその的確さや時間の少なさについて詳細に比較検討したわけではない。この点が第一の課題である。また、今回は数ある心理測定学や統計学の知見の中でも、特に一般化可能性理論に基づくルーブリック改善に限って研究を行ったが、今後は多値型項目反応理論などの知見を取り入れたルーブリック改善法の有効性についても検討する必要がある。この点が第二の課題である。

第二の成果は、「サッカーにおける『資本主義の精神』」において、学習者の読解の総合的な質を評価する分析的ルーブリックを開発できたことである。遠藤(2015)の「最近の評価研究においては、信頼性は妥当性の必要条件になるとされている。つまり、ある程度の信頼性が確保されていないと、妥当性も確保されないのである。」(p.115)という指摘を踏まえれば、評価の信頼性が保証されたこのルーブリック改案は、妥当性が担保されたルーブリックだと言うこともできるだろう。従来、心理測定尺度等に基づく質問紙法においてしばしば信頼性係数が検証されてきたのに対し、ルーブリックに基づく総合的なパフォーマンス評価においてその信頼性が数値的に検証されることは少なかったと思われる。しかし、「どの学級が最も読みの質が高かったか」「昨年の学習者と比べてどちらの読みの質が高かったか」などについて比較検討するための尺度として機能できるのは、妥当性と信頼性が保証されたルーブリックだけである。今回、一般化可能性理論を用いて、質が保証された分析的ルーブリックを開発したからこそ、「サッカーにおける『資本主義の精神』」の授業をより客観的に改善していくことが可能になったと稿者は考える。もちろん、ここで稿者は信頼性係数の測定を行わないような評価の方法をなにかも否定したいというわけではなく、むしろ各教師の主観的観察等を最も重要な反省と授業改善の方法と考えながらも^(註4)、客観的なデータを同時に参照することの重要性を指摘したということである。今後はこのルーブリックを授業改善に生かす具体的な筋道を明らかにしたり、神部(2017)で考察した「学生参加型ルーブリック」で活用していく方法を探るなどしたい。

9. 註

(1) しかしこの手法にも弱点があるとされる。分寺(2016)は「一般化可能性理論では評定者効果の分散を検討対象とし、評定者のパラメータ化を行わないため個人単位で評価するという目的には向いていない。」(p.183)と述べ、宇佐美(2010)も「採点者の個人差を明

確にし、問題となる特定の採点者や項目およびその組み合わせを検討するには不向きである。」(p.164)と述べている。評価の質を高めていくことを目的とする場合、項目反応理論の考え方を組み合わせていくことが重要である。しかしそうした研究は別稿に譲りたい。

(2) 学習者の記述には次のようなものがあつた。「私は指導者の『区切りがないほうが人生を楽しむことができる』という主張に対して反対という立場をとる。なぜなら、区切りこそが人生を作り上げ、彩っているものだと思うからだ。(中略=稿者)要するに、『区切り』とは終わりのみならず物事の始まりでもあるのだ。明日のために今日を終わって仕切り直すという区切りがあるから人生は彩りをもっている。最初に述べた通り、私は指導者の意見には反対だ。エンドロールまでを如何に盛り上げるかを考え抜かれているからこそ映画は面白くあれる。終わりがあからこそ人類はこれまでの発展を成し遂げてきたのだ。『終わりとしての意味』さえも失い、人類が区切りを喪失したとすれば、残るのは穏やかな、そして確実な衰退だろう。」

(3) ここで示したルーブリックは学習指導案の目標と対応するものでもある。なお、このようなルーブリックはあらかじめ学習者に提示しておくことが望ましいが、研究の手順上行うことはできなかった。

(4) 松友(2013)は「教師の『見取り』と呼ばれる観察法による主観的な評価が必要である。こうした教師の観察眼は、指導経験や学力に関する知見によって構成されているが、これを育成する研修も数多くなされ、流動的な学習過程の中で、複数の対象に対して逐次的に観察、評価するというハイ・パフォーマンスに耐えうる実践力と簡素化されてはいるが焦点化された観察の観点を教師に獲得させる事は今後の研究に求められている。」(p.243)と述べているが、一般化可能性理論を適用して得た客観的なデータを参照しつつ評価の改善策等について話し合うことは、そうした教師の観察眼を鍛える上でもなんらかの効果があると思われる。しかしこの点についても検証しておらず課題である。

10. 参考文献

- 池田央(1994)『現代テスト理論』朝倉書店
石井英真(2015)『今求められる学力と学びとは—コンピテンシー・ベースのカリキュラムの光と影—』日本標準
石原千亜紀・山田剛史(2008)「中学生の作文評価のための基準の信頼性と妥当性—分析的項目と総合項目による評価の検討」『兵庫教育大学大学院連合学校教育学研究科 教育実践学論集』,10,pp.15-26

- 宇佐美慧(2010)「採点者側と受験者側のバイアス要因の影響を同時に評価する多値型項目反応モデル—MCMC アルゴリズムに基づく推定」『教育心理学研究』,58,pp.163-178
- 遠藤貴弘(2015)「学力評価の方法」西岡加名恵・石井英真・田中耕治編『新しい教育評価入門—人を育てる評価のために Introduction to Educational Assessment and Evaluation』有斐閣,pp.113-142
- 神部智(2017)「評価基準の内面化を促す国語科学習指導に関する一考察—『学生参加型ルーブリック』の有効性を仮説検証することを通して—」『論叢国語教育学』,13,pp.16-26
- 斎藤有吾(2016)「パフォーマンス評価における項目反応理論を利用したアカデミック・ライティング力の測定」『京都大学大学院教育学研究科紀要』,62,pp.427-439
- 佐々木典彰・村木英治(2005)「口頭発表の評価における信頼性—一般化可能性理論を用いて—」『教育情報学研究』,3,pp.1-4
- 佐藤洋一・岡田智(2010)「小説教材における『習得・活用』の授業・評価開発—村上春樹『青が消える』(高校1年・明治書院)を例に」『愛知教育大学教育実践総合センター紀要』,13,pp.101-110
- 田中洋美(2016)「主体的・協働的に学ぶ言語活動の日常化を目指して—高1国語総合『新書を読もう・伝えよう』実践報告」『椋山女学園大学教育学部紀要』,9,pp.147-166
- 中央教育審議会(2014)「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について(答申)」高等教育局高等教育企画課高等教育政策室
- 平井洋子(2007)「主観的評定における評定基準、評定者数、課題数の効果について—一般化可能性理論による定量的研究」『人文学法』,380,pp.25-64
- 二田貴弘(2009)「メディア・リテラシーを援用して国語科における『読解力』『表現力』を育成する授業の方法—授業の工夫と評価を組み合わせる—」『奈良女子大学研究紀要』,49,pp.39-48
- 分寺杏介(2016)「項目間分散を考慮した項目反応モデルの提案—小論文評定への応用—」『行動計量学』,43(2),pp.181-195
- 間瀬茂夫(2015)「高等学校における高次読解力の評価のあり方—読解力評価問題の活用—」『国語教育研究』,56,pp.219-230
- 松尾知明(2016)「知識社会とコンピテンシー概念を考える—OECD 国際教育指標 (INES) 事業における理論的展開を中心に—」『教育学研究』,83(2),pp.154-166
- 松下佳代・小野和宏・高橋雄介(2013)「レポート評価におけるルーブリックの開発とその信頼性の検討」『大学教育学会誌』,35(1),pp.107-115
- 松友一雄(2013)「読むことの学習に関する評価研究」全国大学国語教育学会編『国語科教育の成果と展望 II』学芸図書,pp.241-248
- 森重里穂・渡部倫子(2017)「一般化可能性理論を用いた外国人児童生徒のための JSL 対話型アセスメント DLA (話す) の検討」『広島大学日本語教育研究』,31,pp.31-34
- 山口陽弘(2015)「今、教育関係者に求められる評価リテラシーとは」『教育目標・評価学会紀要』,25,pp.11-18
- 山口陽弘・清水真紀(2009)「英語学習者のための音読テストの信頼性の検討—一般化可能性理論の適用と第二言語習得への示唆」『群馬大学教育学部紀要 人文・社会科学編』,58,pp.155-168
- 山下洋史・大野高裕・尾関守(1990)「人事考課における寛大化傾向・中央化傾向・厳格化傾向の定量的分析」『日本経営工学会誌』,41,5,pp.336-341
- 山西博之(2005)「一般化可能性理論を用いた高校生の自由英作文評価の検討」『JALT Journal』,27(2),pp.169-185
- 山元隆春(2014)『読者反応を核とした「読解力」育成の足場づくり』溪水社
- 渡部洋・平由実子・井上俊哉(1989)「小論文評価データの解析」『東京大学教育学部紀要』,28,pp.143-164

11. 付記

本稿は、神部による第 69 回中国四国教育学会での自由研究発表(2017年11月26日 於広島女学院大学)「読みの深さを評価するルーブリックの開発と一般化可能性理論による信頼性の検討—『サッカーにおける「資本主義の精神」』授業実践から—」の当日発表用資料に加筆修正を行ったものである。発表後の質疑応答の場や総括討議の場において、多くの方から貴重なご意見をいただきました。心より感謝いたします。

12. 謝辞

本研究を進めるにあたり、本学の宮里智恵教授に貴重なご指導を賜りました。心より感謝致します。また授業の実施に関してご指導していただきました広島県内の高等学校の松本誠司教諭に深く御礼申し上げます。