

博士学位論文審査要旨

申請者：大井洋子

論文題目：Efficacy of Student Assessment as Part of English Writing Instruction for Japanese High School Students

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：マキュワン麻哉 早稲田大学 教育・総合科学学術院准教授

博士（外国語教育学）関西大学

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

※2021年5月8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、日本の高校生を対象に、英作文教育における生徒評価（student assessment）、つまり学習者が評価者として自身もしくは他者の学習成果を評価する活動の有効性をどのように英語授業内の形成的評価に活かすことができるかを探究することである。生徒自身による自己評価（self-assessment）や生徒間の他己評価（peer assessment）などの生徒評価の手法はこれまでも授業内に取り入れられてきたが、18歳以下の学習者、特に日本の高校生を対象とした英語教育に関する調査研究は少ない。また、自己評価と他己評価の類似性や相違点、それぞれの評価タイプが与える英作文や第二言語学習者の情意に与える影響、そして形成的評価との関係に関する先行研究は極めて限られている。一方では、日本の英作文教育については高校での授業回数が比較的少なく不活発であることや、生徒自身の英作文に対する興味が口語英語に対するものに比べて低いことが近年指摘されている。しかし、国際化が進む現代社会においては、英作文能力を向上させることが学校英語教育の責務の一つであると考えられる。そこで本研究では上記を研究の背景とし、形成的な生徒評価の授業への導入が高等学校における英作文教育改革の一助となる可能性に着目した。生徒評価（自己評価と他己評価）の有効性については、両者が与える英作文能力への影響と学習者の情意への影響の両面から探究する必要があるとの立場を取った。また生徒評価の授業への導入により生徒を評価者の立場に置き、学習と評価の責任を教員と生徒が共有することで学習に対する自律や自発が促され、英作文学習不安の減少や英作文能力の向上につながるなどの仮定を立て、二つの研究課題に沿って検証した。

本研究では収斂的混合研究法（convergent mixed-methods design）を用い、質的・量的分析研究の双方向から自己評価と他己評価を比較することで、それぞれの生徒評価手法の特性を明らかにし、形成的評価と生徒評価の関係を探究した。量的分析においては、自己評価と他己評価が共に文法の正確さと学習者の自立に望ましい影響があることが示された。また、自己評価と他己評価はいずれも教員評価ほど採点結果の信頼性は高くない一方、評価の厳格さにおいて自己評価は教員評価と同程度であるとの結果を得た。さらに、両評価手法とも、生徒評価活動前に自立性高かった生徒ほど、指導後の自立性が高い傾向にあることがわかった。質的分析研究結

果もこれらの量的分析研究結果を支持するものであった。自己評価は、学習者の気づきを高めると共に自己内省を促し、英作文力向上に肯定的に働く一方、他己評価は、英作文評価活動がクラス内の相互交流を促し、学習者のメタ認知力を高める可能性が示された。また、本研究では教師の介入は観察者としての役割のみだったことが、生徒評価に対する教師の迅速かつ適度な介入の必要性を浮き彫りにした。このように、高校生の英作文に対する自己評価と他己評価が与える英作文能力と情意への影響には、様々な類似点と相違点があることが明らかになった。また、上記の混合研究結果を形成的評価と生徒評価の関係に関する Leahy and Wiliam (2012) の五つの方策に従って整理した（学習目的や評価基準に対する明確な理解、学習効果の具体的な明示、学習を前進させるためのフィードバック、学習者同志を相互学習資源にすること、学習者を自身の学習の所有者にすること）。その結果、自己評価は外部のフィードバックがないために5つの内3つに強く、一方、他己評価は、全て方策にとって若干の弱さはあるものの全ての方策に有効であることがわかった。また、形成的評価は学習者と教師によって構成され、教師こそが生徒評価を授業内で有効に導入し活用できることから、教師がそれぞれの生徒評価の特質を理解することと、生徒の学習効果を高めるための適度な教師による介入が形成的評価を成功に導く鍵となるとの結論に至った。

2 本論文の構成と目次

Chapter 1 INTRODUCTION

1.1 Background and Study Aims

1.1.1 Learners as Agents of Formative Classroom Assessment

1.2 Why Student Assessment Influences Writing Ability

1.3 The Influence of Student Assessment on Learner Affect

1.4 Research Questions

1.5 Organization of the Dissertation

Chapter 2 LITERATURE REVIEW

2.1 Introduction

2.2 Background of the Study

2.2.1 The Framework of Assessment in the Classroom: Classroom-based Assessment

2.2.2 Formative Assessment

2.2.3 Teacher Assessment

2.2.4 The Implication of Student Assessment

2.3 Previous Studies on Student Assessment

2.3.1 Meta-analysis Results

2.3.2 Writing Assessment and Student Assessment

2.3.3 Holistic Assessment and Analytic Assessment

2.3.4 Student Assessment as Formative Assessment

2.4 Writing Anxiety

2.4.1 What is Foreign Language Learning Anxiety?

2.4.2 The Definition of Writing Anxiety

2.4.3 Previous Studies about L1 Writing Anxiety

2.4.4 Previous Studies about L2 Writing Anxiety

2.4.5 The Development of Scales for L1 Writing Anxiety

2.4.6 Scales of L2 Writing Anxiety

2.5 Learner Autonomy

2.5.1 Learner Autonomy Scales for L2 Learners

2.6 Study Hypotheses

2.6.1 How do Self- and Peer Assessment Compare with Each Other?

- 2.6.2 How can Student Assessment Work as Formative Assessment in the Classroom?
- 2.6.3 Research Questions and Study Hypotheses

Chapter 3 METHODOLOGY

- 3.1 Introduction
- 3.2 Participants
- 3.3 The English Curriculum
- 3.4 Materials
 - 3.4.1 Instrument Development
 - 3.4.2 The Writing Assessment Rubric
 - 3.4.3 Narrative Frame
 - 3.4.4 Writing Tasks
 - 3.4.5 Questionnaire on Writing Anxiety and Learner Autonomy
- 3.5 Procedure
 - 3.5.1 Research Design
 - 3.5.2 Phase 1
 - 3.5.3 Phase 2
 - 3.5.4 Phase 3
- 3.6 Analyses
 - 3.6.1 Quantitative Analysis
- 3.7 Synthesis of Quantitative and Qualitative Results

Chapter 4 QUANTITATIVE STUDY

- 4.1 The Reliability of Student Assessment
 - 4.1.1 Many-facet Rasch Analysis Results on the Composite Score across the Four Analytic Rating Scales
 - 4.1.2 Many-facet Rasch Analysis Results on the Analytic Rating Scales
 - 4.1.3 Conclusion of RQ1: How do Self- and Peer Assessment Compare with each other in terms of: a) the Reliability of Scores versus Teacher Assessment?
- 4.2 The Effects of Student Assessment on the Improvement of Writing Ability
 - 4.2.1 Introduction
 - 4.2.2 The Effects on the Improvement of Writing Ability
 - 4.2.3 Conclusion of RQ1: How do Self- and Peer Assessment Compare with each other in terms of: b) the Effects on Writing Ability?
- 4.3 The Effects of Student Assessment on Writing Anxiety and Learner Autonomy
 - 4.3.1 Introduction
 - 4.3.2 Preliminary analyses
 - 4.3.3 MANCOVA results
 - 4.3.4 Conclusion of RQ1: How do Self- and Peer Assessment Compare with each other in terms of: c) the Effects on Writing Anxiety and Learner Autonomy?
- 4.4 Summary of the Quantitative Analyses

Chapter 5 QUALITATIVE STUDY

- 5.1 Introduction
- 5.2 Analysis of Open-Ended Questionnaire Responses about the Benefits and Challenges of Student Assessment
- 5.3 Semi-structured Interviews with Students and Teachers
 - 5.3.1 Theoretical Codes for the Self-assessment Group
 - 5.3.2 Theoretical Codes of the Peer Assessment Group
 - 5.3.3 Comparison of the Semi-structured Interview Results between Self-assessment and Peer Assessment
- 5.4 Summary of the Qualitative Analyses

Chapter 6 DISCUSSION

- 6.1 Introduction

- 6.2 Research Question 1: How Do Self-Assessment and Peer Assessment Compare with Each Other?
 - 6.2.1 Hypothesis 1.1: Reliability of Student Assessment
 - 6.2.2 Hypothesis 1.2: Effects on the Improvement of Writing Ability
 - 6.2.3 Hypothesis 1.3: Effects on Writing Anxiety and Learner Autonomy
 - 6.2.4 Interpretation of the Results of Research Question 1
- 6.3 Research Question 2: How Can Student Assessment Work as Formative Assessment in the Classroom?
- 6.4 Summary of the Findings

Chapter 7 CONCLUSION

- 7.1 A Review of the Study Objectives
- 7.2 Major Findings
- 7.3 Formative Assessment
- 7.4 Efficacy of Student Assessment
- 7.5 Contributions and Implications of the Study
- 7.6 Limitations to the Study and Recommendations for Future Research
- 7.7 Conclusion

References

3 各章の概要

第1章 序文

本研究では、生徒評価を学習者が評価者として、自身もしくは他者の学習成果を評価することと定義する。生徒評価は、先行研究において学習者の学びに対する責任感を高め、学習意欲をも高めるものと論じられてきた。しかしながら、自己評価と他己評価の類似点や相違点は詳細かつ具体的には検証されてこなかった。また、その研究の大半は、大学生以上の学習者を対象としたもので、日本の高校生を対象とした分析は極めて少ない。特に、生徒評価それぞれのタイプが学習者の英作文を書く能力や情意に与える影響の比較に関する研究は少ない。そこで、生徒評価のそれぞれのタイプの英作文能力や情緒に対する効果に焦点を当て、本研究を進めることにした。本研究の背景には2点ある。第一に停滞する日本の英作文教育の一つの解決策として生徒評価の可能性があるのでないかという点、第二に国際化が進む世界に適応するためには英語で書く力をさらに伸ばす必要があるという点である。本研究では、生徒評価は、改革が必要とされる日本の英作文授業の一つの解決策と仮定した。これは生徒評価の導入は、学習者を受動的から能動的な主体に変え、授業内の形成的評価の発展に寄与すると考えられるためである。このような研究背景をもとに、以下の研究課題を掲げ、生徒評価の効果に関する分析を行った。

- (1) 自己評価と他己評価は、以下の観点から比較するとどのような類似点・相違点があるか。
 - a) 先生評価に比した生徒評価の信頼性
 - b) 英作文能力に与える影響
 - c) 英作文学習不安と学習者の自立性
- (2) 生徒評価は教室内でどのように形成的評価として機能するか。

第2章 先行研究

第2章では、主に教室内評価と形成的評価の観点から先行研究を論じた。また、生徒評価に関する先行研究については、メタ分析的手法をとって自己評価と他己評価のそれぞれについて先行研究の調査主題や研究方法、教員評価に対する自己評価と他己評価の信頼性、その効果や限界などをまとめた。さらに、生徒評価それぞれのタイプが与える学習者情意、つまり英作文学習不安や学習者の自立性に関する調査手法（特に研究手法の有用性や妥当性）や研究の歴史についても先行研究をもとに比較分析した。その結果、18歳以下の学習者に対する研究が極めて乏しいこと、自己評価と他己評価を直接比較対照しその特性を明らかにした研究も少なく、さらに混合研究法は有望である反面あまり採用されていないことが分かった。また、生徒評価と形成的評価との関係を論じた研究も非常に少なかった。よって、本章の最後には、この先行研究の分析結果をもとに、本研究の各研究課題に対する仮説を掲げた。

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、15～18歳の293名の日本の高校生と4名の英語教員（日本人2名、英語母語話者2名）が調査に参加した。また、質的・量的研究法それぞれの弱点を補完し、それぞれの分析結果を統合するために収斂的混合研究法（convergent mixed-methods design）を採用した。本研究の調査の前にはニーズ分析を実施し、調査に運用する評価基準の作成を行った。また、先行研究の分析結果をもとに、英作文学習不安を調査するために the Second Language Writing Anxiety (SLWAI; Cheng, 2004) を、また学習者の学習自立度を調査するために Chang (2007) の質問用紙を採用した。これらの調査用紙は日本語に翻訳し、試行調査を経て修正を加えた。また、本研究は①参加生徒の評価練習、事前テスト、及び調査用紙回答、②5回の英作文とそれに伴う生徒評価・教師による観察、③事後テスト、調査用紙回答（自由回答を含む）、半構造的面接の実施の3段階で実施することとした。

量的研究（研究課題1）においては、多相ラッシュモデルを用いて自己評価と他己評価の信頼性を分析し、同じく多相ラッシュモデルにより英作文能力に対するそれぞれの生徒評価タイプの影響を分析した。また、各生徒評価タイプの英作文学習不安と学習者の自立への影響の分析には多変量共分散分析を用いた。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、事後調査用紙にある自由回答については、回答者が使用する表現に着目する用語抽出法を使って分析をした。また、12名の生徒（自己評価と他己評価の各グループから英作文の評価結果をもとに6名ずつ抽出）と2名の教員（日本人1名、英語母語話者1名）に対して実施した半構造的面接データは、グラウンデッド・セオリーを用いて分析した。分析の最終段階で、生徒評価と形成的評価の関係を考察するために探索的に質的・量的研究分析結果を統合した。

第4章 量的研究分析結果

第4章では、自己評価と他己評価を比較し、研究課題1に関連する三つの観点から量的研究分析結果をまとめた。第一に、教員評価に比した生徒評価の信頼性に関する多相ラッシュモデル分析結果は、英作文の総合得点における自己評価は教員評価と同程度の厳格さであることを示した。一方、他己評価は評価の一貫性と厳格さのいずれにおいても教員評価には及ばなかった。また分析的評価尺度における得点については、自己評価、他己評価のいずれも教員評価と同等の一貫性も厳格さは示さなかった。第二に、生徒評価の各タイプが英作文へ与える影響を同じ

く多相ラッシュモデル分析を用いて検証した。その結果、両評価グループにおいて文法の正確さの得点向上が認められ、また自己評価グループについては、加えて英作文の総合得点向上も確認された。第三に、生徒評価各タイプの英作文学習不安と学習者自立への影響の検証には多変量共分散分析を用いた結果、自己評価・他己評価グループのいずれについても、学習者自立度の向上に伸びが認められた一方、英作文学習不安については、自己評価・他己評価のグループいずれも影響は確認されなかった。このように、量的研究分析結果は、自己評価と他己評価が評価の厳格さと英作文能力の伸びにおいて異なるが、文法の正確さの向上や学習者自立度に肯定的影響を与えるという点では類似していることを示した。

第5章 質的研究分析結果

第5章では、質的研究分析結果の観点から、生徒評価が教室内の形成的評価においてどのように機能するかをまとめた。事後調査用紙の自由回答を用語抽出法によって分析した結果、両生徒評価タイプ共に、用語使用頻度の上位3位が、全用語登場頻度のほぼ半分を占めることを示した。自己評価グループでは、語彙や文法知識不足に関する自己認識や自己評価に対する自信欠如を指摘する用語が上位を占めた。一方、他己評価グループでは、他己評価によって同級生の英作文から学ぶことができる意義や同級生の英作文を読む楽しさを指摘する用語の使用頻度が高かった。また、順位は異なるが、英語力や生徒評価に関する自信のなさに関する用語が両グループ共に認められた。次に、グラウンデッド・セオリーを用いた半構造的面接の分析結果については、両グループ共に、その構成要素は違うものの、英作文力の向上が中心的なカテゴリと特定された。自己評価グループの生徒は、より強く評価基準を意識し、その気づきが内省力や英作文の全体的な向上に役立っていた。一方、他己評価グループの生徒においては、同級生との交流の楽しさが、他者からの学びやメタ認知的能力の向上につながったことがわかった。

第6章 考察

第6章では、量的研究と質的研究結果を統合し、二つの研究課題を考察し、生徒評価の有効性を考察した。まず、第一の研究課題については、自己評価と他己評価は共に文法の正確さと学習者自立の向上においては有効であるが、その影響の質が異なると考察された。これは、量的研究分析結果が示す通り、厳格さにおいて教員評価と同等の自己評価は英作文の評価基準に対する認識を強め、自己規制や内省力を高めたためであると考えられる。自己評価によって醸成されたその強い自己認識は、英語力に対する自信喪失にもつながるが、文法の正確さを向上させるために肯定的な影響を与えたと解釈できる。一方、他己評価においては、生徒同士で英作文を読み合い評価をすることによって、同級生から英語を学ぶことが文法力向上の一助となったと思われる。さらには、励ましや誉め言葉を交換することが、教室内の雰囲気をややかにし、生徒間の交流を高め、そうした交流が英作文へ取り組む態度に肯定的な影響を与えたと考えられる。英作文学習不安に対する生徒評価の影響は、量的研究結果では認められなかったが、質的研究では両グループ共に、不安を課題完成に対する自己達成感や友達との交流による学びへの満足などの肯定的な情意に変えていることが示された。このように自己評価と他己評価が英作文を書く上で必要となる英語運用能力や情緒面に与える影響には類似点や相違点があるが、両生徒評価タイプ共に、英作文を書く能力の向上につながる可能性を示した点では共通してい

た。第二の研究課題は、生徒評価が教室内でどのように形成的評価として機能をするかについてであった。本研究では教師が観察者としてのみ教室内に介在したことが、生徒評価における教員の迅速かつ適切な指導の必要性を明らかにした。これは形成的評価を完成するためには、生徒評価だけではその信頼性において不十分であり、たとえ英文法や学習者自立度が向上したとしても、学習者の生徒評価に対する自信や信頼性には不安が残るためである。形成的評価の成功 (Leahy & Wiliam, 2012) のためには、教師がそれぞれの生徒評価タイプの特性をよく理解したうえで教室内に導入することが求められる。

第7章 結論

本研究は、日本の高校生を対象に自己評価と他己評価という二つの生徒評価の有効性を探究することを目的とした。二つの評価タイプを比較することによって、両者の類似点や相違点を明らかにし、それぞれの特質も具体化した。さらに、収斂的混合研究法の採用により量的研究と質的研究分析結果を統合することで、これまで漠然としていたそれぞれの生徒評価タイプと形成的評価の関係性を明らかにした。それによって、生徒評価は単なる教員評価の代替ではなく、全く異なる視点や特質を持つ評価法であるとの結論に至った。また、本研究結果は、形成的評価の成功に導くためには、生徒評価に教員評価が適切に介入することが不可欠であることにも光を当てた。これは教室内評価を構成するのは教師と学習者であり、それゆえに、教師と学習者が共に教え、学び、評価することの責任を共有することが重要であり、それが教室内で形成的評価をうまく成立させる鍵となることを示すものに他ならない。

参考文献

- Chang, L. (2007). The influences of group processes on learners' autonomous beliefs and behaviors. *System*, 35(3), 322-337. doi: 10.1016/j.system.2007.03.001
- Cheng, Y.-S. (2004). A measure of second writing anxiety: scale development and preliminary validation. *Journal of Second Language Writing*, 13(4), 313-335. doi: 10.1016/j.jslw.2004.07.001
- Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Eds.), *Assessment and learning* (pp. 49-72). London: Sage Publications.

4 総評

本研究は論文としての完成度が高く、その強みは主に以下の2点に集約される。その一点目は、本研究が言語テストや第二言語習得分野にもたらす理論的貢献である。教育現場における言語テストについては、学習者を「消極的な評価結果の受け手」ではなく「積極的な評価の実践者」とし、学習者が評価の担い手となる学習者評価 (student assessment) の重要性が近年指摘されている。その一方で、学習者評価と教員評価との間、また自己評価と他己評価など複数の学習者評価手法間での類似点・相違点が具体的にどこにあるのか、また学習者評価が言語教育の場でどのように形成的評価のツールとして機能するのかについては十分に解明されていない。この現状に答えるべく、大井氏は関連分野における先行研究を系統立ててまとめ、日本人高校生を対象とした混合研究法の数が極めて限られることや、過去の調査研究は採点結果の一貫性など一部のテーマを扱うものが多かった点を指摘している。それを踏まえて混合研究法により多角的に両者を比較する本研究を計画した。その結果①自己評価と他己評価のいずれも評価結

果の厳格さや一貫性の面で教員評価とは同等とはいえないこと、②両者とも英作文力（特に文法的正確さ）の伸びと学習者の自立や書き手としての成長を促す共通点がある一方、自己評価は内省を促し、他己評価は読み手との対話により英作文に対する肯定的態度を養うなど、両者には質的な違いがあること、また③両者を形成的評価手法と捉えた場合、教育現場での役割に違いがあり、その効果的な運用には教員の適度な介入が肝要であることを明らかにしている。このような両評価法に関する具体的な知見は、教室における言語評価研究に貴重な示唆を与えるだけでなく、読み手への気づきを促すことで書き手の成長をサポートする協働ライティング（collaborative writing）手法と他己評価を捉える第二言語ライティング研究の視点を支持するものでもあるといえる。また、高等学校新学習指導要領における望ましい英語ライティング評価の在り方やライティング指導の活性化にも大きな示唆が得られる内容で、教室における教員評価や生徒評価それぞれの意義や役割について改めて考えさせられる論考である。

本論文の第二の強みは、その洗練された研究手法にある。収斂的混合研究法（convergent mixed-methods design）を構成する質的・量的手法それぞれが、研究仮説を検証するうえで効果的であった。量的研究部分は首都圏の公立高校の既存の8クラスの生徒と4名の教員の協力を得て、各クラスを自己評価群と他己評価群に分けてグループ比較する手法を取っている。「コミュニケーション英語I, II」の授業2回において英作文評価トレーニングを行って生徒に各評価法をまず理解させ、続く10日間のライティング指導期間に作文と評価の活動を5回の授業で繰り返している。評価結果の厳格さと一貫性の分析のみでなく英作文・評価活動前後での得点の伸びの検証を目的とする反復測定分析にも多相ラッシュモデル（Many-facet Rasch measurement）を適切に使い、また生徒の情意（英作文に対する不安や学習者としての自立）に関するアンケート回答の多変量共分散分析と合わせ、研究課題への答えをうまく引き出している。また、質的研究部分ではアンケート自由回答の用語抽出分析、成績上位・中位・下位群から抽出した生徒12名と観察を担当した教員2名の半構造化インタビュー結果のグラウンデッド・セオリー分析を組み合わせ、自己評価・他己評価の英作文指導における役割を明らかにした。さらには Leahy and Wiliam (2012) の形成的評価の特徴に関する枠組みに照らして本研究結果に更なる考察を加え、質的・量的研究結果を十分に統合して吟味している。これにより得られた結論は自己評価と他己評価の多面性を浮き彫りにするもので、混合研究手法の特性を十分生かした成果であるといえよう。

一方で、審査員からは研究手法や解釈について複数の指摘もあった。一つは本研究における活動と通常の授業との関係性である。本研究はグループ比較を可能にするため、通常の授業時より制約の多い中で実施されており、普段の授業実践から切り離された活動として完結している印象を受けることは否めない。よって通常クラスでの授業実践（教員の役割や学習目標・指導内容と教員フィードバックの方法など）に結び付けて丁寧に考察することで、普段の英語学習指導への示唆をより明確にすることが必要であろう。特に英作文タスクについては、各ユニットのまとめとして要約を書かせるなど、授業内容との繋がりや実施のタイミングをより明確にする内容・形式を採用することも一案と思われる。また、他己評価グループが使用した評価シートの内容・形式が、自己評価グループが使用した採点尺度との並行性を保つためにフィードバックの具体化を十分促すものではなかったことも今後改善を要する点であろう。加えて生徒の自立度にはこの短期間の活動を通して変化が見られたものの、このような英作文・評価

活動が生徒の英作文に対する不安に与える影響を検証するためには、より長期間の英作文・評価活動が可能な研究デザインが必要とも考えられる。しかし、これらは本論文の価値や意義を損なうものではなく、今後様々な方向に研究を発展させていくうえでの課題と位置づけられた。

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。