

# 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－

石原 千亜紀\*, 山田 剛 史\*\*

(平成20年6月20日受付, 平成20年12月4日受理)

## Reliability and validity of the criteria to evaluate junior high school students' compositions: A Study of evaluation by using evaluation items and general items

ISHIHARA Chiaki\*, YAMADA Tsuyoshi\*\*

The purpose of this study is to make the criteria to evaluate the composition of junior high school students, and to examine the reliability and validity. Japanese class teachers and university students rated junior high school students' compositions by using 17 evaluation items and general items (2 items). On the basis of the attitudes toward compositions, they were each divided into 2 levels : the high favorable and the low favorable. As a result, the high favorable teachers' coefficient is greater than that of low favorable teachers. Generalizability coefficient of high favorable groups is greater than low favorable groups. Rating aspect is different by group from ANOVA estimates. It was considered that, in high favorable groups, the evaluation aspect is similar by group.

Key Words : composition evaluation, reliability, validity, generalizability theory

### I 問題と目的

#### I-1 問題

中学校学習指導要領の国語科 (1998)<sup>(1)</sup>の「書くこと」の目標に、第1学年は、「必要な材料を基にして自分の考えをまとめ、的確に書き表す能力を高めるとともに、進んで書き表そうとする態度を育てる。」、第2学年及び第3学年においては、「様々な材料を基にして自分の考えを深め、自分の立場を明らかにして、論理的に書き表す能力を身に付けさせるとともに、文章を書くことによって生活を豊かにしようとする態度を育てる。」とある。このように、生徒が主体的に論理的に自らの意見を書けるようになることが、中学校国語科における目標として示されているのである。

2003年に実施されたPISA調査の結果では、他の国と比べて顕著な問題点として「自由記述問題の無回答率が際立って高いこと」、特に「自分の意見とその理由・根拠を求められるような問題を不得意としている」が挙げられている(国立教育政策研究所, 2007)<sup>(2)</sup>。さらに、国立教育政策研究所によって行われた教育課程実施状況調査について言及する。この調査では、子どもの学力の現状について、把握・分析が行われた。最近では、平成5～7年度、平成13年度、平成15年度に調査が実施された。調査の結果、学力の経年変化について、国語科においては

記述式の問題の正答率が低下し、中でも文章を深く読んで分析的に理解してその上で論理的に記述する設問では正答率が低下していることが報告されている(中央教育審議会, 2008)<sup>(3)</sup>。以上の結果より、書くことへの苦手意識をもつ生徒が多いことが示唆され、中学校国語科で目標とされている「書く態度」が十分に養われているとはいえない実態があることがわかる。

「自分の考えを自分の言葉で書くこと」は、学習や生活の全てにおいて基本となり、生涯にわたり必要となることである(位藤, 2000)<sup>(4)</sup>。目標の書く態度を養うには、「指導のねらいを明確にし、学校や生徒の実態を踏まえ、また、知識的なことの指導に時間をかけすぎることなく、生徒の書く機会を多くし、書く時間を十分に確保した指導計画を作成することが大切である」(文部科学省, 1998)<sup>(5)</sup>。また、作文指導については、「特に指導者の大きな指導力が要求されるものであること、事後処理に多くの時間を要するものであること」(森本・長谷川, 1980)<sup>(6)</sup>といった見方がなされている。それにもかかわらず、書くことの指導が行われる機会は、「運動会や学芸会、あるいは遠足などのたび」と限られた機会ではないかと指摘されている(菅原, 2000)<sup>(6)</sup>。また、このように書くことの指導機会が限られているだけでなく、書かれた作文の評価についても課題が指摘されている。特に作文の評価が入試の

\*岡山大学大学院教育学研究科学生 (Master program student of Graduate School of Education, Okayama University)

\*\*岡山大学 (Okayama University)

ようなハイスティクスな評価に利用される場面を例に、岩間（2004）<sup>(7)</sup>は「評定が高校入試の資料として使われ進路決定の資料となることを考慮すると、生徒や保護者が納得できるような適切な方法で評定を出さなくてはならない」と述べている。一方、その評価のあり方について、高木（2002）<sup>(8)</sup>は、「論文試験における点数化も、点数として表されたものは、序列をつけるという意味で客観的ではあるが、その評価場面では、評価者の主観が多く入り込んでいる」としている。目標をどの程度達成しているか評価する為、評価基準を準備した作文の評価においても、評価者の主観が入る可能性を完全に排除することは困難である。この評定者の主観を最小限にすることは、作文評価の大きな課題である。

その上、作文を書く際に必要とされる作文力（文章表現力）について、「認識力、思考力、感受性といった、いわば全人間的な力を考える人たちと、その一方、もっと具体的、言語的な文章表現力を考える人たちとの二派の人々が存在する」（井上、1981）<sup>(9)</sup>。このように作文力に対しての考え方も評価者により異なることが考えられ、現場では明確な評価の基準が決められないまま個々の教師に評価が任されている現状がある。

そこで、本研究では「書くこと」の評価に着目する。具体的には、中学生の作文評価のための基準を作成し、その妥当性と信頼性について検討を行う。また、作文に対する好意度と、教師・学生という評定者の違いがどのように評価に影響を及ぼすのかを明らかにする。

## I-2 教育心理学領域における作文評価研究

ここでは、教育心理学領域において行われた作文評価に関する研究を整理する。なお、作文評価に関する研究は、例えば、国語科教育の領域でも行われている。しかし、本稿では、教育心理学的手法を用いて行われた研究を対象とし、国語科教育における作文評価は取り上げないこととする。

梶井（2001）<sup>(10)</sup>は、小学校学習指導要領（1989）に基づいた基準を作成し、小学生の生活文の評価の信頼性と妥当性の検討を行っている。小学校教員21名を対象者として、作文に対する好意度の高い群（高群）と低い群（低群）に分けた。対象者に1年生から6年生までの児童が書いた作品を、1～2年生を低学年、3～4年生を中学年、5～6年生を高学年とした学年区分のもと、学習指導要領（1989）を反映させた18の分析的評価項目と、「うまさ」と「好み」の2つの総合評価項目とで評定させている。どのような要素が作文評価に影響を与えるのかを知るために、総合評価項目に及ぼす影響力が大きい項目を抽出し（妥当性に関する検討）、評定者間で評定結果が一致する傾向について調べている（信頼性に関する検討）。そして好意度の違いが項目の妥当性・信頼性に及ぼす影

響を検討し、評価に影響を及ぼす要因として作文×評定者×項目の3要因の効果を明らかにした。項目の妥当性・信頼性の検討では、多くの項目が「要検討の項目」となり、その傾向は低、中学年で強いとした。さらに中学年の特徴として、「使用の難しい項目」の数が多いこと、 $\alpha$ 値が極端に低い項目の数が多いこと、評価の信頼性（G係数の値）が他の学年に比べ低いこと、総合評価項目の「うまさ」と「好み」とが一致しにくい傾向が強いことを示した。これより、中学年の作文評価が難しく、その上好意度により難しさの中身が異なると述べている。

教員の評価によっても、児童の作文を的確に評価するのは困難であるという梶井（2001）<sup>(10)</sup>の結果を受けて、さらに梶井（2002）<sup>(11)</sup>は、将来の教員を目指す教員養成課程の大学生の評価の実態を明らかにすることによって、教師の作文評価能力の成長過程と評価を困難にする原因を明確にしようとした。そこで、教員養成課程の学生29名に調査を行った。これを好意度及び指導経験により4群に分け、その中の好意度が高くかつ指導経験がある群を高群、好意度が低く指導経験のない群を低群とした。評価基準は梶井（2001）<sup>(10)</sup>で用いられた分析的評価項目18項目を用い、小学生の生活文の評価を行わせ、梶井（2001）<sup>(10)</sup>と同様の分析を行った。この調査から、高群の評価は、「使用できる項目」は検出されず、「使用の難しい項目」が10項目検出された。このことから、分析的評価項目を評価に役立てるのは困難であったと述べられている。さらに調査の際行われたアンケート「分析的評価項目以外に重視する評価項目」から得られた「気持ちや考えがその子らしく表現されているか」、「個性が感じられるか」といった項目が重視され、評定者のもつ価値観が評価に反映されやすいとした。低群では分析的評価項目から「使用できる項目」の候補が多く検出された。よって、グループにより評価方法が異なると考察している。さらに、中学年の作文評価において、高群に比べ低群は総合評価項目の妥当性の検討の結果から、「うまさ」と「好み」とで重視される項目が異なる傾向が強いとした。低群の結果より、作文評価において「うまさ」（客観的総合評価）では「構成力」を重視し、「好み」（主観的総合評価）では「叙述・描写力」を重視するという石田・森（1985）<sup>(12)</sup>の傾向を支持している。これにより、総合評価において「うまい」と「好み」とで異なる評価が行われる傾向が強く、総合評価活動には2極文節化が起こったとした。このことが、中学年における低群の妥当性・信頼性を低める一因となったと考察している。

作文、評定者、項目の3要因の作文評価への影響の検討からは、残差の要因がもっとも大きな割合を占めていたことから、生活文評価は説明困難な部分が大きいとした。評定者の要因から、高群では評定者間で評定結果が一致しにくい傾向が強いことが示されたとされている。

項目の要因からは、高群では低群に比べ評定者間で評定結果が一致しにくく、項目を使い分けて評価することが困難であるとされ、項目の妥当性・信頼性を低めた一因と推察している。よって分散推定値割合から、生活文は群によらず直感的になりやすく、評価項目が反映されにくいとした。好意度別にみると高群では低群に比べて、評定者間で評定結果が一致しにくく、その上、項目を使い分けて評価を行うことが困難であることが、項目の妥当性・信頼性を低めた一因としている。そして低群では高群に比べ作文に対する好みははっきりしており、項目を使い分けて評価を行う傾向が強く、評定者間で評定結果が一致する傾向が強いことを示した。

この調査の結果と梶井(2001)<sup>(10)</sup>の結果から、大学生と教員との間で中学年での評価の取り扱いには注意が必要であること、好意度によって評価観に違いが見られることが共通点として挙げられている。大学生と教員との評価観の違いについて、「使用できる項目」の数からは、大学生低群と教員の高群とで評価傾向が似ていることが示されたが、自由記述の回答より、その中身については教員高群では評定者のもつ児童観が評価に反映されやすいのに対し、大学生低群においては児童の発達段階や年齢に関する、作文の外見を重視した基準が評価に反映されやすいことを示した。さらに、分散推定値割合の大きさからは、大学生の低群では、文章の体裁、作文から予測される発達段階や年齢といった作文の外見重視の作文に対する好みは評価に影響を及ぼすことに加え、分析的評価項目を使い分けて評価を行う傾向が強かったことから、評定者間で評定結果が一致する傾向が強くなったのではとしている。その一方、教員高群では、評定者それぞれのもつ児童観を反映させた好みは評価に影響を及ぼすことに加え、分析的評価項目を評価に役立てるのが困難であったことから、評定者間で評定結果が一致しにくくなったと推察している。よって、好意度が評価に及ぼす影響は、質的に異なることを示した。教員の評価には、単に好意度だけで決まるのではなく、教育実践で培われた児童に関する体験的情報による影響が存在するとしている一方、大学生の結果からは評価が作文の外見に惑わされる傾向や、作文が苦手であるため、項目に忠実に評価しようとする傾向があったとした。この研究から、児童の作文評価の熟達化には好意度に加え、実践場面での子どもへの指導経験が大きく影響し、児童の作文認知も指導経験により異なることが明らかになった。

吉川・岸(2006)<sup>(13)</sup>は、評価課題として小学校の教科書に掲載されている意見文を用いた。評定項目は以下の通りである。①総合評価項目(1項目):「どのくらい面白い作文だと思いましたか」、②教員使用項目(10項目):梶井(2001)<sup>(10)</sup>の高学年用の6項目(学習指導要領項目)に、梶井(2001)<sup>(10)</sup>の調査の自由記述で教師が用いている

と答えた評定項目4項目(その子らしさに関する項目)を加えたもの、③印象項目(本人用印象項目8項目、小学生用印象項目5項目)「わかりやすさ」や「伝わりやすさ」の観点からなる5項目(「伝わりやすさ」の項目)、「説得性」、「賛同・理解」の観点からの2項目、「個性重視」の観点からなる1項目の以上8項目の本人用印象項目と小学生が作文を読んだときにどう感じるかどうかを予測する5項目の小学生用印象項目を印象項目(13項目)。意見文評価に適した評価項目の検討、どのような文章の書かれ方がされたときに評価が高くなるのかの検討、評価の高低は評定者の文章に対する共感性の程度と関連するかの検討を行った。共感経験尺度、文章評価事前調査、文章評価課題の3つの調査を教育学部の大学生75名に対して行った。共感経験尺度は個人の共感レベルを測定するため、文章評価事前調査は個人要因として、文章のテーマ・内容への興味を測定するため行われた。文章評価課題は小学校の教科書に記載されているもの5種類を用い、それを総合評価項目(上記の①)、教員使用項目(上記の②)、印象項目(上記の③)を用いて採点を行う課題である。

評価の結果により、「学習指導要領項目」は内容の良し悪しに関わらず、その文章の書かれ方のみを評価できる項目であることを示した。書かれ方による評価の変化に関して、「学習指導要領項目」は文章の書かれ方に影響を受ける項目であった。評価と共感性の関連について、文章評価項目における「教員使用項目」の「その子らしさに関する項目」は個人の共感性のレベルに影響される。それに対し、「学習指導要領項目」は個人の共感性の影響をほとんど受けなかった。この研究から、意見文を適切に評価するには、「学習指導要領項目」のように文章の構造や論点の明確さ等を評価する項目だけでなく、「その子らしさに関する項目」のように個性を評価する項目、「伝わりやすさ」に関する項目のように文章を読んだ際の印象を評価する項目が必要とした。

平(1995)<sup>(14)</sup>は、妥当性・信頼性の高い作文能力の測定を目指し研究を行った。さらにその作文能力の評価が、どのような要因と関わりが深いのかについて分析を行った。首都圏の国立大学附属中学校1年生1クラス39名、地方公立普通科高校3年生3クラス128名、合計167名を被験者に、教科の「学力」や特別な訓練の影響をあまり受けない、純粋な「作文能力」の測定が可能となるとして「絵を見てお話を作る」物語作成課題を取り組ませた。基礎言語能力をみるため「語彙理解力」、「漢字読み取り能力」を客観式テストで評価を行い、作文能力の評価との関係を吟味した。作文の評価がどのような要因によって説明されるかを分析するため被験者にアンケートを実施した。中高生に作成させた物語文を国語の専門家が相互に独立して、分析的観点による評価と総合的評価(作

文の全体的な印象を評価)を行った。

調査により、作文評価は十分な配慮がなされていれば、妥当性・信頼性の高い評価が可能であること、客観式テストで測定される学力とは異なる能力の評価が期待できることを示した。アンケートからは、作文能力の育成を目指した子どもの教育には、文章を書くことが「好き」という感情を持たせることは大きな意味を持ち、他発的な訓練はあまり有効でないと考察している。

岸・綿井(1997)<sup>(15)</sup>は、マニュアルなどの説明的文章は「わかる」ことを前提条件とした。その上に、読みやすいこと、理解が容易であること、知識を獲得しやすいこと等という要素が加わり、不特定多数の読み手の誰にでもわかりやすい説明表現にするという高次の表現技能が要求されると述べている。その「わかりやすい」説明を書くための規定要因や説明の基礎技能を解明することを目的とした。そのため、予備実験としてわかりやすさの程度がどのような評価基準によって判断されるのかを検討した。本実験で用いる評価項目の選択を行い、本実験において手続き的説明文の産出状況を設定し、説明内容に関する先行知識の量と学習者の知識状態を知ることが説明のわかりやすさとどのように関係するのかを検討した。予備実験から、10項目の評価項目のうち説明文のわかりやすさの基本的な評価項目として、「語彙の適切さ」、「具体例のわかりやすさ」、「説明順序の適切さ」、「目的との適合」の4項目が選ばれた。予備実験をもとに、本実験では「語彙の適切さ」、「段落の意味のまとまり」、「説明順序の適切さ」、「目的との適合」の4項目を用いて手続き的説明文を評定させた。手続き的説明文は「テニスのシングルス・ゲームの進め方」を説明するという課題を、テニスは知っているがゲームの進め方は知らない未経験者(14名)にゲーム進行に必要な内容をあらかじめ伝えさせた上で書かせた(未経験群)。そして、ゲーム経験者(14名)にはまず説明文を書かせ(経験1群)、その4～5日後にゲーム進行に必要な内容に関する質問を受けるという面接を行った後にもう一度文章を作成させた(経験2群)。

この3群の説明文のわかりやすさの違いをみたところ、わかりやすい手続き的説明文を書くための技能には、説明内容に関する先行知識量とともに、読み手の知識量を把握することが関連するとした。また先行知識が少なくても、説明内容に関する事前説明により、わかりやすい説明文が書ける可能性があるとして述べている。

### I-3 目的

評価する対象について、梶井(2001, 2002)<sup>(10)(11)</sup>や吉川・岸(2006)<sup>(13)</sup>では小学生の作文に対して、岸・綿井(1997)<sup>(15)</sup>では大学生の文章に対して評価を行っている。これまでの教育心理学の立場からの作文の評価研究は、小学生児童

の作文を対象としたものが多く、中学生の作文が対象の研究は、平(1995)<sup>(14)</sup>の研究などのように、あまり行われていない。だが現在、「書くこと」の指導は、文章を書くことに苦手意識を持っている中学生、書き方のトレーニングを必要とする生徒にとっては重要な指導である。中学校においても、作文指導の充実とその評価方法の確立が急務といえよう。

そこで本研究では中学校国語科の学習指導要領(1998)<sup>(1)</sup>を反映させた中学校の作文評価の基準を作成し、その基準の妥当性と信頼性について検討を行う。

本研究では、項目の信頼性と妥当性の検討を、梶井(2001, 2002)<sup>(10)(11)</sup>と同様の手続きにより行う。具体的には、中学校国語科の学習指導要領(1998)<sup>(1)</sup>、国語科教育で考えられている作文力、教育心理学での作文研究を反映させた評価項目(分析的項目)と総合項目で中学生の作文を評定したものを3つの基準を満たす項目の抽出を行う。3つの基準は①作文の直感的よしあしを識別できる、②評定者間で評定結果が一致する傾向が強い、③総合評価と関係が強い、の3つである。3つの基準の①は作文を読んだ後の作文に対する印象を反映する項目をみるため、②は評定者の主観が影響する項目を排除するため、③は総合評価に及ぼす影響力が大きい項目を抽出するため設定する。3つの基準を同時に満たす項目を「使用できる項目(妥当性の高い項目)」、満たさない項目を「使用の難しい項目(妥当性の低い項目)」とする。

次に、評価に影響を及ぼす要因として作文、評定者、評価項目の3要因を考え、これらの要因がどのような様相で、どの程度、評価に影響を及ぼしうるのかを明らかにする。具体的には、一般化可能性理論を用いて、作文×評定者×評価項目の3相データから信頼性係数(G係数)とそれぞれの要因の分散推定値割合を算出し、作文に対する好意度と教師・学生の別の違いによりそれらの値がどのように異なるかを明らかにする。

## II 方法

### II-1 評価対象の作文

2007年10月15日～18日、兵庫県の公立中学校3年生の2クラス59名に対し各クラスの授業時間内に実施した。実施する際、課題に取り組みせる前に意見文のテーマ以外の自分の興味のある事柄について概念地図を書かせた。その後意見文課題に取り組みさせた。概念地図とは、「人間の持つ知識を図に表現させることで、その構造を評価し、精緻化させる方法」(藤村, 2006)<sup>(16)</sup>である。今回、予備調査の結果、書くことを困難としている生徒の存在があった。彼らが課題に取り組み際、書こうとする概念に対して自分の知っていることを図に表現するという方法を用いることで、書くことへの難しさを和らげられるよう実施した。課題とその教示は、図1, 2に示した。

この課題は、国立国語研究所（2007）の実施した「特定の課題に関する調査」の国語の問題を用いたものである。教示として提示したものは、この調査に付されていたものと、ほぼ同じ内容のものとなっている。

ある雑誌に「小学生はテレビを見すぎている」という記事が載っていました。これを見て、田中さんの学校では学年全体で「小学生のテレビの見方」について、話し合うことになりました。あなたが田中さんのクラスの一人だとして、自分の考えを書きましょう。

図1 意見文課題

#### 「作文の条件」

- ① 四百字以上六百字以内で書きましょう。
- ② 読む人に分かりやすい文章になるよう、工夫して書きましょう。
- ③ 見たり聞いたりしたこと、体験したことを区別して書きましょう。
- ④ 段落を分けて書きましょう。
- ⑤ 原稿用紙には題名は書かず、一行目から本文を書きましょう。

図2 教示

### 評価対象とする作文の選択

得られた作文の中から、本研究における評価課題となる作文を選ぶため、著者・研究協力者（大学4年生）の2名がそれぞれ、得られた作文59編に対して1～5点で評価を行った。その結果、評価の一致する作品から、それぞれに3作品ずつ選び、合計15編の作品を選んだ。2名の評定者間の全59編の作文についての評定値のピアマンの順位相関係数を求めたところ、.578 ( $p<.05$ ) となり中程度の相関が見られた。

### II-2 作文への評価の調査

#### 被調査者

中学校教師を中心とする国語指導経験者（以下、教師群）18名、国語教育を専門とする大学生（以下、学生群）17名に調査を依頼した。その結果、返信があり、かつ回答が有効だったもの（教師群：15名、学生群14名）を調査の対象とした。被調査者全体に対し、「作文が得意であるか否か」、「作文が好きであるか否か」という2つの質問を実施した。教師群、学生群それぞれに対して、「作文が得意であるか否か」、「作文が好きであるか否か」に関する2つの質問からそれぞれを2グループに分類した。作文が得意もしくは好きと回答したグループを、好意度が高いグループと判断し「高群」、得意でも好きでもない

と回答したグループを「低群」と定義した。以下、教師群・学生群を好意度の違いで分類し、この違いが、評価項目の妥当性と信頼性に及ぼす影響を検討する。

#### 調査期間・場所

調査期間は、2007年10月31日～11月30日であった。中学生に示した課題文と教示文、15編の作文とそれぞれの評価項目を調査冊子として配布し、各評定者が個別に任意の時間、場所で行った。

#### 評価項目

評価項目は、分析的項目・総合項目を用意した。総合項目は石田・森（1985）<sup>(12)</sup>の「うまさ」と「好み」の2項目を用いた。

石田・森（1995）<sup>(12)</sup>の「うまさ」に関する項目を「客観的総合評価」、「好み」を「主観的総合評価」と定義し、大学生の「うまさ」と「好み」を基準とした評価構造に違いが見られることを明らかにすることを目的とした。具体的な項目は表1に示した。

次に、分析的項目について述べる。中学校国語科の学習指導要領（1998）<sup>(1)</sup>の「書くこと」では、第1学年において日常の経験や周囲の様々な情報に基づき、自分の見方・考え方を明確にすることから「書くこと」の能力を高めようとしている。さらに、第2、3学年では日常の経験や周囲の様々な情報に対する自分の立場を明確にし、文章を論理的に書き起こすことで、言語で認識し思考する能力の発展が期待される。自分の伝えたい事柄や意見が、相手に誤解されずに効果的に伝わるよう、事柄・意見の根拠を明確にし、論理の展開を工夫するような内容表現や文章構成に関わる力が必要とされる。この際に、文や文章の形を整えてより説得力のある文章にする能力も必要とされる（文部科学省、1998）<sup>(1)</sup>。

国語科研究では、研究者により作文力の規定の仕方が異なる（井上、1981）<sup>(9)</sup>。1つは、「作文力」とは文章を書くために必要な様々な力量、例えば認識力、思考力、感受力といった力の総体であると規定する立場がある（鈴木、1984）<sup>(17)</sup>。もう1つは、コンポジション・スキルと呼ばれる、主題、材料、構成、段落、文、語句、文法、つづり、表記に関する能力を文章表現力（作文力）とするものである（森岡、1977）<sup>(18)</sup>。

このように、作文力として様々な力が規定されている。本研究では、国語科の学習指導要領（1998）<sup>(1)</sup>より、自分の伝えたい事柄や意見が、相手に誤解されずに効果的に伝わるよう、事柄・意見の根拠を明確にする力として「内容表現」、論理の展開を工夫する力として「文章構成」、文章を整えてより説得力のある文章にする力として「表記」を作文力の構成概念とすることとした。国語科教育における作文力との関係は、具体的に、「内容表現」では、書き手の認識力や思考力、感受力といった力が含まれるものとする。「文章構成力」・「表記力」については、森

岡 (1977)<sup>(18)</sup>のコンポジション・スキルに示される基礎能力 (主題の展開, 段落, 効果的な文, 用語, 文字・表記) が含まれるものとする。

そこで, 具体的には, 先行研究 (梶井, 2001<sup>(10)</sup>; 岸・綿井, 1997<sup>(15)</sup>; 平, 1995<sup>(14)</sup>; 吉川・岸, 2006<sup>(13)</sup>) で用いられた項目, 学校図書<sup>(19)</sup>, 教育出版<sup>(20)</sup>, 光村図書<sup>(21)</sup>の3社の教科書の「書くこと」の領域の目標をもとに, 本研究で使用する評価項目を表2のように作成した。内容表現には感覚力・知覚力, 思考力, 情意力に関わる項目として項目1~9を, 表記については語彙力, 表記力に関わる項目として項目10~14, 文章構成には文(章)を構成する力に関する項目として項目15~17を用意した。評価項目の後の記号は, 参考または引用元の文献を示す。a: 平 (1995)<sup>(14)</sup>, b: 吉川・岸 (2006)<sup>(13)</sup>, c: 梶井 (2001)<sup>(10)</sup>, d: 岸・綿井 (1997)<sup>(15)</sup>, e: 教科書<sup>(19)(20)(21)</sup>を参考に作成した評価項目とした。

表1 総合項目

1. どのくらい面白いと感じるか
2. どのくらい好きか

表2 分析的項目

1. 内容に独創性・目新しさがある(a)
2. 生き生きと書けている(b)
3. 自分なりの人生経験や知識を背景としている(b)
4. ものごとを多面的に捉え, 論点がしぼられている(c)
5. 自分の立場をしっかりと持っている(a)
6. 語り口を工夫するなどして, 読み手や伝える目的を意識している(c)
7. 最も伝えたいことがよく分かる文章である(a)
8. 読み手をひきつける文章である(a)
9. 自分の意見の根拠を明らかにしている(e)
10. ことばの選び方, 使い方が正しく行われている(e)
11. 使用している漢字は正確に, また適切に用いられている(a)
12. 伝えたい事柄が, 相手に効果的に伝わるような言葉を選んでいる(e)
13. 主語と述語の照応(対応)がなされている(c)
14. 接続語(句, 単語, 文間, 段落間のつながりを明確にする語)の使い方が適切である(d)
15. 読み手に違和感を感じさせない話の展開になっている(e)
16. 書き出しから結びまでが一貫した整った文章である(c)
17. 書き出し, 中心部分の展開, 結びに, 適切な構成を工夫している(e)

### 評価課題

意見文15編を, 2つの総合項目と17つの分析的項目に沿って評定してもらった。19項目の配列は, 先に総合項目を「うまさ」, 「好み」の順で行うよう構成した。分析的項目は表2に示す番号順に構成した。評定は各々1~

5の5件法とした。調査冊子は表紙に「調査の主旨」, 1頁に「中学生が取り組んだ作文課題とその教示」, 2頁以降に, 見開きの左頁に生徒の作文, 右頁に評価項目を載せ, 最終頁に「作文が得意であるかどうか」, 「作文が好きであるか否か」の2つの好意度に関する質問を載せた。さらに, 自由記述として今回の調査の分析的項目17項目の内, 特に評価の際に重視した項目を4つ挙げてもらった。また, 17項目以外に重視する項目を自由記述してもらった。調査冊子は, 評定結果に順序効果が及ばぬよう, 15編を作品1~15の順で並べたものと作品15~1の順で並べた2種類を準備した。

### II-3 分析について

#### 評価者の好意度による分類

教師群, 学生群それぞれに対し, 「作文が得意であるか否か」, 「作文が好きであるか否か」の2つの質問から各々を2グループに分類した。作文が得意もしくは好きと回答したグループを, 好意度が高いグループであるとし「高群」, 得意でも好きでもないと回答したグループを「低群」と定義した。各々の質問への回答の分布を, 以下の表3, 4に示す。「作文は好きですか。」の質問に対し, 学生低群に分類されている被調査者の中に4と回答したものがあつた。その被調査者を分類することが, 表3, 4のみでは困難であったため, 2つの質問の回答を合計したもの(表5)から判断することとした。2つの質問を合計したものから, 高群は7~10, 低群は2~6と明確に得点幅が分かれた。よって, 教師高群8名, 教師低群7名, 学生高群8名, 学生低群6名を被調査者とした。

表3 「作文は好きですか。」に対する回答一覧

	1	2	3	4	5	合計
教師高群	0	0	0	5	3	8
低群	0	4	3	0	0	7
学生高群	0	0	1	5	2	8
低群	2	3	0	1	0	6
合計	2	7	4	10	5	29

表4 「作文は好きですか。」に対する回答一覧

	1	2	3	4	5	合計
教師高群	0	0	4	2	2	8
低群	2	4	1	0	0	7
学生高群	0	0	1	7	0	8
低群	3	3	0	0	0	6
合計	5	7	6	9	2	29

## 分析方法

### 評価項目の妥当性・信頼性

#### a. 分析的項目の妥当性

中学生の作文はどのような評価がされるのかを、評定者に意見文を「うまさ」と「好み」の総合項目と「内容表現」・「表記」・「文章構成」を作文力の構成概念として作成した17項目により評定してもらった。これ以降の全ての分析には、統計パッケージR (Ver.2.6.1) を用いた。

表5 「作文は好きですか。」「作文は得意ですか。」の回答の合計

	2	3	4	5	6	7	8	9	10	合計
教師高群	0	0	0	0	0	3	2	2	1	8
低群	0	1	4	1	1	0	0	0	0	7
学生高群	0	0	0	0	0	2	4	2	0	8
低群	2	1	2	0	1	0	0	0	0	6
合計	2	2	6	1	2	5	6	4	1	29

まず、評価項目の妥当性の検討を行う。各々の分析的項目の評定値に影響を及ぼすと考えられる因子を「作文」、「評定者」として、作文×評定者の乱塊配置モデルの下、限界水準を吟味する。乱塊配置モデルでは、

$$y_{(a)jk} = \mu_{(a)} + \alpha_{(a)j} + \beta_{(a)k} + e_{(a)jk} \quad \dots(1)$$

という構造模型を考える。 $\omega$ は分析的項目の項目、 $j$ は作文、 $k$ は評定者を示す。 $y_{(a)jk}$ は、作文 $j$ に、評定者 $k$ が分析的項目 $(a)$ に対し「評定した真の値(期待値)」である。 $\mu_{(a)}$ は分析的項目 $(a)$ の全体の評定平均、 $\alpha_{(a)j}$ は作文 $j$ の「真の評定値」、 $\beta_{(a)k}$ は評定者 $k$ の「評定の甘さ」を表すと考える。 $e_{(a)jk}$ はその際の測定誤差を表す確率変数である。評定結果より、17項目に対し分散分析を実施した(手続き1)。限界水準の値から、作文の要因が有意で、かつ評定者の要因が有意でない、つまり作文のよしあしに散らばりはあるが、評定者の甘さには散らばりがない項目を「使用できる項目」、作文の要因が有意でなく、評定者の要因が有意な項目を「使用の難しい項目」とした。

次に、総合項目「うまさ」・「好み」のそれぞれがどの分析的項目と強い関係をもつのかを検出するため、総合項目の「うまさ」・「好み」のそれぞれの評定値を基準変数、分析的項目の評定値を説明変数としたstepwise法による重回帰分析を行った(手続き2)。この手続きは2つの総合項目「うまさ」(客観的評価項目)と「好み」(主観的評価項目)のそれぞれに分析的項目がどのような様相で影響を与えているのかをみるために行う。stepwiseのための有意水準は.05として総合項目と有意な相関が見られる項目を抽出した。

これらの2つの手続きから分析的項目の妥当性の検討を行う。分散分析により作文の直感的よしあしの識別を

おこない、重回帰分析により評定者間で評定結果が一致する傾向をみた。2つの手続きを通過できた項目を「使用できる項目」、どちらか片方の手続きを通過した項目を「要検討の項目」、どちらの手続きも通過できなかった項目を「使用の難しい項目」とした。この分散分析と重回帰分析の2つの手続き通過したかどうかを表6に一覧にして示した。手続き1・2ともに通過した項目を◎、手続き1のみ通過した項目は○、手続き2のみを通過した項目は\*、手続き1・2ともに「使用の困難な項目」は×と表した。空欄は、手続き1において「要検討の項目」となり、手続き2を通過しなかった項目として表した。

#### b. 分析的項目の信頼性

a.の分析的項目の分散分析の結果より、項目ごとの信頼性係数を算出した。評定値の分散に占める要因の効果の分散の割合を、測定の信頼性の指標とする。分散分析を利用して測定の信頼性を推定する方法を「一般化可能性理論」という(例えば、豊田, 1994)<sup>(22)</sup>。「1つの作文に1人だけの評定者が得点を与え、別の1つの作文にまた別の評定者が得点を与える」場合の信頼性の指標は、

$$\rho = \frac{\sigma^2_{\alpha}}{\sigma^2_y} \quad \dots(2)$$

で推定される。ここでは、「1つの作文に $n$ 人の評定者が得点を与え、別の作文にまた別の $n$ 人の評定者が得点を与える」場合を考える。この場合、誤差の分散は評定者数で割るので、信頼性の指標は、

$$\rho = \frac{\sigma^2_{\alpha}}{\sigma^2_{\alpha} + \left(\frac{\sigma^2_e}{n}\right)} \quad \dots(3)$$

で推定される。これから推定される推定値はクロンバックの $\alpha$ 係数とも一致する。教師・学生、好意度別に、各項目の $\alpha$ 係数の推定を行う。結果は表6に示す。

#### c. 評価モデルの信頼性

繰り返しのない3因子実験で、「作文」、「評定者」、「分析的項目」を因子として分散分析を行う。この3因子の組み合わせでは1つの得点しかないので、「繰り返しのない3因子実験」である。この構造模型は、

$$y_{jkl} = \mu + a_j + b_k + c_l + (ab)_{jk} + (bc)_{kl} + (ac)_{lj} + e_{jkl} \quad \dots(4)$$

と表される。「分析的項目」の因子は観点間の難易度の差を表している。3因子実験では繰り返しのない場合でも交互作用を考えるので、 $(ab)_{jk}$ 、 $(bc)_{kl}$ 、 $(ac)_{lj}$ が構造模型に含まれている。これのもと、分散成分とG係数の推定を行う。作品も評定者も評価項目も、たまたま選ばれたものとする。本研究でのG係数は、教師高群( $b'=8$ )、教師低群( $b'=7$ )、学生高群( $b'=8$ )、学生低群( $b'=6$ )がそれぞれ、

15編の作文を17の分析的項目(c'=17)に沿って評価を行った際の評定結果であることから、

$$\rho = \frac{\sigma_a^2 + \sigma_{ab}^2 + \sigma_{ac}^2}{b'} + \frac{\sigma_e^2}{b' \times c'} \dots(5)$$

(5)式から得られた値をG係数とする。結果は表6に示す。

#### d. 総合項目の妥当性

教師と学生、好意度別で分析的項目の合計と総合項目の相関をみる。相関係数は梶井(2001)<sup>(10)</sup>と同様、教師か学生か、及び好意度に関わらず「うまさ」は強い相関となった。「好み」については好意度に関わらず、教師では強い相関が見られ、学生では中程度の相関となった。

#### 評価に影響を及ぼす要因

評価モデルの信頼性を(4)式から推定した。(4)式には(ab)jk, (bc)kl, (ac)ljの交互作用が含まれる。そこで、主効果及び交互作用の解釈を行うため、3因子実験のもと変動要因間の分散成分の推定値を教師・学生、好意度別に算出した。分散成分の比較のため、分散推定値割合を用いる。

$$\text{分散推定値割合(\%)} = \frac{\text{当該要因の分散推定値}}{\text{全要因の分散推定値合計}} \times 100 \dots(6)$$

表7より、作文、作文×評定者の割合が他の要因よりも大きく、項目の割合が小さいことが示された。作文×評定者×項目の割合も大きい。繰り返しのない3因子実験では、作文×評定者×項目の要因と誤差は分けて考えないので、作文×評定者×項目を誤差とする。

表6 項目の信頼性・妥当性の一覧表

分析的項目	教師高群			教師低群			学生高群			学生低群		
	うまさ	好み	α係数	うまさ	好み	α係数	うまさ	好み	α係数	うまさ	好み	α係数
1. 内容に独創性・目新しさがある		*	.842	*	*	.655			.695		*	.682
2. 生き生きと書いている		*	.823			.683	×	×	.295			.434
3. 自分なりの人生経験や知識を背景としている			.729			.675			.849			.700
4. ものごとを多面的に捉え、論点がしぼられている	*		.836	*		.740			.800	*		.769
5. 自分の立場をしっかりと持っている			.844			.663		*	.768			.799
6. 語り口を工夫するなどして、読み手や伝える目的を意識している			.879	*		.775		*	.755	*		.873
7. 最も伝えたいことがよく分かる文章である			.868			.683			.817		*	.807
8. 読み手をひきつける文章である			.844		*	.734		*	.736		*	.781
9. 自分の意見の根拠を明らかにしている	*		.814	*		.721			.588			.759
10. ことばの選び方、使い方が正しく行われている	*		.919	*		.892			.888	*		.911
11. 使用している漢字は正確に、また適切に用いられている	*		.840	*		.841			.898	*		.845
12. 伝えたい事柄が、相手に効果的に伝わるような言葉を選んでいる			.902			.818			.852			.840
13. 主語と述語の照応(対応)がなされている	*		.870	*		.831		*	.899			.830
14. 接続語(句、単語、文間、段落間のつながりを明確にする語)の使い方が適切である			.841		*	.796			.887	*		.828
15. 読み手に違和感を感じさせない話の展開になっている	*		.852	*	*	.656	*	*	.829	◎	○	.898
16. 書き出しから結びまでが一貫した整った文章である			.879			.705			.790			.754
17. 書き出し、中心部分の展開、結びに、適切な構成を工夫している	*		.866			.836	*		.888	*		.864
相関	.919	.835		.841	.825		.865	.672		.830	.519	
G係数		.920			.865			.921			.884	

表7 分散推定値割合の一覧表

変動要因	教師高群	教師低群	学生高群	学生低群
作文(作文のよしあしのばらつき)	34.45%	20.00%	25.13%	25.41%
評定者(評価の評定者間不一致度)	9.62%	16.51%	8.50%	10.15%
項目(項目を使い分けて評価を行う傾向が強い)	2.09%	1.99%	2.71%	3.50%
作文×評定者(作文への好みの評定者により異なる)	22.45%	20.60%	15.53%	18.29%
作文×項目(作文ごとの評価項目への得手不得手)	2.75%	3.78%	9.25%	5.86%
評定者×項目(評価項目に対する重要度の違い)	2.42%	6.33%	4.50%	3.16%
残差(3要因では説明できない割合)	26.22%	30.78%	34.39%	33.63%

#### III-4 17項目以外に重視する項目の調査の結果

評定冊子の最後に、アンケートで④評価で用いた分析的項目のうちで、特に重視した項目を4つ挙げさせた。また、⑥分析的項目以外に評定者が重視する項目があれば記述を求めた。その結果、④の回答より、特に重視された項目の上位4項目を表8に挙げる。⑥の回答から得られたものを表9に示す。

#### IV 結果と考察

作文×評定者×評価項目からなる3相の作文評価データに対して、a~dの分析を行った。以下に、その分析結果ごとに考察を行う。

##### IV-1 評価項目の検討

###### a. 分析的項目の妥当性の検討

分析的項目について、分散分析で作文によって評定が



表8 重視する項目の回答数の多かった項目

	教師高群	教師低群	学生高群	学生低群
重視する項目	4 (6)	3 (4)	2 (5)	3 (4)
	7 (6)	10 (4)	3 (5)	3 (4)
	9 (4)	11 (4)	10 (4)	7 (4)
	5 (3)	5 (3)	5 (3)	9 (3)

※ 表の括弧内の数値は回答件数を示す。

表8 重視する項目の回答数の多かった項目

好意度	項目	件数	項目例
教師低群	内容表現	1	自分の体験が十分に書けているかどうか、その上で結論を出しているかどうか
		1	主題の質
		1	常体、敬体の統一
	表記	3	字数はあるか、分量はあるか
		1	原稿用紙の使い方
		1	符号、句読点の使い方
学生低群	内容表現	2	自分の言葉で書けているか
	表記	2	句読点の使い方

※ 教師高群、学生高群は略した。

異なり、評定者間で評定が一致する項目の検出、重回帰分析によって総合項目と関係の強い項目の検出を行い、分析的項目の妥当性を検討した。その結果は、多くの項目が「要検討の項目」となった。これは、今回用いた評価項目を基準に用いて評価するのは困難であったことを示している。アンケートの回答の中に、評価項目にあいまいな部分が含まれており評価者により評定に違いが生じるのではという回答があった。項目1「内容に創造性、目新しさがある」、項目2「生き生きと書けている」、項目8「読み手をひきつける文章である」、項目15「読み手に違和感を感じさせない話の展開になっている」の創造性、目新しさ、生き生き、ひきつける、違和感等の語を含む項目への指摘が多かった。よって評価基準には、より明確な基準が必要とされ、創造性、目新しさ、生き生きのような言葉を含む評価項目は、注意を要すると示唆された。

#### 教師高群の評価

教師高群では手続き1を通過できた項目はなかったが、手続き2において「うまさ」と「好み」各々で通過する項目があった。総合項目「うまさ」に影響を与える分析的項目として抽出された項目は項目4「ものごとを多面的に捉え、論点がしぼられている」、項目9「自分の意見の根拠を明らかにしている」、項目10「言葉の選び方、使い方が正しく行われている」、項目11「使用している漢字は正確に、また適切に用いられている」、項目17「書き出し、中心部分の展開、結びに適切な構成を工夫している」

であった。「うまさ」に影響を及ぼすものとして抽出された項目（項目4、項目9、項目10、項目11、項目17）は、「好み」に影響を与える項目（項目1、項目2、項目13、項目15）と比較すると、構成要素の「表記」に分類した項目（分析的項目の項目10～14）が多く含まれていた。「好み」は「うまさ」に比べ、構成要素の「内容表現」に分類した項目（分析的項目の項目1～9）が多く含まれていた。「うまさ」と「好み」の各々に影響を及ぼす項目として抽出された項目は異なり、「うまさ」は表記に関わる項目、「好み」は内容表現に関わる項目に影響を受けていると考えられる。以上より、教師高群における主観的総合評価と客観的総合評価の要素は異なることが示された。

#### 教師低群の評価

教師低群でも手続き1を通過できた項目はなかったが、手続き2においてはそれぞれで通過する項目があり、特に「うまさ」においては項目1「内容に創造性・目新しさがある」、項目4「ものごとを多面的に捉え、論点がしぼられている」、項目9「自分の意見の根拠を明らかにしている」、項目10「言葉の選び方、使い方が正しく行われている」、項目11「使用している漢字は正確に、また適切に用いられている」、項目13「主語と述語の照応（対応）がなされている」、項目15「読み手に違和感を感じさせない話の展開になっている」と8項目が抽出された。「好み」については、項目1、項目8「読み手をひきつける文章である」、項目14「接続語（句、単語、文間、段落間のつながりを明確にする語）の使い方が適切である」、項目15の4項目が抽出された。「うまさ」においては、構成要素「内容表現」に関わる項目（項目1～9）が4項目、「表記」に関わる項目（項目10～14）が3項目と両構成要素に関わる項目が複数検出された。よって教師低群の「うまさ」（客観的総合評価）に、今回用いた評価基準の多くが影響を与えている、中でも「内容表現」・「表記」に関わる項目が大きく影響を与えていることが推察される。

#### 学生高群の評価

学生高群では手続き1において、項目2「生き生きと書けている」が通過できない「使用の難しい項目」の候補となった。手続き2においては「うまさ」と「好み」それぞれで通過する項目があった。手続き1を通過できなかった項目2は手続き2においても通過できず、「使用の難しい項目」となった。「生き生き」といった表現が非常に主観的であることから、項目2は評価に役立てにくく、総合評価に反映されにくかったと考えられる。総合評価に影響を及ぼす項目に関して、客観的総合評価「うまさ」は「文章構成」に関する項目15「読み手に違和感を感じさせない話の展開になっている」、項目17「書き出し、中心部分の展開、結びまでが一貫した整った文章である」に、主観的総合評価「好み」は「内容表現」に関

する項目5「自分の立場を持てている」、項目7「最も伝えたいことがよくわかる文章である」、項目8「読み手をひきつける文章である」に影響を受けていることが示された。

#### 学生低群の評価

学生低群では手続き1で、項目15「読み手に違和感を感じさせない話の展開になっている」、項目16「書き出しから結びまでが一貫した整った文章である」で「使用できる項目」が抽出された。手続き2では「うまさ」と「好み」各々で通過する項目があった。手続き1と手続き2の両方を通過した項目は「うまさ」の項目15「読み手に違和感を感じさせない話の展開になっている」であった。総合項目に影響を及ぼした項目を見ると、客観的総合評価「うまさ」には複数の項目が影響していた。主観的総合評価「好み」には内容表現に関わる評価項目（項目1「内容に創造性・目新しさがある」、項目7「最も伝えたいことがよくわかる文章である」、項目8「読み手をひきつける文章である」）が影響していることがわかった。

教師の結果からは、客観的総合評価「うまさ」、主観的総合評価「好み」ともに、今回定義した作文力の構成要素「内容表現」、「表記」、「文章構成」の項目にそれぞれ少なくとも1項目から影響を受けている可能性があることが示された。これより教師は作文力の構成要素全体から作文を評価していると推測される。

学生の結果をみると、学生高群の客観的総合評価「うまさ」には文章構成に関わる項目が、学生低群の主観的総合評価「好み」には内容表現に関わる項目が、大きく影響を与えていることが示された。このことに関して、石田・森(1985)<sup>(12)</sup>は、大学生の評価が客観的には、文章全体の「構成力」を重視するのに対し、主観的には児童の気持ちがよく伝わるかといった「説得力」(本研究においては内容表現に含む)を重視する傾向が強いことを明らかにした。この結果は梶井(2002)<sup>(11)</sup>でも支持され、評価対象学年が中学生となったが本研究においても支持されることとなった。

#### b. 分析的項目の信頼性について

a. 分析的項目の分散分析の結果より、項目ごとの信頼性係数を算出した。結果は、教師高群で $\alpha = .729 \sim .977$ 、教師低群においては $\alpha = .655 \sim .892$ 、学生高群では $\alpha = .588 \sim .899$ (項目2の.295を除く)、学生低群で.682~.911(項目2の.434を除く)の幅の値をとっており、どの群においても、各評価項目は信頼性の高い評価が行われたことを示している。項目2については、教師高群を除いた3つの群で $\alpha$ 値が低い値が算出された。(教師低群 $\alpha = .683$ 、学生高群 $\alpha = .295$ 、学生低群 $\alpha = .434$ )以上の値より3つの群における項目2の評価の信頼性は低いことが示された。教師高群については、項目2の $\alpha$ 値は.823であっ

た。「生き生き」といったあいまいな語を含む項目であったが、教師高群では「生き生きと書けている」ということに教師間で似た評価観を持っていたため、それぞれの評価項目において信頼性の高い評価が行われたと推測される。

#### c. 評価モデルの信頼性について

評価モデルの信頼性は、教師高群において $G = .920$ 、教師低群では $G = .863$ 、学生高群では $G = .921$ 、学生低群においては $G = .884$ という値が得られた。このことから、どの群においても信頼性の高い評価が行われたということが示された。今回の調査において評定者は全て専門を国語とする人であったことから、評定者は国語という教科について似た考えを持っていたと考えられる。そしてさらに、好意度ごとに群に分けたため、似た評価観を持つ評定者により各群が構成された。このことから梶井(2001, 2002)<sup>(10)(11)</sup>のどの信頼性係数よりも高い値が算出されたものと考えられる。実際の中学校の指導現場では、作文評価は国語科の教師により行われると考えられることから、国語科の専門家複数人での評価が行われれば、評価モデルの信頼性はより高まると推察される。好意度別にみると、高群>低群という関係が教師・学生ともにみられる。これより、高群に属する評定者間では低群に比べ、より似た評価観を持っていることが示唆された。

#### d. 総合項目の妥当性について

各群で、分析的項目の評定値合計と2つの総合項目との相関を求めた。すると、教師の群においては「うまさ」・「好み」ともに強い相関がみられた。学生の群においては「うまさ」において強い相関が、「好み」において中程度の相関がみられた。これは、教師は客観的総合評価(「うまさ」)、主観的総合評価(「好み」)の両方の観点を評価に反映させていること、学生は主観的総合評価よりも客観的総合評価のほうが評価に大きく影響していると思われる。

#### IV-2 分散推定値割合について

群の違いにより、作文×評定者×項目の3要因が、評価にどのような様相で、どの程度影響を及ぼしうのかを一般化可能性理論より算出した分散推定値割合(表7)により明らかにした。どの群でも、作文の要因、作文×評定者の要因の割合が残差をのぞく他の要因よりも大きく、学生低群をのぞいて項目の要因の割合が小さいことが分かる。学生低群では評定者×項目の割合が最も小さかった。残差は、作文、評定者、項目の3つの要因では説明できない変動要因に相当し、教師高群では26.221%、教師低群で30.781%、学生高群では34.385%、学生低群では33.634%であった。この値は梶井(2001, 2002)<sup>(10)(11)</sup>の

割合より小さい割合を占め、意見文評価の説明困難な部分は比較的小さい値となった。

作文の要因は、作文のよしあしのばらつきを意味し、この割合が大きければ作文のよしあしが作文によって異なる度合いが大きいことを意味する。

作文×評定者の要因は、評定者の作文に対する好みを意味し、この割合が大きければ作文に対する好みの評定者ごとに異なる度合いが大きいことを意味する。この割合は4つの群で比較すると、教師高群が最も大きい割合を占めていた。

項目の要因は、項目間の識別性を意味し、この割合が大きければ、項目を使い分けて評価を行う傾向が強いことを意味する。項目の要因は教師の2群よりも学生の2群の方が大きい割合となった。中でも学生低群が4群の中で最も大きい割合を示した。

評定者×項目の要因は、評定者が持っている評価項目に対する重要度の評価の違いを意味し、この割合が大きければ評定者ごとに重要視する項目が異なる傾向が強いことを意味する。この評定者×項目の割合を教師高群とそれ以外の群で比較すると、教師高群において割合が最も小さかった。このことより、他の群に比べて教師高群において評定者間で項目に関する重要度の評価が似る傾向があると考えられる。

評定者の要因については、評定の評定者間不一致度に相当し、この割合が大ききことは評定結果が評定者間で一致しにくいことを示す。評定者の割合について好意度の違いによって異なるかをみても、教師・学生ともに高群<低群の関係があり、高群の評定者の間の評定は一致する傾向が強いことを示した。

以上から、教師高群は他の群と比べ、作文への好みの違いが評定者間で大きい傾向が強いが、項目に関する重要度の評価が似る傾向があると思われる。これに対し教師低群においては評定者×項目の要因の割合が他の群に比べ、もっとも大きく項目に関する重要度の評価が異なる傾向があることが示された。教師高群でどの項目においても $\alpha$ 値が高かったのは、項目に関する重要度が似ていることから、評定者間で似た評価がなされたと考えられる。学生低群では他の群と比較して、項目を使い分けて評価する傾向が強いことが示唆された。これは、作文への苦手意識や教師と比べた際の指導経験の少なさのために、項目に忠実に評価しようとしたためではないかと考えた。教師・学生の別で検討を行うと、教師は学生に比べ、作文、評定者、評価項目の3つの要因で説明できる割合が高かった。項目の識別性の割合がどの群においても小さかったことについては、本研究の評価項目にあいまいさを含んだ項目の存在のため、項目を使い分けた評定が困難になったのではと推察される。

### IV-3 教師と大学生に見られる評価の特徴

教師と学生の評価の特徴について、項目の信頼性・妥当性の検討の結果、一般化可能性理論から得られた分散推定値割合の大きさ、自由記述により得られた内容の2点から考察する。妥当性の検討の手続き2において通過した項目の様相が、教師低群、学生低群で似ていた。自由記述の回答からも、両群において「自分の体験が十分に書けているかどうか、その上で結論を出しているかどうか」や「自分の言葉で書けているか」といった書き手の考えに関する項目、「句読点の使い方」の項目を挙げており、両低群の評価の様相は似ていると推察される。さらに、学生高群の分散推定値割合の大きさを他の群と比べると項目を使い分けて評価を行う傾向の強いこと、評定者によって作文の好みの分かれる程度が小さいことから、この群において評価モデルの信頼性( $G=.921$ )が高くなったと考えられる。

図2の①の字数制限について、中学生が課題に取り組む際にどの程度書けばよいのかの目安として示した。そのため、今回の評価項目の中には「字数が守れているか」というような項目は含んでいなかったのだが、字数の制限など作文を書く上での制限を設けているなら、普段の作文指導の際にはそれを守れているのかどうかを評価するとの回答があった。その他にも「原稿用紙の使い方」といった項目は教師高群・低群で重視されている項目であった。実際に指導体験を積んだ群で「表記」に関する項目の指摘が多くみられた。しかし、「表記」だけでなく、「内容・表現」に関わる項目も複数挙げられていることからさまざまな角度から作文の評価を行っていると思われる。

### IV-4 まとめ

本研究の結果、明らかになったことを整理してみると、以下ようになる。

- (1) 教師の評価は構成要素全体から評価がなされた。
- (2) 学生の評価においては、客観的評価項目「うまさ」では「文章構成」が、主観的評価項目「好み」では「内容表現」が重視されることが示された。
- (3) 項目ごとの信頼性は、いくつかの項目を除いて、高い信頼性が得られたことから、各群内で似た評価観をもっていたと考えられる。
- (4) 教師高群では、あいまいさを含む評価項目でも高い信頼性が得られたことから、教師高群はあいまいさについても似た評価観をもっていたと推察される。
- (5) 評価モデルの信頼性の結果より、中学校で行われる作文評価は、国語教師が複数人で行えば、信頼性の高い評価が行えることが示唆された。
- (6) 総合項目の妥当性の検討から、学生群は主観的総合評価よりも客観的総合評価のほうが、評価に及ぼす影響が強い。

(7) 学生低群で、項目を使い分けて評価を行う傾向が強いのは、作文への苦手意識や教師と比べた際の指導経験の少なさのために、項目に忠実に評価しようとしたためではないか。

#### IV-5 今後の課題

今後の課題は、評価項目のあいまいさを小さくすることが挙げられる。項目2「生き生きと書けている」という項目の $\alpha$ 値を見てみると、教師高群以外では低い値が示された。評価項目にあいまいな部分が残ることにより、評定者によって評価項目に対する解釈に違いが生じ、それが評価の差につながると考えられる。よって評価を行う際は、評価項目に関して評価者間での共通理解が必要とされるであろう。

そして、評価課題について本調査において意見文を取り上げたが、説明文課題等、その他の課題について検討の必要があると考える。特に、今回用いた評価項目「内容・表現」に関する項目は、課題により求められる項目は異なると考えられる。課題ごとの基本的な評価項目を検討していく必要があるだろう。

分析方法について、今回は作文×評定者×評価項目の3相データに対し、分散分析モデルによる一般化可能性理論を用いたが、作文評価データを分析するための統計的方法とし、いくつかの方法が提案されている(山内, 1999など)<sup>(23)</sup>だが、それらの方法間の比較はまだ十分に行われていない。そこで今後は、様々な統計的モデルに関する比較・検討を行う必要があるだろう。

最後に、次の学習に繋がるような評価活動を目指すことである。本研究は、作文への評価に主眼をおいた研究であり、学習指導に重点をおいた研究ではなかった。今後は、本研究の評価基準を用いることで、中学生の作文学習指導にどのような貢献ができるのかを考えていくことが、課題となるだろう。

#### V 参考文献

- (1) 文部科学省『中学校学習指導要領 解説 一 国語編一』東京書籍, 1998
- (2) 国立教育政策研究所『生きるための知識と技能 3 OECD生徒の学習到達度調査 (PISA) 2006年調査国際結果報告書』ぎょうせい, 2007
- (3) 中央教育審議会『中央教育審議会答申全文と読み解き解説』明治図書, 2008
- (4) 位藤紀美子「自分なりの考え、自分の言葉で書く文章表現の学習指導の評価」小森茂(編)『生きてはたらく国語の力を育てる授業の創造 第8巻 自分なりの考えを持ち自分の言葉で書くことができる学習指導』ニチブン, pp.264-270, 2000
- (5) 森本正一・長谷川孝士(編)『初等教育 国語科教

- 育の研究』建帛社, pp.77-91, 1980
- (6) 菅原稔「適切に書くことによって育成される文章表現力」小森茂(編)『生きてはたらく国語の力を育てる授業の創造 第6巻 適切に書くことができる学習指導』ニチブン, pp.200-205, 2000
- (7) 岩間正則「指導と評価の一体化を目指した評価づくり(中学校)」清水敬(編)『日本語学』第23巻 第13号 明治書院, pp.54-66, 2004
- (8) 高木展郎「国語科教育における評価の考え方」三樹讓(編)『日本語学』第21巻 第7号 明治書院, pp.13-26, 2002
- (9) 井上敏夫「文章表現力の実態をどうとらえ、どう伸ばすべきか」『国語教育』28, p.3, 1981
- (10) 梶井芳明「児童の作文はどのように評価されるのか?—評価項目の妥当性・信頼性の検討と教員の評価観の解明—」『教育心理学研究』49, pp.480-490, 2001
- (11) 梶井芳明「大学生は児童の作文をどのように評価するのか?」『日本教育工学会論文誌』26, pp.33-44, 2002
- (12) 石田潤・森敏昭「小学生の文章表現の発達的变化」『広島大学教育学部紀要』第1部, 33, pp.125-131, 1985
- (13) 吉川愛弓・岸学「作文の評価項目に関する検討—意見文の評価は何に影響を受けるのか—」『東京学芸大学紀要 総合教育科学系』57, pp.93-102, 2006
- (14) 平直樹「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, pp.134-144, 1995
- (15) 岸学・綿井雅康「手続的知識の説明文を書く技能の様相について」『日本教育工学会論文誌』21, pp.119-128, 1997
- (16) 藤村宣之「概念(認知)地図」森敏昭・秋田喜代美(編)『心理学キーワード』有斐閣, pp.206-207, 2006
- (17) 鈴木敬司『「作文力」—そのウィーク・ポイント—』『月刊国語教育研究』140, p.12, 1984
- (18) 森岡健二『文章構成法 文章の診断と治療』至文堂, pp.421-438, 1977
- (19) 野地潤家・安岡章太郎 他24名『中学校国語 1～3』学校図書, 2007
- (20) 木下順二・加藤周一 他32名『伝え合う言葉 中学国語 1～3』教育出版, 2007
- (21) 宮地裕『国語 1～3』他32名 光村図書, 2005
- (22) 豊田秀樹『違いを見ぬく統計学 実験計画と分散分析入門』講談社, pp.222-246, 1994
- (23) 山内香奈「論文評定データの解析における多相Raschモデルと分散分析モデルの比較」『教育心理学研究』47, pp.383-392, 1999