

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 18 日現在

機関番号：34416

研究種目：挑戦的萌芽研究

研究期間：2014～2015

課題番号：26580121

研究課題名(和文)日本人大学生を対象とした技能統合型英文要約タスクの評価尺度の開発

研究課題名(英文)Developing a Rubric for Integrated Summary Writing Tasks in English

研究代表者

山西 博之(YAMANISHI, Hiroyuki)

関西大学・外国語学部・准教授

研究者番号：30452684

交付決定額(研究期間全体):(直接経費) 1,900,000円

研究成果の概要(和文):本研究は、日本の大学英語教員のニーズを満たす、技能統合型英文要約タスクの評価尺度の開発を目的とし、その第一段階として、海外で開発された当該タスク用の総合的評価尺度と分析的評価尺度の比較を行った。その結果、両尺度は総合点のみを扱う限り、遜色ない評価結果が得られるが、日本の文脈に合う評価尺度にするためには記述子等に改善が必要であることがわかった。また、英文要約タスクの指導において、代表者らが開発した評価尺度を書き直しの際の到達目標として明示した結果、評価尺度を参照した学習者群はモデル要約を参照した学習者群よりも要約の自己採点結果が有意に高く、評価尺度は指導のツールとして効果があることが示唆された。

研究成果の概要(英文):This study aims to develop a rubric for integrated summary writing tasks in English, which meets needs from English language teachers in Japanese's universities. As a first step, we compared a holistic rubric and an analytic rubric developed for the tasks outside of Japan. Results showed both rubrics were valid and reliable for evaluation of task performance only when the overall scores of each rubric were taken into consideration. Therefore, descriptors of the rubrics need refinements so that the rubrics would be more appropriate to the Japanese contexts. In teaching English summary writing tasks, the rubric developed by the researchers was shown to learners explicitly as a task goal of revision in order to examine the effectiveness of the rubric. Consequently, those who referred to the rubric showed significantly higher self-evaluation scores of their summaries than those who referred to model summaries. Hence, the rubric seems to be an effective tool to teach summary writing tasks.

研究分野：英語教育学

キーワード：要約研究 技能統合型タスク

1. 研究開始当初の背景

本研究では、英語で書かれた学術的な文章を読み、あるいは読んで聞き、その要点を英語でまとめる「技能統合型 (Skill-integrated) 英文要約 (Summary writing) タスク」を取り扱う。英文要約に関する研究は主に大学生以上を対象として行われているが、その多くは海外においてのものであり、日本の大学生を対象にした体系的な英文要約の研究はほとんど行われてきていない。しかしながら、昨今の我が国では、政府による「グローバル人材育成戦略」や教育再生実行会議による提言などで、技能統合型の課題を課す TOEFL iBT の大学教育への導入 (入試や卒業要件として) が議論されたり、平成 25 年度入学生からの新高等学校学習指導要領で「4 技能を総合的に育成する指導」が求められたりするなど、要約に限らず技能統合型の指導や評価への関心やニーズが高まっているといえる。

このような関心やニーズの高まりにもかかわらず、日本人を対象とした技能統合型タスクの指導や評価の実態はほとんど明らかになっていない。そのような状態では、例えば、大学の入試や卒業要件に TOEFL iBT を導入すべきという動きに関しても、指導や評価における実態の把握や、それに基づいた体系的な指針もないまま導入されることになり、教育現場では混乱が起きることが予想される。そこで代表者らは、技能統合型タスクのうち、日本人大学生を対象とした技能統合型英文要約タスクに焦点を絞り、その結果を評価可能な尺度を開発するという着想を得た。これは、評価のみの使用に留まらず、指導のための到達目標としても利用可能なもので、教育現場における適切な指導と評価のためのツールという位置づけになる。

2. 研究の目的

本研究は「日本人大学生を対象とした技能統合型英文要約タスクの評価尺度の開発」という課題を取り扱い、その目的は、以下の 2 点にまとめられる。

目的 : 日本人大学生が取り組む技能統合型の英文要約タスク、より具体的には英文を「読んで、要約を書く」「聞いて、要約を書く」、そして TOEFL iBT の様な「読んで、聞いて、要約を書く」といったタスクに対応することが可能な評価尺度を開発する (研究)

目的 : 評価尺度による評価と連動した、技能統合型英文要約タスクの指導を実践し、その効果を検証する。具体的には、評価尺度を到達目標として明示した指導の効果を検証する (研究)

3. 研究の方法

研究

目的 に挙げたうち、TOEFL iBT でも課される「読んで、聞いて、要約を書く」という

3 技能統合型の英文要約を評価するための尺度開発に従事した。そのような評価尺度開発のための第 1 段階として、海外で開発された 3 技能統合型ライティング課題用の総合的評価尺度と分析的評価尺度の比較を以下の通り行った。これにより、既存の評価尺度における困難点と改善点を検討した。

(1) 協力者

日本の大学に在籍する日本人大学 1、2 年生 36 名 (男 13 名、女 23 名) であり、英語インテンシブ・コース (履修者の多くは留学を目指す) で週 4 回英語の授業を履修する学生であった。協力者の専攻は法律学または政治学であった。

(2) 手順とマテリアル

TOEFL iBT の技能統合型ライティングを実施するにあたり、タスクに修正を加え、ライティングの制限時間を 20 分から 30 分に延長してタスクの難易度を下げた。使用した技能統合型ライティングのマテリアルは Wadden・Hilke・早川 (2014) より抜粋した。

協力者は、まずリーディングとして 225 語のパスセージを 3 分間で読み、リスニングではレクチャーを 2 回聞いた。その後、ライティングでは提示されたキーワードをもとに、パスセージとレクチャーの要点がどのように関連しているかを書いた。タスク終了後、協力者は当該タスクの難易度や所感を質問紙に回答した。この質問紙 (17 項目、そのうちの 1 項目は自由記述) は Cho, Rijmen & Novák (2013) で使用された質問項目に改良を加え和訳して使用された。

(3) 採点方法

採点者は日本の大学英語教員 7 名 (英語母語話者 3 名、日本人 4 名) であった。英語母語話者教員は 9~24 年の英語教授経験があり、日本人教員は 4~13 年の英語教授経験があった。日本人教員 1 名のみ Educational Testing Service (ETS) の技能統合型ライティングの尺度の使用経験があった。採点者は総合的評価尺度 (ETS, 2008) と分析的評価尺度 (Yang & Plakans, 2012) を使って 36 篇の文章をそれぞれ採点した。各尺度は 6 段階評価 (0~5 点) で、分析的評価尺度には、「内容 (content)」「構成 (organization)」「言語使用 (language use)」「言い換え (verbatim source use)」の 4 観点が含まれた。採点の慣れによる影響を防ぐため、採点する文章の順序は採点者ごとに並べ替えられた。採点者は、採点終了後に、各評価尺度の使用における難易度や改善点を省察して質問紙 (14 項目、そのうちの 9 項目は自由記述) に記述した。

(4) データ分析

分析対象として 2 種類のデータを扱った。まず、総合的評価尺度と分析的評価尺度を基

にそれぞれ採点された5点満点の評定値である。採点者間信頼性を確認し、総合的評価尺度と分析的評価尺度における評定値の相関を分析した(自由記述データについては、別機会に取り扱うものとする)。

研究

目的 においては、代表者らが開発済みの「読んで、要約を書く」という2技能統合型の英文要約の評価尺度を用いた指導実践を行った。

(1) 協力者

日本の大学に在籍する日本人大学3年生82名(39名のクラスAと43名のクラスB)であり、学際的な領域の学部において必修科目で週2回英語を学ぶ学生であった。クラスAとクラスBは、同一学部内の学生がランダムに振り分けられたものであったため、両クラスには、ほぼ同程度の英語熟達度を持つ学生が所属していた。

(2) 手順とマテリアル

協力者は194語から成る文章(Oshima & Hogue, 2007)を読んで、その内容を50語~60語の英語でまとめた。要約を作成後、代表者らが開発した英文要約評価尺度を用いて、以下の2条件の比較を行った。この評価尺度は「内容」「言い換え(量)」「言い換え(質)」「言語使用」「全体的な要約の質」という5観点が含まれた。

モデル要約参照群(クラスA): 協力者は、要約(1回目)を作成した後、モデル要約(Oshima & Hogue, 2007)を参照し、自らが作成した要約と比較することで改善点を把握し、要約の書き直しを行った。その際、「書き直しで気が付いたこと」について、自由記述形式で振り返った。その後、協力者は配布された評価尺度を用いて、書き直した要約(2回目)を4段階(1~4点)で自己採点した。自己採点時には、その評点を出した採点理由を観点ごとに自由記述形式で回答した。

評価尺度参照群(クラスB): 協力者は、要約(1回目)を作成した後、配布された評価尺度を用いて要約を4段階(1~4点)で自己採点し、採点理由を観点ごとに自由記述で回答した。その後、評価尺度の各項目の記述子の内容が到達目標であることを明示的に指導した。指導後、協力者は評価尺度を参照しながら要約の書き直しを行い、書き直した要約(2回目)を再び4段階(1~4点)で自己採点した。その際には、採点理由および「書き直しで気が付いたこと」を観点ごとに自由記述形式で回答した。

(3) データ分析

ここでは、分析対象として、モデル要約参照群の2回目の要約自己採点結果、評価尺度参照群の1回目と2回目の自己採点結果を取り扱った(自由記述データについては、別機

会に取り扱うものとする)。

具体的には、評価尺度参照群の1回目の要約自己採点結果をベースラインのデータとして、モデル要約参照群と評価尺度参照群の2回目(書き直し分)の自己採点結果をそれぞれの指導の効果検証用のデータとした。

4. 研究成果

研究

総合的評価尺度における英語母語話者3名の採点者間信頼性は $\alpha = .81$ 、日本人4名の採点者間信頼性は $\alpha = .82$ であり、どちらも高い信頼性であった。同様に、分析的評価尺度における英語母語話者群の採点者間信頼性は $\alpha = .79$ 、日本人群は $\alpha = .83$ で、どちらも高い信頼性係数が確認された。

次に、総合的評価尺度と分析的評価尺度における評定値の相関は、英語母語話者群(図1)、日本人群(図2)の通りとなった。

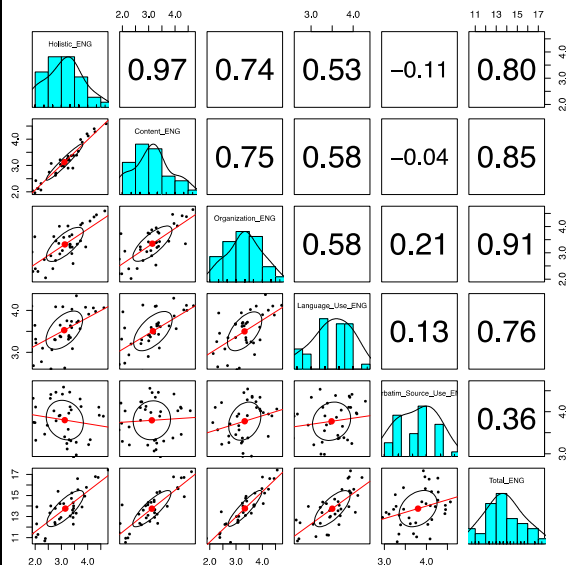


図1. 英語母語話者群の評価の相関

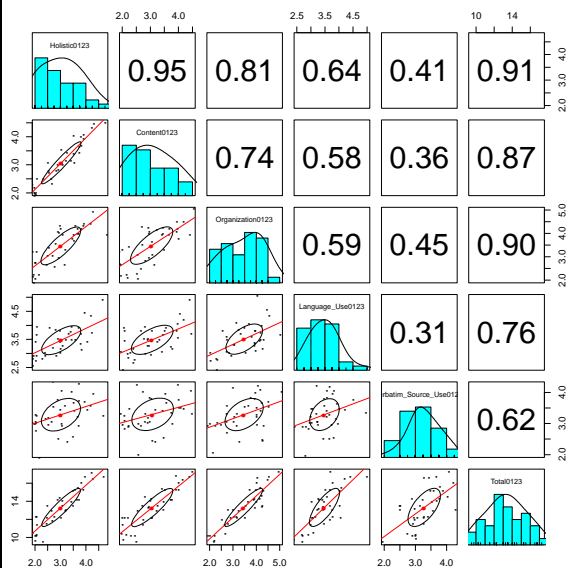


図2. 日本人群の評価の相関

図1、2から、英語母語話者と日本人の双方

とも、総合的評価尺度と分析的評価尺度（総合点）の相関が高いことが確認された（それぞれ、 $r = .80$ 、 $r = .91$ ）。また、分析的評価尺度の下位項目である「内容 (content)」「構成 (organization)」「言語使用 (language use)」「言い換え (verbatim source use)」の4観点は、概ね関連し合っ
て一つの尺度を構成していると解釈することができた。しかしながら、特に英語母語話者の評価における「言い換え (verbatim source use)」については、他の観点との相関がほとんど見られないことが分かった。

この結果から、今回用いた総合的評価尺度 (ETS, 2008) と分析的評価尺度 (Yang & Plakans, 2012)は、評価結果のうち総合点のみを問題にする限り、遜色ない評価結果が得られる評価尺度であったといえる。

しかし、実際の評価は、採点者がどれほど評価の困難度を感じるのか、あるいは評価にどれほどの時間がかかるのか、といった事情によって評価尺度の使い分けがなされることとなる。また、評価の目的や、評価結果からどの程度の診断的なフィードバック情報を得ることを目指すのかによって、総合的評価尺度（単一スコアという診断情報のみ）と分析的評価尺度（観点ごとの診断情報を得られる）を使い分けることになる。そして同時に、日本の大学英語教育の文脈に合った、より適切な評価尺度になるよう、評価尺度の記述子等を改善していく必要もある。そのような適切な使い分けや評価尺度の改善を可能にするためには、評価結果に対する量的分析と、採点者および学習者による当該タスクと評価尺度の省察に対する質的分析の両者に基づいて、それぞれの評価尺度の性質を捉えるための検討が不可欠である。

研究

要約作成（書き直し）におけるモデル要約参照と評価尺度参照の効果を検証するために、表1に示した自己採点結果の比較を行った。

表1. 自己採点結果（5観点の合計20点）

評価尺度参照群 (1回目)	モデル要約参照群 (2回目)	評価尺度参照群 (2回目)
11.53 (2.55)	13.28 (2.88)	14.45 (2.17)

注. 評価尺度参照群 ($n = 43$)、モデル要約参照群 ($n = 39$)
数値は、平均値（標準偏差）

評価尺度参照群（1回目）とモデル要約参照群（2回目）の平均値の差に対しては、対応のないt検定を行った。その結果は、 $t(80) = -2.91$ 、 $p = .003$ 、 $d = -0.64$ であり、モデル要約参照群（2回目）の平均値がベースラインである評価尺度参照群（1回目）よりも高いことが確認された。

次に、評価尺度参照群（1回目）と評価尺

度参照群（2回目）の平均値の差に対しては、同じ学習者の繰り返しデータであるため、対応のあるt検定を行った。その結果は、 $t(42) = -12.00$ 、 $p < .001$ 、 $d = 1.21$ であり、評価尺度参照群（2回目）の平均値がベースラインの平均値を上回ったことが確認された。

最後に、モデル要約参照群（2回目）と評価尺度参照群（2回目）の平均値の差に対して、対応のないt検定を行った。その結果は、 $t(80) = -2.09$ 、 $p = .040$ 、 $d = -0.46$ であり、評価尺度参照群（2回目）の平均値がモデル要約参照群（2回目）の平均値を上回ったことが確認された。

これらの分析から、要約書き直し時の到達目標として、モデル要約の参照と評価尺度の参照という異なる指導方法は共に有効な方法であったこと、そして両指導方法の間では、評価尺度参照の方がより高い効果があったことが分かった。

しかしながら、この結果は協力者である学習者の自己採点結果のみに基づいているため、今後は実際に学習者が書いた要約を第三者が評価した結果を用いた検証も必要である（その際には、評価尺度参照（1回目、2回目）・モデル要約参照（1回目、2回目）の4種の要約を評価することになる）。同時に、学習者の自由記述回答を質的に分析することで、より具体的にこれらの指導がどのような影響や効果をもたらしたのかを学習者の視点から把握することも望まれる。

5. 主な発表論文等

〔雑誌論文〕(計2件)

山西 博之、大学でのサマリー・ライティングの指導と評価、英語教育(大修館書店) 査読無、12巻(12月号)、2015、22-23

Yuko HIJIKATA-SOMEYA, Masumi ONO, & Hiroyuki YAMANISHI, Evaluation by native and non-native English teacher-raters of Japanese students' summaries, English Language Teaching, 査読有、8巻(7号)、2015、1-12

DOI: <http://dx.doi.org/10.5539/elt.v8n7p1>

6. 研究組織

(1)研究代表者

山西 博之 (YAMANISHI, Hiroyuki)
関西大学・外国語学部・准教授
研究者番号: 30452684

(2)研究分担者

大野 真澄 (ONO, Masumi)
慶應義塾大学・法学部・講師
研究者番号: 50704657

(3)研究協力者

土方(染谷) 裕子 (HIJIKATA-SOMEYA, Yuko)