

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月 29日現在

機関番号：12101

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22520550

研究課題名（和文）中学校英語授業での代替評価の効果

研究課題名（英文）Effects of Formative Assessment in Junior High School English Lessons

研究代表者

齋藤 英敏（Hidetoshi Saito）

茨城大学・教育学部・准教授

研究者番号：20318695

研究成果の概要（和文）：本研究の目的は英語授業で6か月間にわたり代替（形成的）評価を授業活動で使用したクラス（実験群）と使用しなかったクラス（統制群）の学習での成果を比較し、評価活動の影響を調査するものである。リスニングでは統計的には授業内評価活動の影響は見られなかった。事後テストでは実験・統制群の違いに関わらず、統計的に有意な伸びがみられた。自己評価でも、実験・統制群の違いは事後ではみられなかった。スピーキングに関しては現在分析中である。

研究成果の概要（英文）：The study examined the effects of using formative (peer and self-reflection) assessment for six months in junior high school EFL lessons. All students were engaged in pair speaking activities which another pair observed. In the experimental groups, students filled in a peer assessment and self-reflective comment sheet after the activity, while the control groups repeated practice instead of filling the form. There was no statistically significant difference in the post-listening test scores between the experimental and control groups, although both groups show statistically significant improvement from the pre-test scores. There was no statistically significant difference in post-self-assessment measures between the experimental and control groups. Speaking test data is currently being analyzed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	500,000	150,000	650,000
2011年度	600,000	180,000	780,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	1,800,000	540,000	2,340,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：代替評価、ピア（相互）評価、形成的評価、英語教育

## 1. 研究開始当初の背景

これまで様々な実践が報告されているが代替評価の学習への効果を調査したものは極めて少ない。唯一、外国語教育分野ではRoss (2005)がある。この論文は、ある大学の

英語プログラムで代替評価（論文では形成的評価と呼ばれている）を主に用いた4ヵ年と通常の総括的な評価を用いた4ヵ年に授業履修した生徒の英語の成績、習熟度テスト結果を比較したものである。この英語プログラム

は二年にわたるもので、それぞれの学生は4回分の成績と3回分の習熟度テスト結果を持ちそれが分析された。結果としては代替評価を用いた学生たちのほうが、リスニングにおいて顕著な習熟度の伸びがみられたが、リーディングには違いがなかった。その理由は明らかではないが、代替評価の効果を示す最初の大規模( $n = 2215$ )な研究である。しかしいくつかの問題点もある。例えば二つのグループははっきりと代替評価、総括的評価グループとわけがたい部分がある。また代替評価グループの一年目はほとんど代替評価は行われていなく、二年目も徐々にその代替評価使用率が上がっていくという形である。また総括的評価グループでもわずかな割合ではあるが代替評価が使用されている。また長期にわたっているので教師が入れ変わり、教科書も変更されるなど多用な要素が説明要因としても考えられる。

本研究では相互評価、自己の振り返り活動を行ったグループと行わなかったグループの比較で、期間はおよそ6か月で短いグループ間の違いは明確である。

## 2. 研究の目的

本研究の目的は中学校英語授業で一年間にわたり代替(形成的)評価を授業活動で使用したクラス(実験群)と使用しなかったクラス(統制群)の学習での成果を比較し、評価活動の影響を調査するものである。つまり代替評価を使用した実験群は、リスニング、スピーキング、自己評価の点数において、代替評価を用いない統制群より顕著な学習の伸びがみられるかどうかを調査した。評価活動を行う学習活動は四人組でのスピーキングで二名が話題に関して話をし、残り二名が観察を行うものである。それは主にスピーキングの活動となるが、リスニングも含まれているのでスピーキング、リスニングテストでその影響を見ることとした。リスニングテストを含むもう一つの理由として、先行研究の

Ross(2005)が影響が見られたスキル領域として報告していることも挙げられる。また、もう一つの目的として、相互評価のコメント内容は回数を重ねるにつれてどのような変化が起きたかを調査することとした。

## 3. 研究の方法

中学校二年生の四クラス(データが不完全なものを除いたあとの147名)がランダムに二クラスは実験群、二クラスは統制群にわけられる。両群ともに週一回、教科書から少し離れたスピーキング活動の授業を行った。四

クラスのうちランダムに二クラスを統制群、二クラスを実験群(うち一つは教師の担任のクラス)とする。6月にリスニング、スピーキングテスト、自己評価の事前テストを両グループで行い2月に事後テストを行った。自己評価は自らの学習の変化を自己申告をする。また、スピーキング、リスニングテストともに別の中学生グループで妥当性などの検討を行い、ラッシュ分析を用いて等価されている。これ以外に校外模試4回、校内試験2回のデータが補足データとして入手された。

実験群では授業内で半年間(7月から2月)通して相互評価、自己の振り返りの代替評価活動を用いた。またピア評価はスピーキングの活動を四人グループで行い、二名がペア活動を行っている際に残りの二名が評価表を用いて観察評価を行い、活動者にフィードバックを返すことをおこなう。これは半年間を通して十二回行った。相互評価・自己の振り返りのワークシートはファイル化して、2月に振り返りの活動を行った。その際、実験群はアンケートで相互評価の有効性について聞かれた。統制群では上記のいずれの代替評価活動も用いない。その代わりにその時間分の4人組での会話活動時間が増えることとなる。相互評価は同様に四人グループで行うが二名が行っている際、残りの二名は観察するだけでフィードバックは返さないこととする。最後のアンケートで統制群はこの活動の有効性についてたずねられた。リスニングテスト、自己評価、は分散分析によって分析された。スピーキングに関しては現在分析中である。

生徒の相互評価のコメントはすべて入力され、SPSS Text Analytics for Surveysで分析されたあと、二名の評価者によりカテゴリ化(コーディング)が行われた。データはその後コレスポネンス分析を行った。

## 4. 研究成果

分散分析の結果、リスニングでは統計的には授業内評価活動の影響は見られなかった( $F(1,145) = 2.53, p > .10$ )。事後テストでは実験・統制群の違いに関わらず、統計的に有意な伸びがみられた( $F(1,145) = 8.96, p < .001$ )。しかしながら、上位( $n=29$ )中位( $n=80$ )下位( $n=38$ )と三群に補足テストデータをもとに分けると、上位、下位群は実験(下記 Figure 1)・統制(下記 Figure 2)群両方で伸びがみられたが、中位に関しては統制群では変化がみられない傾向が見られた。自己評価でも、実験・統制群の違いは事後ではみられなかった

。また統計的に有意な伸びもみられなかったが、上位中位下位の三群にわけると、実験群では上位・中位に上昇傾向が見られたが、統制群にはそれがなかった。しかし、いずれも実験群への評価活動の統計的効果はこの二つの指標ではみられなかった。

Figure 1. Results of Listening Test (Treatment)

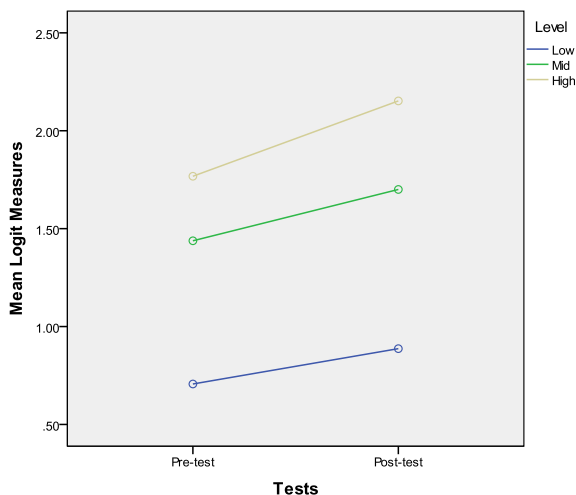
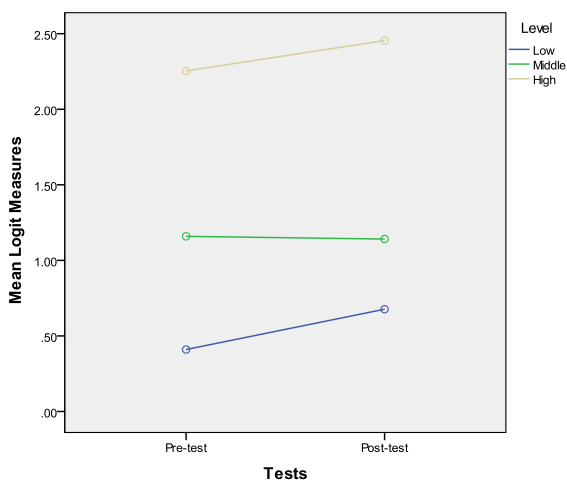


Figure 2. Results of Listening Test (Control Group)

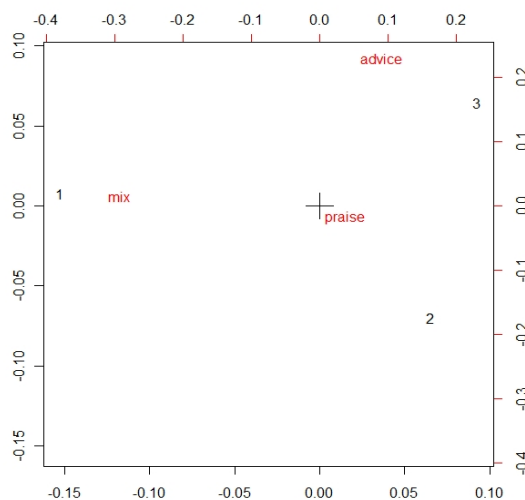


生徒の相互評価コメントの変化を見るためにコメントのカテゴリ化は「機能（誉め、アドバイス、誉め・アドバイスの混合）」と「内容（データに基づいた記述的な23カテゴリ）」の二通りのコーディングを行い、前期（1-4週）・中期（5-7週）・後期（8-11週）に分けて頻度を分析した。結果として、コメントの機能では期間にかかわらず、「誉め」が圧倒的に多い。また後期に「アドバイス」が少し上昇し「誉め・アドバイスの混合」は前期に比較的多く見られた（下記Figure 3を参照）。

コメントの内容の面では「話の内容」「流

暢さ」「アイコンタクト」などが期間に関わらず多い。これらの頻度の高いカテゴリは、生徒たちが使用した相互評価表の中にある項目と同一であった。後期になるとこれらの、評価項目にあるカテゴリの頻度が減少し、「質問をする・答える」「話をつなぐ」などの評価表項目にないカテゴリが上昇する傾向が見られた。また、学内外の到達度テストスコアとはコメントの「内容」は相関がないことがわかった( $r = -.03$ )。つまり、生徒は到達度に関わらず、多様なコメントを受けているということになる。

Figure 3. Results of Correspondance Analysis



注：1=前期, 2=中期, 3=後期

スピーキングテストは現在分析中であるが、リスニング、自己評価からは明確な評価活動の影響を本研究ではしめすことはできなかった。理由として考えられることは以下の通りである。(1)相互評価、振り返りなどの評価活動はリスニングテストや自己評価には明確な影響を与えるほどのインパクトがない。(2)本実験で使用したリスニングテストや自己評価は評価活動の影響に対して敏感な測定ではない。(3)相互評価活動でのコメント自体がリスニングに対してのものではないので、リスニングには影響がないのは当然である。(4)統制群は評価の代わりに練習を1回程度多く行っていた。この回数が評価活動の影響の埋め合わせをした可能性がある。

また、上記の Figure 1, 2 から人数の最も多い中位群で実験群では上昇傾向が見られ、統制群でその傾向が見られないことは、評価活動は中位群に対して何らかの働きがある可能性がある。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- ① Saito, H., Kino, I., & Saito, T. Development of Listening Prochievement Tests for Third- Year Japanese Junior High School Students Studying English as a Foreign Language (Part II), 茨城大学教育学部紀要(教育科学), no.61 p.207-215, 2012, 査読無  
<http://hdl.handle.net/10109/3191>
- ② Saito, H., Saito, T., & Kino, I. Development of Listening Prochievement Tests for Third- Year Japanese Junior High School Students Studying English as a Foreign Language (Part D). 茨城大学教育学部紀要(教育科学), no.60 p.111-121, 2011, 査読無  
<http://hdl.handle.net/10109/2556>

[学会発表] (計 2 件)

- ① Saito, H. Qualitative Change in Junior High School Students' Peer Assessments on Spesking Performances. The 35<sup>th</sup> Language Testing Research Colloquium, 2013. 7. 4. Seoul (韓国)
- ② Saito, H. What Makes Peer Comments More "Formative" in Speaking Activities? 第 16 回日本語テスト学会, 2012. 10. 26. 専修大学.

## 6. 研究組織

### (1) 研究代表者

齋藤 英敏 (SAITO HIDETOSHI)

茨城大学・教育学部・准教授

研究者番号：20318695