

博士学位論文審査要旨

申請者：松村香奈

論文題目：A Mixed Methods Study on Formative Assessment of EFL Writing of College Students in Japan: Focusing on the Effectiveness of Schematized Teacher Feedback on Coherence

主査：澤木泰代 早稲田大学 教育・総合科学学術院教授 Ph.D. カリフォルニア大学

副査：原田哲男 早稲田大学 教育・総合科学学術院教授 Ph.D. カリフォルニア大学

副査：佐々木みゆき 早稲田大学 教育・総合科学学術院教授 Ph.D. カリフォルニア大学

副査：中野美知子 早稲田大学 名誉教授 Ph.D. エディンバラ大学

※2023年4月29日に公開発表会を Zoom により実施した。

1 本論文の目的と概要

本論文の目的は、日本語母語話者である日本の大学生を対象に、学生の書いた英語論証文の構造を樹状図 (tree-diagram) の形式で図示化し、教室での形成的評価における教員フィードバックとして提供し、その有用性を探究することである。英語学習者による英語での論証文の構成 (organization) 評価は、一貫性 (coherence)、結束性 (cohesion)、論理的展開 (logical development)、主題-題述構造 (theme-rheme structure) など、様々な構成要素から成る。そのため、その評価および評価内容の結果の提示が難しいことが指摘されている (Knoch, 2007; Lee, 2002)。教室における英語学術ライティング教育では、初稿の書き直し作業を繰り返して英作文を完成させるプロセスライティングが主な手法の1つとして用いられているが、その際の修正作業における適切かつ理解しやすい評価結果に関する情報 (フィードバック) の提示は、学習者の修正作業において重要な役割を果たす。教室での教員による従来のライティング・フィードバックは、学習者の英作文に対し、教員がテキストに直接下線や色付け、矢印などを施し、修正案やコメントを書き込む方法が一般的である (Biber et al., 2011)。一方で、分析的評価尺度や診断的評価による詳細な記述子をフィードバックとして提示する方法も試みられている (Jacobs et al., 1981; Kim, 2011)。しかし、これらの分析的・診断的評価は、表現が抽象的な場合があること、あるいはテキスト全体に対する総評であるため、具体的な原文の問題となる箇所を特定し、修正することが難しい。そこで本研究では上記を研究の背景とし、学習者の書いた英語論証文の構造を樹状図化できるソフトウェアを用いてフィードバックを提供し、「形成的フィードバック導入が学習者に与える影響」の検証という言語評価的視点から実証研究を行った。談話注釈タグ付けツール (discourse annotation tool, 以下アノテーション・ツールとする) を用いることで、教員による「構成」に関する分析的評価を可能にし、分析結果として得られた樹状図を形成的フィードバックとして用いることが、「構成」の中でも特に「一貫性」に対する学習者の理解を促し、修正作業を効率的に促進させるという仮説を基に、三つの研究課題に沿って検証した。本研究では、英作文能力への影響と学習者の作文修正行動や情意への影響の両面から探究する必要があるという立場から、介入を伴う収斂的混合研究法 (convergent mixed-methods design) を用い、量的・質的分析研究の双方向から検証を試みた。量的分析研究

においては、分析的評価尺度の信頼性についても検証した。研究全体の設計は、言語テスト妥当性検証の枠組みの一つである Assessment Use Argument (AUA; Bachman & Palmer, 2010) に基づいて行った。なお、有用性の検証は、図示化された形成的フィードバックを提示された介入群と、従来のテキストにコメントや修正案を書き込まれた従来型のフィードバックが提示された対照群の比較によって行った。

量的分析研究結果においては、混合型多変量分散分析の結果、初稿、初稿修正、および転移・保持 (transfer/retention) の3回のタスク間の主効果は有意であったが、介入群と対照群の2群間での主効果では、統計的有意差は確認されなかった。ただし、難易度の高い方の課題Bでの交互作用の結果が有意であったため、単純主効果の検定の結果、転移タスクでの語彙項目および初稿修正タスクでの言語使用項目において、介入群が対照群より有意に得点が高かったのは注目すべき点である。また、初稿修正タスクの作業時間における両群の比較を行ったところ、課題A、Bともに両群での産出語数と得点と同程度であるにもかかわらず、介入群の作業時間が大幅に短く、効率的に作業が行われたことが示された。次に質的分析研究結果では、アノテーション・ツールでのタグ付けで、一貫性の乱れ・破調 (coherence break/anomaly) と判断された文 (sentence) の数が対照群の方が多くに加えて、その内容において「論理的展開の不味さ」や「不適切な支持文」といった項目において有意に数が上回ることがわかった。標本サイズが小さいため評価の判断は慎重であるべきだが、一考に値する結果である。同じく、質的分析結果において、転移タスクでの介入群について、特徴的にバランスの取れた樹形図の形状の多いことが確認された。また、修正作業後の質問紙自由記述回答のテーマ分析において、教員のフィードバックに関する「わかりやすさ」や「構成」や「一貫性」についての気づきの多さや「客観的視点の獲得」などが介入群の回答に特徴的であった。

以上、量的および質的分析研究結果を総合的に解釈すると、日本語母語話者である日本の大学生を対象に、学生の書いた英語論文の構造を樹状図の形式で図示化、教室での形成的評価における教員によるフィードバックを行うことは、英語論文の作文能力の得点評価においては、総合的にみて必ずしも有用性は確認できなかったが、書き直し作業における時間対効果が上がることで、また点数には表れにくい一貫性の破調といったテキスト単位での評価、あるいは図示化されたテキストの構造の形状といった点、更には学生の英作文における「構成」や「一貫性」への理解や気づきといった情意面においてもその有用性については評価できるという結論に至った。これらの効果の要因として、推察ではあるが、図示化した形成的フィードバックは、修正の際の読み直し作業において学習者の認知的負荷が軽減されることで、作業時間の短縮を促し、「構成」以外の「語彙」や「言語使用」に対する作文能力資源の分配ができる可能性が示唆されたのではないかと考える。

2 本論文の構成と目次

Chapter 1 INTRODUCTION

1.1 Background and Research Purpose

1.2 Terminology Related to Coherence

1.3 Issues in Formative Assessment in EFL Writing Classes

1.3.1 What Is Required for Formative Feedback in Classroom-based Assessment

1.3.2 Issues in Writing Feedback in Terms of Assessment of Coherence

1.4 Theoretical Background and Research Design of the Present Study

- 1.4.1 Assessment Use Argument as the Validity Framework
- 1.4.2 AUA Claims
- 1.4.3 Mixed Methods Design in the Study
- 1.5 Overview of the Dissertation

Chapter 2 LITERATURE REVIEW

- 2.1 The Current Issues of EFL Writing in Japan
 - 2.1.1 Background of EFL Writing Instruction in Japan
 - 2.1.2 The Writing Instructions in Upper Secondary Schools
 - 2.1.3 The Writing Instruction in Universities
 - 2.1.3.1 Target Language Use and Admission Tests in Japanese University
 - 2.1.3.2 Writing Instructions in EFL Writing Classes and EFL Students' Perceptions of Their English Proficiency in Universities in Japan
- 2.2 L2 Writing Feedback
 - 2.2.1 Overview of the Trends of L2 Writing Instructions and Feedback
 - 2.2.2 Types of Research Design Incorporating Revisions Based on Feedback
 - 2.2.3 Types of L2 Writing Feedback
- 2.3 Diagnostic Language Assessment on EFL Writing
 - 2.3.1 Direct and Indirect Writing Tests
- 2.4 Coherence in Writing
 - 2.4.1. L2 Organizational Pattern or Logical Development and Contrastive Rhetoric
 - 2.4.2 Conceptual Definitions of Coherence Through Descriptors in Language Assessment Framework
 - 2.4.3 Conceptual Definitions of Coherence in L2 Writing Studies
 - 2.4.3.1 Writing Connectors in Metadiscourse Perspectives
 - 2.4.3.2 Coherence in Topical Structure Analysis
 - 2.4.3.3 Development of Checklists for Diagnostic Assessment of L2 Writing
 - 2.4.3.4 Taxonomy of Knowledge-Based Academic Writing Skills
 - 2.4.3.5 Defining L2 Writing Components Through Raters' Decision-making Behaviors
 - 2.4.4 Identifying Coherence Break/Anomalies in Formative Assessment
 - 2.4.4.1 Coherence Breaks/Anomalies in L2 Writing
 - 2.4.4.2 Identifying Coherence Breaks/Anomalies in L2 Writing With Annotation Tool
- 2.5 Text Annotation Tools and Graphic Displays as Structural Representation
 - 2.5.1 Graphic Displays in Language Learning
 - 2.5.1.1 Theoretical Background of Graphic Displays in Learning
 - 2.5.1.2 Graphic Organizer as a Scaffolding for Reading Comprehension
 - 2.5.1.3 Graphic Displays with Annotation Tool for Analysis of L2 Learners' Written Products by Researchers
 - 2.5.1.4 Differences Between Graphic Displays Between Graphic Organizer in Reading Research and Annotation Diagram in Writing Research
 - 2.5.1.5 Types of Annotation Tool for Generating Annotation Diagrams
- 2.6 Relevant Rating Scale for Diagnostic Assessment
 - 2.6.1 The Target of the Rating Scale: (a) Who is the Rating Scale for?
 - 2.6.2 Types of Rating Scales Desired:(b) What Aspects of Writing Are Most Important, and How Will They Be Divided up?
 - 2.6.3 The Basis of the Criteria (c) What Are the Criteria Based on?
 - 2.6.4 Descriptors and Scoring Levels in the Rating Scale: (d) What Will the Descriptors Look Like and How Many Scoring Levels Will Be Used?
 - 2.6.5 Score Report Obtained by Rating Scale: (e) How Will Scores Be Reported?
- 2.7 Study Hypothesis and Research Questions

Chapter 3 METHODOLOGY

- 3.1 Student Participants
- 3.2 Materials
 - 3.2.1. Writing Tasks

- 3.2.2 Rating Scale
 - 3.2.2.1 The Base Rating Scale
 - 3.2.2.2 Rating Scale in the Present Study
 - 3.2.2.2.1 Modification of the Analytic Rating Scale
 - 3.2.2.2.2 Conversion of Raw Scores Into Level Sores
 - 3.2.2.2.3 Modification of the Descriptors in the Rating Scale
- 3.2.3 Raters and Task Assignment
 - 3.2.3.1 Raters
 - 3.2.3.2 Rater Training Sessions and Post-Evaluation Interview
 - 3.2.3.3. Rater Task Assignment
- 3.2.4 Annotation Tool
- 3.2.5 Annotators
- 3.2.6 Teacher Feedback (Control and Intervention Groups)
- 3.2.7 A Sample of Schematized Feedback Sheet (Intervention Group)
- 3.3 Procedure
 - 3.3.1 Research Design
 - 3.3.1.1 Mixed-methods Research Design
 - 3.3.1.2 Establishment of Research Questions and Procedure in the AUA Framework
 - 3.3.2 The Procedure in Terms of Research Questions
- 3.4 Data Analysis
 - 3.4.1 Quantitative Analysis (Preliminary Study, Study 1, and Study 2)
 - 3.4.1.1 Preliminary Study: Conversion of Analytic Sores to Six-Level Scores
 - 3.4.1.2 Many-Facet Rasch Measurement (MFRM) (Preliminary Study)
 - 3.4.1.3 Parallel Coordinate Plots
 - 3.4.1.4 Mixed-Between-Within Multivariate Analysis of Variance (MANOVA)
 - 3.4.1.5 Operation Time for Each Task Among the Group (Study 2)
 - 3.4.2 Qualitative Analysis (Study 3)
 - 3.4.2.1 Coding Scheme Based on Rhetorical and Ideational Anomaly Detection
 - 3.4.2.2 Thematic Analysis on Students' Responses on the Questionnaire

Chapter 4 RESULTS

- 4.1 Quantitative Research Results
 - 4.1.1 Quantitative Study Results for Study 1: RQ 1, RQ2, RQ3(1), and (2)
 - 4.1.1.1 Analysis of Appropriateness of the Rating Scale
 - 4.1.1.1.1 Assumption Checks for the MFRM Analysis
 - 4.1.1.1.2 Examining Each Fit Statistics in MFRM
 - 4.1.1.1.3 The Wright Map for the Joint Calibration of the Four Facets
 - 4.1.1.1.4 Functioning of the Scale
 - 4.1.1.2 Measuring Difference in Difficulty Between Task A and B in Pre-condition
 - 4.1.1.3 RQ 2: Interpretation of Students Overall Performance in Terms of Composite Score
 - 4.1.1.3.1 Descriptive Statistics over Three Writing Occasions with Total Scores
 - 4.1.1.3.2 Parallel Coordinate Plots with Total Scores Over Three Occasions in Tasks and Between Groups
 - 4.1.1.4 RQ 3: Examining the Effects of Teacher Feedback Type on Learner Performance Scores
 - 4.1.1.4.1 Mixed Between-within MANOVA
 - 4.1.1.4.2 Primary Assumption Checks for Mixed MANOVA
 - 4.1.1.4.3 Descriptive Statistics
 - 4.1.1.4.4 Mixed MANOVA Results on Analytic Scores
 - 4.1.1.4.5 Topic A Results
 - 4.1.1.4.6 Topic B Results
 - 4.1.1.4.7 Control vs. Intervention in Terms of Integration of the Four Dependent Variables
 - 4.1.1.4.8 Summary of the Results Obtained From Mixed MANOVA
 - 4.1.2 Quantitative Analyses Results for Study 2: RQ3(2)
 - 4.1.2.1 Comparison of Student Revision Work Time

- 4.1.3 Summary of the Quantitative Analyses
- 4.2 Qualitative Research Results
 - 4.2.1 Qualitative Analyses Results for Study 3-1
 - 4.2.1.1. A Taxonomy of Ideational and Rhetorical Coherence Summarized for the Study
 - 4.2.1.2 Analysis of Anomalies Identified by the Annotators with the Annotation Tool
 - 4.4.1.3 Relationship Between Organizational Elements and Anomalies for All Student Transfer Task Writing Samples
 - 4.4.1.4 Probable Cause, Location and Frequency of Anomalous units Tagged as "Questionable" (Control vs. Intervention)
 - 4.4.1.5 Causes of Anomalies (Control Versus Intervention)
 - 4.2.2 Linking Quantitative and Qualitative Data for Interpretation
 - 4.2.2.1 Relationship Between Anomalies and Analytic Writing Scores Focusing on Organization and Content Criteria
 - 4.2.2.1.1 Relationship between Organization and Content Criteria in Terms of Frequency of Anomalies
 - 4.2.2.2 Relationship Between the Location of Occurrence of Organizational Elements of Anomalies and the Organization and Content Scores
 - 4.2.2.3 Relationship Between the Causes of Anomalies and the Organization and Content Scores
 - 4.2.2.3.1 Category 1: Reasoning, Logic, or Topic Development
 - 4.2.2.3.2 Category 2: Deviation From the Topic.
 - 4.2.2.3.3 Category 3: Redundant Ideas
 - 4.2.2.3.4 Category 4: Lack of Organizational Constituent(s)
 - 4.2.2.3.5 Category 5: Misunderstanding the Task/Task Incompletion
 - 4.2.2.3.6 Category 6: Inappropriate Support Due to Insufficient Knowledge on the Subject – Relevance of Support
 - 4.2.2.3.7 Category 8: Insufficient Rebuttal
 - 4.2.2.3.8 Category 11: Incomprehensible Sentence – Related to Language Use
 - 4.2.2.4 Analysis of the Overall Structure/Shape of Annotated Tree-Diagrams
 - 4.2.2.4.1 Examples of Tree-diagrams of Samples That Received High Scored Writings
 - 4.2.2.4.2 Examples of Tree-diagrams of Samples That Received Low Scores
 - 4.2.2.5 Comparison of the Tree diagrams Between the Control and Intervention Groups
 - 4.2.3 Summary of Results of Study 3-1 (Qualitative Results with Annotation Diagrams)
 - 4.2.4 Qualitative Analyses Results for Study 3-2
 - 4.2.4.1 Students' Behaviors during Rewriting Based on the Responses to a Questionnaire (Control vs. Intervention)
 - 4.2.4.1.1 What the Students Revised in the Second Draft with Feedback
 - 4.2.4.2 Student' Perceptions on Teacher Feedback
 - 4.2.4.3 Creating a Tentative Story Line Derived From the Summary of Themes Summary of Themes

Chapter 5 DISCUSSION

- 5.1 Introduction
- 5.2 RQ 1: The Reliability of the Rating Scale and Annotation Scheme (Claim 4 in AUA)
- 5.3 RQ 2: The Meaningfulness of Interpretation of Students' Performance Over Writing Occasions (Claim 3)
- 5.4 RQ 3: The Effectiveness of Schematized Teacher Feedback (Claim 1 in AUA)
 - 5.4.1 RQ 3-1 Performance on Writing Skills Criteria Across the Occasions (Claim 1 of the AUA)
 - 5.4.2 RQ3-2 Revision Time on Task (Claim 1 of the AUA)
 - 5.4.3 RQ3-3 Ideational and Rhetorical Coherence in Transfer Task (Claim 1 of the AUA)
 - 5.4.4 RQ3-4 Students' Rewriting Behaviors and Perceptions (Claim 1 of the AUA)
- 5.5 Limitation of the Study
- 5.6 Pedagogical Implications
- 5.7 Directions for Further Research
- 5.8 Summary of Discussion

Chapter 6 CONCLUSION

References

Appendices

3 各章の概要

第1章 序文

日本語母語話者である英語学習者が英語で自分の意見を記述する論証文では、母語とは異なる言語体系に則った言語的正確さが求められるだけでなく、英語に特有の談話形式に従って論を展開する必要がある (Connor, 1996; Kaplan, 1966; Oi, 1986, 1999, 2005)。日本の英語学習者の多くは、大学の英語学術的ライティング授業におけるパラグラフライティング (paragraph writing) で、主題文、支持文、結語文の基本的な論理の展開を系統的に学習することになる。英語学習者が、いわゆる英語らしい論理的構成で、論証文を書く際に問題になるのは、文化的な差異だけでなく、主語と述語の選び方、言語情報の提示する順序や構成、接続詞の選択や文と文のつながり方などといった、言語的な差異を踏まえ、まとまりや一貫性のある (coherent) 英語パラグラフを書く難しさがある。それに加え、ライティングにおける構成 (organization) 評価は、様々な評価観点から成るため、教員や評価者の理解も十分であるとは言えない。一貫性はライティングの質を確保するために効果的な要素である (Bamberg, 1984; Richards, 1990; Wilkinson, 1990) としながらも、その評価および評価結果を効果的に提示することの難しさが指摘されており (Knoch, 2007; Lee, 2002)、適切な形成的評価に基づくフィードバック作成による介入が望まれる。一方、近年の日本の英作文教育においては、新学習指導要領 (2018) の高校での達成目標として、詳細な根拠を伴った論理的な展開で構成の論証文を産出する能力の育成を掲げており、英作文における論理的破綻や逸脱のない一貫性のある論証文を産出する力の育成が課題となっている。本研究では、「構成」の要素である「一貫性」を、他の文の解釈に基づく談話の意味的特性 (van Dijk, 1977) とした上で、文章が適切な順序で配置され、切れ目なく繋がり、文の間で淀みなく次々とアイデアが提示されることと定義し、学生への英語ライティング指導における形成的評価としての図示化された形成フィードバックが学生の英作文能力および作文修正行動や一貫性の観点への理解に及ぼす影響を探究した。

第2章 先行研究

第2章では、第一に、主に形成的・診断的評価の観点から、Black & Williams (1998) や Alderson (2015) による先行研究に基づき、本研究での学習者の作文能力の改善に求められる適切かつ効果的な形成的フィードバックについて論じた。また、ライティング評価における構成や一貫性の定義を明確にするため、ヨーロッパ言語共通参照枠 (以下 CEFR とする) ならびに日本の主要な標準化テストでのライティングテストの構成、一貫性に関する評価尺度における記述子を示し、具体的にどのような基準で評価がなされているのかを比較分析した。評価記述子の先行研究の比較分析から、論理的・一貫性を評価する記述子は CEFR の B1 レベル以上に求められる能力として出現し、さらには、「詳しく (detailed)」や「十分に構成が整った (well-structured)」といった、形容詞や副詞を用いた程度を表わす抽象的な表現が用いられていることが分かる。テストでのフィードバックコメントはこういった記述子を基に作成されているこ

とから、学生には、実際の修正行動で参照できる具体的なフィードバックの必要性が確認された。次に、教育における図の提示の効果について先行研究を論じた。読解におけるテキストの関係性を表す図の提示（graphic organizer）の教育的効果の先行研究では、図示化により学習者の認知負荷の軽減が予想されることが指摘されている（Vekiri, 2002）。ライティングにおける修正作業は初稿を読み直す作業であることから、本研究による図示化のフィードバックへの効果も期待できた。具体的なテキストの図示化については、英語教育での談話分析でも良く知られる修辞構造理論（Rhetorical Structure Theory; Mann & Thompson, 1988）および本研究で使用した Tool for Interactive Argument Annotation（TIARA, Putra et al., 2020）他を比較し、その操作性、利便性、機能性から TIARA を採用した。量的評価のための評価尺度選定に関する先行研究から、本研究では Jacobs et al. (1981) の評価尺度に修正を加えたものの採用を決定した。以上の先行研究を踏まえ、本研究では以下三つの研究課題を設定した。

（1）本研究で使用した評価尺度に基づく採点結果は一貫しているか。

（2）評価尺度による得点とアノテーション・ツールでの談話注釈樹状図を通して、生徒の総合的なライティング・パフォーマンスと構成はどのように特徴付けられ、解釈され得るか。

（3）教師からのフィードバックの種類（従来型か図示型か）は次の4つの観点に対し、どの程度、どのような影響を与えるか。

①初稿、初稿修正、転移タスクの各場面における各ライティング観点での得点

②初稿修正タスクでの作業所要時間

③転移タスクにおける内容・構成視点（ideational and rhetorical）での一貫性（アノテーション・ツールによる情報分析を通じての分析）

④学生のライティング修正タスクでの行動と認識

第3章 研究方法

第3章では、本研究の研究・調査方法を記した。本研究には、東京の私立大学で教育学を専攻する大学2年生ニクラス計50名（うち研究対象者45名）と大学で英語ライティングを教えた経験のある6名の日本人英語教員が調査に参加した。ライティング課題として、難易度の異なる英検2級（課題A）と英検準1級（課題Bおよび転移課題C）から学生の興味関心に配慮したテーマ（大学生が社会人になった際に考える内容）を選んだ。また、質的・量的研究法のそれぞれの長を活かしながら分析結果を統合するために、介入を伴う収斂的混合研究法（convergent mixed-methods design）を採用した。

量的研究の研究課題1においては多相ラッシュモデルを用い、本研究のために既存の尺度に修正を施した評価尺度について、評価者間の採点やの信頼性やおよび学生の能力やタスク難易度の観点での尺度使用の適切さを分析し、評価尺度がうまく機能しているかを検証した。また、研究課題2では、総合的な得点の特徴の分析および解釈のため記述統計に基づく平行座標プロットを分析した。また、図示化フィードバックの効果の分析（研究課題3）のためには介入群と対照群（2水準）での従属変数4観点（内容、構成、言語使用、語彙）における採点結果比較を初稿、初稿修正、転移タスクの反復測定（3水準）で行い、分析には2要因混合型多変量分散分析の結果を用いた。なお、分析は難易度の異なる2つの課題A、Bで個別に実施した。一方、質的研究においては、二種の質的データを以下の方法で分析した。まず、アノテーション・ツールを用いた学習者のテキストを注釈タグ付けした上で文同士の関係を樹状図に図示化し、タグ

付けから一貫性の破調（coherence anomaly）特定および構成の形状の特徴を分析した。また、両群の修正行動とフィードバックの受容の違いを確認するため、自由記述での質問紙から得られた回答をテーマ分析の手法で探索的に分析、検証した。分析の途中および最終段階で、質的・量的研究分析結果を統合して課題の検証を行った。

第4章 分析結果（量的および質的研究）

第4章では、前章で示した手法に従って行った結果をまとめた。混合型多変量分散分析の結果、反復測定である3回のタスク間での主効果は有意であったが、介入群と対照群の2群間での主効果では、統計的有意差は確認されなかった。ただし、課題Bでの単純主効果の検定の結果、転移タスクでの語彙項目および事後タスクでの言語使用項目において、介入群が対照群より統計的に有意に得点が高かったのは注目すべき点である。次に、量的結果として、修正タスクの作業時間における両群の比較を行ったところ、課題A、Bともに介入群と対照群の産出語数と得点は同じにもかかわらず、介入群の作業時間が大幅に短く、効率的に作業が行われたことがわかった。次に質的分析研究結果では、アノテーション・ツールでのタグ付けで一貫性の乱れ・逸脱（coherence break/anomaly）と判断されたテキスト単位である文（sentence）の数が対照群の方が多いことのみならず、更には、その内容において「論理的展開の不味さ」や「不適切な支持文」といった項目で有意に数が上回ることがわかった。標本サイズが小さいため評価の判断は慎重であるべきだが、一考に値する結果である。また、産出された転移タスクにおける樹形図の形状において、介入群において特徴的にバランスの取れた整った形状が多いことが確認された。修正作業後の質問紙自由記述回答のテーマ分析においては、介入群が修正作業内容の報告の回答数および修正の種類において、対照群より有意に高い項目があること、更には、教員のフィードバックについて、「わかりやすさ」や「構成」や「一貫性」についての気づきの多さや「客観的視点の獲得」などが介入群の回答に特徴的に観られたことなど、介入群での修正作業や情意での有用性が明らかになった。

第5章 考察

研究課題3の①における、図示化の形成フィードバックの量的分析（混合型多変量分散分析）による有用性については興味深い結果となった。同分析においては、介入群と対照群の間には統計的に有意な差は確認されなかった。しかし、単純主効果の検定で、課題Bでの初稿修正タスクにおける「言語使用」観点および転移タスクにおける「語彙」の観点で、介入群が対照群より有意に得点が高かった点は注目に値する。図示化フィードバックの効果として、「構成」観点での得点での差を予想していたが、実際には「言語」と「語彙」の得点で差があったからである。この現象の原因を特定するのは難しいが、推測としては、フィードバックに使われた樹形図がテキスト単位で区切られているため、一文に注目しやすく、文レベルでの誤りを見つけやすかった可能性が考えられる。また、同様の理由から個々の単語により注意が払われた結果、初稿および初稿修正タスクの2種類での語彙使用の定着により、語彙の再利用が行われ、転移タスクでの語彙得点が高かった可能性もある。このことは、質問紙自由記述回答で、修正箇所を思い出して挙げてもらう質問に対して、介入群の方がより頻繁に、より詳細に回答していることから、介入群で、ライティング作業での学習内容がより深く印象に残って

いる可能性があることと関連しているのかもしれない。しかしながら、この要因の解釈の適切性を確認するためには、更なる調査が必要となる。また、研究課題3の④における、学習者の転移タスクにおける行動の変化として、作文課題の計画の段階で樹状図を作成して構成を考えていた学生の存在が確認された点も興味深い。本研究の限界点の一つは、教室での形成的評価研究であるため、著者が授業者である既存クラス2つを比較した点である。この研究デザインには、学習者を詳細に観察できる利点があると同時に、バイアスがかからない工夫をすることの難しさが挙げられる。更に、アノテーション・ツールによるフィードバック作成には一定時間を要する点も、教員負担の観点から実際の教室での使用では留意する必要がある。今回の研究では、難易度の異なる課題2種で実施をしたため、2回のフィードバック作成となったが、学生のコメントからも、図示化フィードバックは強く印象に残ることから、1学期に1回程度の実施でも効果があると考えられる。

第6章 結論

本研究では、日本語母語話者である日本の大学生を対象に、これまで教室での英語作文教育において、指導と学習、さらには評価や評価の結果の効果的な提示に難しさが指摘されていた「構成」における「一貫性」の観点で、プロセスライティングの修正作業に際して教員による形成的・診断的評価およびフィードバックにおいて、アノテーション・ツールを用いて図示化した文の関係性を樹状図で表したものを提供することの有用性について探求した。量的および質的分析結果を統合する混合研究法により、図示化による形成的フィードバックが初稿修正文や転移タスクのライティングに与えるさまざまな肯定的な影響が明らかになった。日本語母語話者である日本の大学生を対象に、学生が書いた英語論証文の構造を樹状図の形式で図示化、教室での形成的評価における教員によるフィードバックを行うことは、英語論証文の作文能力の得点評価において内容や構成の観点では対照群と比較して有用性における統計的有意差は確認できなかったが、初稿修正タスクでの言語使用ならびに転移タスクの語彙では有意差が確認された。書き直しにおける作業における時間対効果が上がること、点数には表れにくい、一貫性の破調といったテキスト単位での評価、あるいは、図示化されたテキストの構造の形状といった点、更には、学生の英作文における「構成」や「一貫性」への理解や気づきといった情意面においてもその有用性については評価できるという結論に至った。

参考文献

- Alderson, C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Bamberg, B. (1984). Assessing Coherence: A Reanalysis of Essays written for the National Assessment of Education progress, 1969-1979. *Research in the Teaching of English*, 18 (3), 305-319. National Council of Teachers of English. <https://www.jstor.org/stable/40171021>
- Biber, D., Nekrasova, T., & Horn, B. (2011). The Effectiveness of Feedback for L1-English and L2-Writing Development: A Meta-Analysis. *TOEFL iB™ Research Report, TOEFL iBT-14*. <https://doi.org/10.1002/j.2333-8504.2011.tb02241.x>
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. DOI: 10.1080/0969595980050102

- Connor, U. (1996). *Contrastive Rhetoric: Cross-cultural aspects of Second-language writing*. Cambridge University Press.
- Jacobs, H., Zinkgraf, S. Wormuth, E., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16, 1-20. <https://doi.org/10.1111/j.1467-1770.1966.tb00804.x>
- Kim, Y. H. (2011). Diagnosing EAP writing ability using the Deduced Reparameterized Unified Model. *Language Testing*, 28 (4), 509-541. DOI: 10.1177/0265532211400860
- Knoch, U. (2007). Diagnostic writing assessment: the development and validation of rating scale. [Doctoral dissertation, University of Auckland]. Researchgate. <https://www.researchgate.net/publication/37986429>
- Mann, W., & Thompson, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization, *Text ---Interdisciplinary Journal for the Study of Discourse*, 8(3), 243-281. Retrieved from https://www.sfu.ca/rst/05bibliographies/bibs/Mann_Thompson_1988.pdf
- Oi, K. (1986). Cross-cultural Differences in Rhetorical Patterning--A Study of Japanese and English. *JACET Bulletin*, 17, 23-48. Retrieved from
- Oi, K. (1999). Comparison of argumentative styles: Japanese college students vs. American college students--An analysis using the Toulmin model. *JACET Bulletin*, 30, 85-102. Retrieved from <https://dl.ndl.go.jp/pid/10501296/1/1https://dl.ndl.go.jp/pid/10500387/1/1>
- Oi, K. (2005). Teaching argumentative writing to Japanese EFL students using the Toulmin model. *JACET Bulletin*, 41, 123-140. Retrieved from <https://dl.ndl.go.jp/pid/10501483/1/1>
- Putra, J. W.G., Teufel, S., Matsumura, K. & Tokunaga, T. (2020). TIARA: A Tool for Annotating Discourse Relations and Sentence Reordering. *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, pp. 6912–6920.
- Richards, J. C. (1990). *The Language Teaching Matrix*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667152>
- van Dijk, T.A. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. Addison-Wesley Longman.
- Vekiri, I. (2002). What Is the Value of Graphical Displays in Learning? *Educational Psychology Review*, 14(3), 261-312. <https://doi.org/10.1023/A:1016064429161>
- Wilkinson, A. M. (1991). *The scientist's handbook for writing papers and dissertations*. Edgewood Cliffs, NJ: Prentice Hall.

4 総評

本研究は、大学における英語ライティング教育基礎科目において広く指導対象となる論証文パラグラフ作文指導において、学習者が作成したパラグラフ内の文同士の関係を樹状図として示すフィードバック方法と教員によるコメント書き込み型の従来のフィードバック方法を、作文の「一貫性」や作文修正行動への効果の点から比較するものである。本研究はこの2つのフィードバック方法（樹状図あるいは書き込み）を形成的評価の手段と位置付け、言語テスト妥当性検証の枠組みである Bachman and Palmer (2010)の Assessment Use Argument (AUA) を基に計画・実施されている。この2タイプのフィードバック方法の機能を比較検証するうえでの AUA を用いた研究課題設定は適切で、収斂的混合研究法 (convergent mixed-methods study) により量的研究と質的研究手法を組み合わせ、各研究課題について多角的かつ精緻なデータ分析を重ねている。AUA に基づく結果の総括も有効で、十分なエビデンスと共に本研究の形成的評価の文脈での樹状図フィードバック使用は概ね支持できるとの結論を導き出している。

本論文は特に以下の点で優れていると評価する。その一点目は、第2章の先行研究調査において、第二言語ライティング評価における「一貫性」の構成概念を、膨大な資料の整理により

明確にしたことである。「一貫性」は「構成」「結束性」「論理展開」など様々な類似の構成概念と密接に関係しており、理論的定義も一様ではない。本論文では先行研究による理論的定義や分析的枠組みから英語外部テスト等で一般的に使用されている採点尺度における操作化に至るまで、資料に幅広くあたって関連の概念を整理したうえで本論文における「一貫性」の定義を明確に示し、それに見合ったフィードバック方法や分析方法を提案している。この面から、本論文は今後のライティング評価における「一貫性」の構成概念定義に貢献すると考えられると共に、効果的な作文に不可欠である一方で、その概念的な曖昧さゆえに難しいと捉えられがちな「一貫性」の指導に役立つ知見を提供するものであるといえよう。

本論文の強みの2点目は、樹状図を第二言語ライティング指導における学習者への個別フィードバック作成に応用したことの新規性と、本研究から得られた、第二言語習得研究の更なる発展に貢献しうる研究結果を得たことである。第二言語リーディング指導等においてテキスト構成の図示を学習者への足場かけとして活用する先行研究事例があるが、ライティングやスピーキングにおいて学習者に個別フィードバックを与える手段としての樹状図の利用は他に例を見ない。過去には樹状図を作成するソフトウェア自体が重く、研究のためのツールにはなりえても普通の授業による利用は現実的ではなかった。しかし、本論文では自然言語処理研究において近年開発され、容量が軽いソフトウェア TIARA (Putra et al., 2020) を使用している点で、まずライティング指導における樹状図フィードバック実践例としての価値が見いだせる。また、学生が作文したパラグラフを樹状図で表した際に見られた上位群と下位群の違い、特に上位群はバランスの取れた樹状図が多かった一方で、下位群の樹状図には主に「平たい」「バランスが取れていない」「縦に長い」などのタイプがあることを特定した。これは先行研究では指摘されていない、樹状図を使用した研究であるからこそ得られた知見であり、学習者作文の「一貫性」に関しての理論的発展に貢献しうるものである。

一方で、審査においては更なる検討を要する点も指摘された。一点目は第二言語習得の観点から興味深い結果を複数得ている一方、その理論的意義に関する指摘が十分ではなかった点である。上述の受講者のライティングカレベルと樹状図の特徴の関係性をはじめ、樹状図の提示による認知負荷の軽減が作文修正時間の短縮や作文の推敲促進につながる可能性、また樹状図提示群では後に転移・保持タスクにおいて、学習者が自ら樹状図を描くという学習方略の使用が確認されるなど、本研究では複数の方向に研究の発展を期待できる結果を得ている。その理論的貢献や示唆に関する考察については一層の充実を求めたい。2点目は、著者自身が今回研究対象となった2クラスの授業担当教員であったことである。本研究の量的分析で従属変数とした採点尺度による評価は独立した6名の採点者が行い、十分な採点者間一貫性も得られている点では、量的分析結果への直接的影響は少ないと考えられる。一方で、指導者が学習者をよく知ることが学習者のパフォーマンスや教員の指導法の質的变化につながる可能性は否めず、本論文の結果の一般化可能性については追試での確認が不可欠であろう。しかし、これらは本論文の質を脅かすものではなく、今後の研究の中でより発展させていくべき事項と位置付けられた。

以上の観点から、審査員一同、本論文が「博士（教育学）」に値するという結論を得た。