

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 22 日現在

機関番号：32620

研究種目：若手研究(B)

研究期間：2010～2013

課題番号：22720216

研究課題名(和文)日本人英語学習者におけるスピーキング能力の発達 メタ分析による先行研究の統合

研究課題名(英文)Development of speaking ability among Japanese learners of English: A meta-analysis

研究代表者

小泉 利恵 (Koizumi, Rie)

順天堂大学・医学部・准教授

研究者番号：70433571

交付決定額(研究期間全体)：(直接経費) 1,900,000円、(間接経費) 570,000円

研究成果の概要(和文)：日本人英語学習者のスピーキング能力がどのように伸びるかについて、中学校・高校・大学の文脈において研究されてきたが、伸びのパターンについてまだ一定の傾向が見えてきていない。本研究の目的は、メタ分析を用いることにより、複数の先行研究における結果を統計的に統合し、全体的な発達傾向と、発達に影響すると思われる要因ごとに、発達パターンを吟味することである。

13件の関連文献での情報を用いて分析した結果、流暢さと統語的複雑さにおいて、小から中程度の伸びが見られた。正確さについては、有意な伸びが見られなかった。調整変数分析を行ったが、校種間・指導期間の違いでは、伸びの程度の違いを有意に説明できなかった。

研究成果の概要(英文)：Research into how Japanese learners of English (JLE) develop speaking ability has been conducted, but with mixed results. This study used meta-analysis to explain divergent results among previous studies and to clarify developmental patterns of speaking ability of JLE, taking into consideration moderator variables regarding differences in research contexts (e.g., junior or senior high school) and in instructional lengths (e.g., nine months). After searching for relevant documents using databases, I found 13 studies suitable for meta-analysis, using degrees of differences between pretest and posttest scores.

Results obtained using the random-effect model showed degrees of development to small or medium degrees in fluency and syntactic complexity but not in accuracy. The moderator variable analyses did not show significant effects on the degree of development, showing the need for further investigation in this area.

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：英語 英語教育学 第二言語習得 言語テスト

### 1. 研究開始当初の背景

日本人英語学習者のスピーキング能力がどのように伸びるかについて、中学校・高校・大学の文脈において研究されてきた。伸びを見る研究手法としては、縦断的に同じ学習者の伸びを見る方法と、横断的に異なる能力の学習者の比較を行う方法があり、それぞれ一長一短がある。Koizumi & Katagiri (2009) では、2つの手法を組み合わせで行い、ある SELHi (Super English Language High School) 指定校英語科において、学年が上がるにつれて流暢さが先に伸び、正確さが後に伸びる傾向を示した。しかし、高校生を対象とした他の研究では流暢さ・正確さ・統語的複雑さ(節使用の多さや文の長さ)が同時期に伸びた研究(Egusa, 2009)や、正確さが伸びなかった研究(Egusa, 2005, 2008)もあり、まだ一定の傾向が見えてきていない。これは中学校(例:小泉・山内, 2003; Kosuge, 2003)・大学(例:Koizumi & Fujimori, 2010; Yashima & Viswat, 1997)での先行研究でも同様である。本研究は、先行研究の不一致を説明し、発達にどのようなパターンがあるかを吟味するために、メタ分析を用いる。

### 2. 研究の目的

本研究は、メタ分析を用いて、日本人英語学習者のスピーキング能力がどのように伸びるかについて、先行研究をまとめ、今後の方向性を示すことが目的である。メタ分析とは、関連した先行研究の知見を統計的に統合する手法(Cooper, Hedges, & Valentine, 2009)である。本研究では、メタ分析を用いることにより、複数の先行研究における結果を統計的に統合し、全体的な発達傾向を示すだけでなく、対象校種による違い・指導期間の長さの違いなど、発達に影響すると思われる要因ごとに、発達パターンを吟味する。

### 3. 研究の方法

第1に、先行研究を包括的・系統的に収集するために、データベースを使用した。特に、Scholarly and Academic Information Navigator (CiNii)、Education Resources Information Center (ERIC)、Linguistics and Language Behavior Abstracts (LLBA)で、[longitudinal, Japanese, speaking]または[longitudinal, Japanese, oral]のキーワードを含む文献を検索し、関連する本や論文の引用文献等も調べた。既出版の研究と未出版の研究を両方対象とした。関連する研究のアブストラクトを読み、少しでも関連しそうな場合には論文を入手し、本文を読んでメタ分析で使用できるかを確認した。

次に、メタ分析に入れるのに十分な情報が載っているかを調べた。今回は、日本人英語学習者が日本の学習環境で2か月以上学んだ際の、事前と事後(複数回測定がある場合は最終測定時)テストにおける変化の程度を分析した。収集した文献を読み、先行研究ごと

にメタ分析に必要な情報(平均値・標準偏差・サンプルサイズ)を抽出し、効果量を算出した。不明な箇所は著者に問い合わせた。本報告書では、(事後テストでの平均値 - 事前テストでの平均値)/(2回分をまとめて計算した標準偏差)の式で計算するCohenの $d$ を用いた。必要な情報が掲載されていない研究を除いた結果、13件の文献が残った。

第2に、タスクに影響されない能力値を算出するために、先行研究(Koizumi, 2008, 2009; Koizumi & Katagiri, 2009)で使用した発話データを、多相ラッシュ分析を用いて再分析した。

第3に、先行研究から得た効果量の量的な統合を行った。統合の際には、Comprehensive Meta-Analysis (Ver. 2.2.050)とランダム効果モデルを使用した。先行研究により、測る観点によって結果が異なることが予想されたため、観点ごとに統合した。効果量の解釈は、Plonsky & Oswald (in press)による同一受験者におけるCohenの $d$ の基準である(0.60~0.99が小、1.00~1.39が中、1.40~が大)を使用した。

### 4. 研究成果

#### (1) 多相ラッシュ分析による再分析

Koizumi & Katagiri (2009)ではスピーキングを3回測定したが、各回で使用するタスクが同じではなく、タスクの違いによって伸びの見え方が異なっている可能性もあった。また、Koizumi (2008)では同一タスクを3回使用したが、タスク1個のみで、使用タスクの影響を受けた伸びの結果が報告されている可能性もあった。その問題点を解決するために、それぞれの測定時のデータを、観点ごとに同一の尺度上に位置付け、厳密な比較が可能になるように、Facets (Version 3.71.1; Linacre, 2013a)を使用し、多相ラッシュ分析を行った。

流暢さは「1分あたりの語数」、正確さは「1 AS ユニットあたりのエラーのない AS ユニット数」、統語的複雑さは2指標使い、「1 AS ユニットあたりの節数(統語的複雑さ1)」と、「1 AS ユニットあたりの語数(統語的複雑さ2)」を用いた。

分析対象者は、以下であった。第一に Koizumi & Katagiri (2009)における高校生で、縦断的に測定した39名と、横断的に測定した115名で、それぞれ4つのタスクの一部のみを受けた。第二に Koizumi (2008)における高校生で、縦断的に測定した8名と、横断的に測定した約73名で、1タスクを受けた。第三に Koizumi (2009)での大学生・社会人の英語初級者から上級者、英語母語話者の計55名で、全員が4タスクを受けた。測定間に1年の間があったため、同じ受験者からのデータを別個に独立と捉えた(Dawson, 2000)。計547個のデータが得られ、それを同一尺度上に値を変換する、尺度化・リンクという手順を取った。欠損値はあったが、

ラッシュ分析では分析可能なため問題なかった。

信頼性については、受験者信頼性は4観点で.78～.81、タスク信頼性は正確さ以外で.61～.80だった。正確さのタスク信頼性は.00で、タスクの違いがほとんどなかったためと思われたが、その後のタスク難易度の分析により、それが支持された。

データがラッシュモデルに適合したかについては、Infit Mean Squaresの値が1.5以内であれば適合したと考えた(Linacre, 2013b)。全ての観点において、タスクの不適合のケースはゼロだった。一方、モデル不適合の受験者については、流暢さにおいては15名(2.74%)、正確さにおいては17名(3.11%)、2指標で見た統語的複雑さでは、共に20名(3.66%)だった。タスクの不適合のケースがゼロだったことから、タスクはモデルに適合し、受験者は少し適合していない者も見られたが、全体には適合していた。そのため、尺度化することで、各観点の、流暢さ、正確さ、統語的複雑さという次元での測定が可能であることが分かった。

本分析は、タスクの影響を調整するために行った。流暢さのタスク難易度を見ると、-0.04～0.06であり、全て標準誤差は0.01だった。つまり、全てのタスクの難易度は平均的で、引き出した「1分あたりの語数」にはほとんど違いがなかったが、誤差の範囲を越えた違いはあった。

正確さのタスク難易度を見ると、-0.01～0.00であり、全て標準誤差は0.02だった。全てのタスクの難易度は平均的で、誤差を考慮しても、引き出した正確さにはほとんど違いがなかったことが分かる。

統語的複雑さ1では-0.07～0.11であり、全て標準誤差は0.05～0.06だった。つまり、全てのタスクの難易度は平均的で、この指標で引き出した統語的複雑さにはほとんど違いがなかったが、誤差の範囲を越えた違いはあった。

統語的複雑さ2では-0.11～0.17であり、4タスクの中では難易度の差が最大だった。全て標準誤差は0.06～0.08だった。全てのタスクの難易度は平均的だったが、この指標で引き出した統語的複雑さには、誤差の範囲を越えた少しの違いがあった。

まとめると、誤差を考慮しても、引き出した正確さにはほとんど違いがなかったのが正確さで、誤差の範囲を越えたがほとんど違いがなかったのが流暢さと統語的複雑さの1指標で、違いが少し見られ、誤差の範囲を越えた少しの違いがあったのが統語的複雑さの1指標であった。全体的には、大きなタスク難易度の違いはなかったものの、タスクの違いは多少あり、本分析を行って値を調整する必要があったことが裏付けられた。

## (2) メタ分析

以下の研究をメタ分析に含めた。流暢さの

分析では、Egusa (2005, 2008, 2009), Gilmore (2011), Koizumi (2008 再分析), Koizumi & Fujimori (2010), Koizumi & Katagiri (2009 再分析), 小泉・山内 (2003), Kosuge (2003, 2011), Nakamura (2006), Taguchi (2011), Takiguchi (2003) の13研究を入れた。正確さの分析では、Egusa (2005, 2008, 2009), Koizumi (2008 再分析), Koizumi & Katagiri (2009 再分析), 小泉・山内 (2003), Kosuge (2011), Takiguchi (2003) の8研究を含めた。1 AS ユニットあたりの節数で測定した統語的複雑さではEgusa (2009), Koizumi (2008 再分析), Koizumi & Katagiri (2009 再分析), Takiguchi (2003) の4研究を入れた。1 AS ユニットあたりの語数で測った統語的複雑さでは、Egusa (2005, 2008, 2009), Koizumi (2008 再分析), Koizumi & Katagiri (2009 再分析), Kosuge (2011), Takiguchi (2003) の7研究を含めた。

表1に示したように、メタ分析の結果、流暢さと統語的複雑さの2指標においては、小から中程度の有意な伸びが見られた。「ユニット当たりの語数」の指標で測定した統語的複雑さが最も伸びていた。正確さについては、95%信頼区間に0を含み、有意な伸びがあったとは言えなかった。

表1 ランダム効果モデルを用いた結果

変数	N	研究数	d (95%信頼区間)	I <sup>2</sup>
流暢さ (時間あたりの語数)	360	13	0.75 (0.49, 1.02)	62.01
正確さ (1単位当たりのエラーのない単位)	233	8	0.04 (-0.41, 0.49)	79.68
統語的複雑さ1 (ユニット当たりの節数)	113	4	1.03 (0.53, 1.54)	66.56
統語的複雑さ2 (ユニット当たりの語数)	162	7	1.38 (0.76, 2.00)	80.88

注. N = サンプルサイズ総数.

一般的に、有意な研究結果は出版されやすく、入手しやすいため、メタ分析では統合結果にバイアスが入る可能性がある。その出版バイアスがどの程度あるかを、Funnel plotと、Duval and Tweedie's trim and fillの値で調べた。その結果、流暢さのみで出版バイアスが見られた。流暢さにおいては、プラスの伸びがなかった研究が5件抜けており、それを考慮して修正すると、 $d = 0.49$  (0.18,

0.79) となり、流暢さは有意だがほとんど伸びなかったという結果だった。そのため、厳密には、流暢さについては結論がまだ明確には出せないと考えられる。しかし、正確さについては、伸びがほとんど見られず、統語的複雑さについては中程度ぐらいの伸びが見られるという全体的な傾向が示されたところに本研究の意義があると考えられる。

研究間の効果量のばらつきの大きさを調べる  $f^2$  値では、25~49 が小、50~74 が中、75 以上を大きな程度と解釈するが、4 観点の値は 62.01~80.88 であり、中程度または大きな程度のばらつきが見られた。表には記載がないが、研究間の効果量のばらつきが有意に大きいかを調べる、 $Q$  値も調べた。その結果、全ての観点において、有意差が見られ、研究間の効果量のばらつきが有意に大きいことが示された (流暢さ:  $Q=31.58$ ,  $df=12$ ,  $p<.01$ ; 正確さ:  $Q=34.44$ ,  $df=7$ ,  $p<.01$ ; 統語的複雑さ 1:  $Q=8.97$ ,  $df=3$ ,  $p=.03$ ; 統語的複雑さ 2:  $Q=31.38$ ,  $df=6$ ,  $p<.01$ )。

高めの  $f^2$  値と有意な  $Q$  値から、他の要因が関わることが示唆され、先行研究からも他要因の存在が導けるため、調整変数分析を行った。今回は、校種・指導期間の 2 変数で分析した。校種の違いは、3 件以上見つかった校種のみを対象にしたところ、表 2 のように、比較が可能だったのは、流暢さでの中高のみだった。そこでも校種間で有意な違いは見られなかった。

表 2 校種での違い (混合効果モデル使用)

変数	校種	研究数	$d$ (95%信頼区間)	$Q$ , $df$ , $p$
流暢さ	中学	7	0.89 (0.29, 1.50)	0.19, 1, .67
	高校	3	0.74 (0.40, 1.07)	
正確さ	高校	5	0.39 (-0.09, 0.87)	--
統語的複雑さ 1	高校	3	0.80 (0.51, 1.10)	--
統語的複雑さ 2	高校	5	0.96 (0.39, 1.52)	--

指導期間の調整変数分析では、単位を月とし、指導期間の長さで、伸びの違いの分散を有意に説明できるかをメタ回帰分析で調べた。結果は、表 3 の Model の行が示すように、どの観点でも有意に説明できたものはなかった。

本報告書では紙面の関係上省略したが、メタ回帰分析の図を見て、メタ回帰直線からかなり離れている、外れ値的なデータを特定した。流暢さでは Egusa (2005) と Kosuge

(2011) が、正確さでは Kosuge (2011) が、統語的複雑さ 1 では Takiguchi (2003) が、統語的複雑さ 2 では Kosuge (2011) が、回帰直線から離れていた。これらの研究を吟味すると、全て、事前と事後の測定時に使用したタスクが同一ではなく、タスクの違いが、スピーキング測定値の違いの一つの要因となって、伸びが大きくまたは小さく出ている可能性が示唆された。しかし、例えば Egusa (2005) のように、流暢さでは回帰直線から離れていたが、正確さではその傾向が見られなかった研究もあり、タスクの違いの影響は一律に出るのではないことも分かった。

表 3 指導期間の違い (ランダム効果回帰の method of moments 使用)

変数	傾き	$p$	切片	$p$
流暢さ	0.01	.38	0.57	.02
Model	$Q=0.78$	$df=1$	$p=.38$	
正確さ	-0.03	.21	-0.52	.26
Model	$Q=1.55$	$df=1$	$p=.21$	
統語的複雑さ 1	-0.06	.30	2.10	.048
Model	$Q=1.06$	$df=1$	$p=.30$	
統語的複雑さ 2	0.01	.78	1.20	.14
Model	$Q=0.08$	$df=1$	$p=.78$	

注. 研究数は表 1 と同じ。

まとめると、調整変数分析を行ったが、校種間・指導期間の違いでは、先行研究間の伸びの程度の違いを有意に説明できなかった。これは、分析で使用できた研究数が少ないためと、本来調整変数による影響が少ないためという 2 つの解釈が考えられる。そのため今後は、メタ分析に含められる、日本人英語学習者のスピーキングの伸びを扱う再現研究を増やし、さらに調べていく必要がある。元 SELhi 校などの研究指定校では、同様の研究結果を報告書にまとめているが、そのための基礎データを持っている場合もあるだろう。全国的に未出版でデータベースにも入っていない報告書を共有したり、研究指定校等の教員と共同研究という形でスピーキングの伸びを確認したりすることで、今後のメタ分析に加えられる研究も増えてくるのではないかとと思われる。

また、メタ回帰分析で行ったように、全体的なパターンとは異なる個々の研究について再度見直し、何か特定されていない要因や複数の要因の交互作用がないかも調べる必要がある。

加えて、今回の報告書では、複数回測定がある場合は最終測定時を事後とし、事前テストとの結果を比較した。今後は複数測定の場合の 2 回目測定、3 回目測定の場合などを分けていくことで、スピーキングの伸びのパターンをさらに詳細に示せるだろう。

## 5. 主な発表論文等

[雑誌論文](計15件)全て査読有

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4, 900-913. doi:10.4304/jltr.4.5.900-913. Retrieved from <http://ojs.academypublisher.com/index.php/jltr/article/view/jltr0405900913/7555>

Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40, 522-532.

doi:<http://dx.doi.org/10.1016/j.system.2012.10.017>

Koizumi, R., & Hirai, A. (2012). Comparing the Story Retelling Speaking Test with other speaking tests. *JALT (Japan Association for Language Teaching) Journal*, 34, 35-59. Retrieved from

<http://jalt-publications.org/jj/articles/1507-comparing-story-retelling-speaking-test-other-speaking-tests>

Koizumi, R., Sakai, H., Ido, T., Ota, H., Hayama, M., Sato, M., & Nemoto, A. (2011). Toward validity argument for test interpretation and use based on scores of a diagnostic grammar test for Japanese learners of English. *Japanese Journal for Research on Testing*, 7, 99-119.

Koizumi, R., Sakai, H., Ido, T., Ota, H., Hayama, M., Sato, M., & Nemoto, A. (2011). Development and validation of a diagnostic grammar test for Japanese learners of English. *Language Assessment Quarterly*, 8, 53-72. doi:10.1080/15434303.2010.536868

[学会発表](計13件)

小泉利恵 (2014). 「スピーキングの分析指標と妥当性: 流暢さ・語彙的複雑さを中心に」外国語教育メディア学会関西支部第7次基礎理論研究部会主催 招待講演. 於: 関西学院大学大阪梅田キャンパス. 2月11日

小泉利恵 (2013) 「教室におけるスピーキングの評価 生徒同士のロール・プレイを中心に」『東京国際大学第11回英語教育改革フォーラム 英語で表現する力の育成と評価 パネルディカッション「産出スキルの指導と評価のあり方を考える」』2013年11月23日

Koizumi, R., & In'nami, Y. (2013). Longitudinal and cross-sectional investigation into the development of speaking ability at Japanese schools. Paper presented at the 35th Language Testing Research Colloquium, Seoul, Korea. 2013年7月5日

Koizumi, R., & In'nami, Y. (2012). Factor structure of fluency, accuracy, syntactic complexity, and lexical diversity of speaking performance: A structural equation modeling approach. Paper Presented at American Association of Applied Linguistics 2012 Conference, Boston, Massachusetts, U.S. 2012年3月27日

[図書](計6件)

Koizumi, R. (2013). Vocabulary and speaking. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* [online edition]. Hoboken, NJ: Wiley-Blackwell: Wiley-Blackwell. doi:10.1002/9781405198431.wbeal1431

In'nami, Y., & Koizumi, R. (2013). Structural equation modeling in educational research: A primer. In M. S. Khine (Ed.), *Applications of Structural equation modelling in educational research and practice* (pp. 23-51). Rotterdam, the Netherlands: Sense Publishers. Retrieved from <https://www.sensepublishers.com/media/1694-application-of-structural-equation-modeling-in-educational-research-and-practice.pdf>

小泉利恵 (2012). 「テキストの難易度の尺度」. 卯城祐司 (編). 『英語リーディングテストの考え方と作り方』(pp. 20-24). 東京: 研究社

小泉利恵 (2011) 「リスニングの測定・評価」石川祥一・西田正・斉田智里 (編) 『英語教育学体系第13巻 テスティングと評価 4 技能の測定から大学入試まで』(pp. 173-187). 東京: 大修館書店

印南洋・小泉利恵・平井明代 (2012). 「メタ分析 複数の研究を統合する」平井明代編著 『教育・心理系研究のためのデータ分析入門 理論と実践から学ぶ SPSS 活用法』(pp. 224-248) 東京: 東京図書

## 6. 研究組織

(1) 研究代表者

小泉 利恵 (KOIZUMI, Rie)  
順天堂大学・医学部・准教授  
研究者番号: 70433571