

小学校外国語科の話すこと（やり取り）の評価を左右する要因は何か

What Factors Influence the Assessment of Speaking Interactions in Elementary School Foreign Language Education?

佐藤 剛*・内海 里菜**・大島梨理香**・大高 智英**

Tsuyoshi SATO*・Rina UTSUMI**・Ririka OSHIMA**・Tomohide OTAKA**

北向 周平**・佐々木駿介**・中村 尚平**・畠山 大輝**

Shuhei KITAMUKI**・Shunsuke SASAKI**・Shohei NAKAMURA**・Taiki HATAKEYAMA**

要旨

本稿は、小学校外国語科における話すこと（やり取り）のパフォーマンステストにおいて、採点者が重視する評価の観点や根拠を明らかにするものである。国立大学教育学部に所属する大学生71名を対象に、児童のパフォーマンスを再現したインタビューのビデオを、総合評価および「文法の正確さ」、「発音の正確さ」、「流暢さ」、「ジェスチャー」、「アイコンタクト」、「声の大きさ」の6つの観点別に評価してもらい重回帰分析によって分析した。結果、決定係数は.53であり、6つの要因で総合評価の半分程度を説明できることが明らかになった。また、標準化偏回帰係数の値から、「文法の正確さ」、「アイコンタクト」、「流暢さ」の順で評価に影響していることが明らかになった。非言語的要因である「アイコンタクト」が評価の大きな要因であることは、話すこと（やり取り）の能力を測定するための評価方法としての妥当性に疑問を残す結果となった。

キーワード：小学校英語 スピーキング 評価

1 はじめに

平成29年に改訂された学習指導要領下で小学校においても外国語が教科化された。それにより、文章によって児童の到達度を評価していたこれまでの外国語活動とは異なり、数値による客観性と透明性の高い評価が求められることになった。外国語科においては「読むこと」と「書くこと」の指導がなされているものの、「聞くこと」と「話すこと」を中心とした授業が行われている小学校においては、ペーパーテストではなく、インタビューやプレゼンテーションなどのパフォーマンステストをもとに評価を行うことが多い。

このような状況下で実施されるスピーキングのパフォーマンステストにおける課題のひとつは、全受験者に対して、妥当性・信頼性の高い平等な評価をすることの難しさにある。ペーパーテストのように学習者

のパフォーマンスを形に残すことが難しいスピーキングのパフォーマンステストの場合、採点者間では言うまでもなく採点者内でも信頼性のある一貫した評価を行うことは非常に難しい。

さらに、評価にあたりどのような観点を設定するかについても、教師の言語観や授業で重視している指導観などによって、個人差があることが多い。文法的な正確さ、流暢さ、発音など言語的側面を重視する評価者がいる一方で、小学校ではアイコンタクト、ジェスチャー、笑顔など非言語的要因を重視する評価者が少なくないこともまた事実である。さらに、上記の非言語的要因を英語の「話すこと」の評価項目とすることが妥当であるかどうかについても疑問が残る。

実際に、評価をする際の観点やそれに基づく基準にも大きなばらつきがあることも課題のひとつである。例えば、2023年度東京都立高等学校入試のスピーキン

* 弘前大学教育学部

* Department of English Education, Faculty of Education, Hirosaki University

** 弘前大学教育学部学校教育教員養成課程

** Teacher Training Division, Faculty of Education, Hirosaki University

グテストに活用された ESAT-J の評価基準を見ると、音声の録音を評価する形式のテストであることの影響もあるかもしれないが、英語の発話技能のみが評価されるようになってきている。その一方で、大学入試共通テストへの活用も検討された、実用英語技能検定のスピーキングテストでは、コミュニケーションを図ろうとする態度を評価する項目が設けられているなど試験によってそれぞれである。特に、小学校におけるスピーキングテストでも声の大きさやアイコンタクトなどが評価の対象となる事も多い。しかし、前述の通り、これらの非言語的要因が「妥当性・信頼性の高い評価が可能な英語力」と言えるのかについては検討の余地が残されている。

以上のことから本研究では、事前に明確な評価基準を示さなかった場合、評価者が総合評価をする際に、どんな観点に影響を受けるのかについてインタビューテストの再現ビデオとその評価を用いて調査する。

2 先行研究

妥当性と信頼性の高いスピーキングの評価が難しいことは多くの先行研究で指摘されている。稲毛他(2006)では、日本人英語教師と英語母語話者の英語教師の間で、また日本語教師間でも、中学英語教師、高校英語教師、大学英語教師の間で何らかの差が出てくるか、また教師経験年数などによりどのような評価の差が出てくるのかを分析することを目的とし、大学生15名(学部生3~4年生、大学院生1年生)を対象にスピーキングテストを実施し、スピーキングテストを実施する際には、事前にいかに慎重に評価の個人差を縮小するかが、評価の客観性に大きく影響すると述べている。

萬谷(1982)は、Fluencyには様々な要素が絡み合っているため、定義づけを行うことは極めて困難であると指摘している。理由として、話すというのは、不規則で予想困難であり、どの観点から見ても良い話し方というのを決めることが難しいと考えられるだろう。

片桐(1999)は、日本人英語教師による英語学習者のスピーキングの評価に影響を与える可能性のある主な要因をAETの評価と比較しながら追及した。結果から、日本人英語教師とAETの英語学習者のスピーキング評価は、発話速度が評価に影響を与える最も重要な要因であるとされ、ローカルエラーの数が2番目に重要な要因であるとされた。一定の傾向はあるものの、評価者によってばらつきがあるため、スピーキン

グ評価の信頼性のためには、複数の評価者で行う必要があるとしている。

小学校における英語教育の関心の高まりから、小学生を対象にした研究は近年増加傾向にある。話し言葉中心の指導が行われている小学校英語に関する研究にはスピーキングに関する研究も数多くなされている。岡崎他(2021)は、小学校外国語科で実施されているスピーキング・パフォーマンス評価が授業における学習活動や海外交流を通して培った児童のスピーキング力を測るのに適しているかを明らかにすることを目的のひとつとして実践研究を行った。態度(アイコンタクト、声の大きさ)、流暢さ、正確さ、応答、質問の5つの観点から採点が行われたものを、妥当性、信頼性、実用性、波及効果の面から分析検討した。その結果、テストの妥当性、信頼性、実用性が確保されていること、波及効果として児童の学習及び英語コミュニケーションへの意欲向上に影響を与えたことがわかった。

田村(2022)は教科としての英語の評価に関して、小学校教員96名に意識調査を行った。その結果、質問紙には「話すこと」に関してのものが多く寄せられた。具体的には「ペーパーテストで評価ができない「話すこと」についてどのように見取ればよいか」などといった意見が見られた。また、話すことに関して、「話すこと(やり取り)」の指導・評価について難しいと感じている教員は全体の64.4%となり、「話すこと(発表)」の指導・評価について難しいと感じている教員は全体の38.7%となった。田村は児童が(やり取り)を行っているような場合、個々の児童を見取ることができないと感じている教員も多く、記録方法について学ぶ研修機会を設けていく必要があるとし、評価を行う際は、学年内で十分にモデレーションを行うなど、評価の妥当性・信頼性を高める取り組みを行うべきであるとした。

妥当性と信頼性の高い評価には、どのような観点で評価を行うのが重要である。言い換えれば、スピーキングの構成要素は何かを明確にすることが求められる。多くの研究によって採用されるのは、正確さ(accuracy)、複雑さ(complexity)、流暢さ(fluency)(CAF)である。多くの先行研究において話し言葉の熟達度とCAFとの間に強い相関関係があることが示されている(Najmi, et al., 2021; Foster & Skehan, 1996; Housen & Kuiken, 2009)。正確さは「特定の基準からの逸脱の程度」(Housen & Kuiken, 2009, p. 463)、または「エラーの少なさ」に焦点を当てたものであ

る (Foster & Skehan, 1996, p. 304)。複雑さは、構文や語彙の面で言語の緻密さ、豊かさ、多様性に関する基準である (Foster & Skehan, 1996; Housen & Kuiken, 2009)。最後に、流暢さは、話しや文章の流れ、流麗さ、滑らかさに焦点を当てた基準である (Housen & Kuiken, 2009, p. 463)。

横山 (2019) は、日本語の Academic Speaking の評価基準を作成すること、その妥当性を検証することを目的として研究を行った。話し方、語彙、文法という言語面に対して流暢さ、複雑さ、正確さの指標で分析を行い検証した結果、評価には発話の流暢さよりも、発話の複雑さ、また文法的正確さが関連してくることが明らかになった。

今井・吉田 (2007) によると、スピーキングテストにおける評価の観点は「機能 (Function)」、「発話内容 (Content)」、「発話の複雑さ (Text type)」、「発話の理解度 (Comprehensibility)」であり、「機能」は「ことばを使って何ができるか」、「発話内容」とは「どれくらい多様な内容について話すことができるか」、「発話の複雑さ」は「発話がどのくらい複雑な構造であるか」、「発話の理解度」は「聞き手にとってのわかりやすさの観点」であり、これには発音・文法の正しさ・正確さ・流暢さの下位観点がある。

このように CAF など客観的な数的データに基づいて厳密な基準に評価をすることができれば、妥当性・信頼性の高いスピーキングの評価をすることは可能であるように考えられる。しかし、実際の小学校におけるスピーキングのパフォーマンスを毎回このような厳密なデータに基づいて評価することはその現実性の点から難しいこともまた事実である。さらに、先に述べたように、小学校では CAF のような言語的要因だけでなく、アイコンタクト、ジェスチャー、笑顔などのような非言語的要因を重視する評価者が少なくないことも、その課題のひとつである。

3 リサーチクエスション

本稿は、小学校外国語科における話すこと（やり取り）のパフォーマンステストにおいて、採点者が重視する評価の観点や根拠を明らかにするものである。そのため、以下の2つのリサーチクエスション (RQs) を設定した。

RQ 1. 小学校の外国語科における話すこと（やり取り）のパフォーマンステストにおいて、採点者が重視する観点は何か？

RQ 2. 小学校の外国語科における話すこと（やり取り）のパフォーマンステストにおいて、採点者はどのようなことを根拠に高い評価／低い評価をしているか？

4 研究方法

4.1 実験協力者

本研究の実験協力者は、国立大学・教育学部小学校コースに所属する大学生2年生である。実験協力への同意を得られた大学生は合計71名であり、その多くは、将来小学校の教員志望をしており、教育への関心が強く、小学校外国語の教育法の授業を履修しスピーキングの評価についてもある程度の背景知識をもっている。

4.2 実験器具

本研究を行うにあたり、実験協力者が評価するパフォーマンスのビデオの作成を行った。ビデオは小学校での児童と教師との英語のパフォーマンステストを教師と大学生が演じる形をとった。

作成したビデオは、文法の正確さ、発音の正確さ、流暢さ、ジェスチャー、アイコンタクト、声の大きさの6つの観点のいずれかが著しく劣っている状況を再現した6種類のパフォーマンスに加えて、すべての観点を満たしたパフォーマンスを再現したものの合計7種類である。7人それぞれのパフォーマンスの詳細は以下のとおりである。

- (1) 6 観点全てを満たした児童のパフォーマンス (以下 PF)
- (2) 発音が正確かつ流暢であり、アイコンタクトを取りながらジェスチャーをつけて十分な声量で発話しているが、語順の間違いを多く含む文法の正確さが著しく劣っている児童のパフォーマンス (以下 GA)
(例) *Very much volleyball I like. Yes, do I.*
- (3) 文法的な面では正確かつ流暢であり、ジェスチャーをつけながらアイコンタクトを取りながら十分な声量で発話しているが、カタカナ発音での応答する発音の正確さが著しく劣っている児童のパフォーマンス (以下 PA)
(例) ハロー、アイ ライク バレーボール
- (4) 文法的・発音の面でも正確であり、アイコンタクトを取りながらジェスチャーをつけて十分な声量で発話しているが、教師から質問をされ

た後、答えるまでに毎回5秒程度の間が入る流暢さが著しく劣っている児童のパフォーマンス (以下 *FL*)

- (5) 文法的・発音の面でも正確かつ流暢であり、アイコンタクトを取りながら十分な声量で発話しているが、一切ジェスチャーをつけずに応答する児童のパフォーマンス (以下 *GT*)
- (6) 文法的・発音の面でも正確かつ流暢であり、ジェスチャーをつけて十分な声量で発話しているが、パフォーマンスを通して、一度も教師とアイコンタクトをとらない児童のパフォーマンス (以下 *EY*)
- (7) 文法的・発音の面でも正確かつ流暢であり、アイコンタクトを取りながらジェスチャーをつけて十分な声量で発話しているが、全体をとおして声量が不十分な児童のパフォーマンス (以下 *VL*)

本研究は上記6つの観点が採点者の総合評価に与える影響を調査するものである。よって発話内容、発話の構成、発話の長さ、発話における統語的・語彙的複雑さなど、他の要因の影響を避けるために、話す内容は7人とも以下に示す共通のものとした。

教師役：Hello.

児童役：Hello.

教師役：What is your name?

児童役：I'm ○○.

教師役：How do you spell your name?

児童役：○○○○.

教師役：What sports do you like?

児童役：I like XXX.

教師役：Can you play XXX well?

児童役：No, I can't.

教師役：Do you watch XXX on TV?

児童役：Yes, I do.

教師役：Question please.

児童役：Do you like XXX?

教師役：Yes, I do. Thank you very much.

児童役：You're welcome.

さらに、7人の児童役のやりとりの総合評価と、6つの観点についての観点別評価を実験協力者に行ってもらうため、グーグルフォームを用いて4件法(4=とても高く評価をつける、3=高く評価をつける、2

=低く評価をつける、1=とても低く評価をつける)のアンケートを作成した。加えて、なぜそのように評価したのかの理由を記述してもらうための自由記述欄を設けた。

4.3 実験手続き

本実験はすべてMicrosoft Teamsを使ったオンデマンド形式で実施した。前述の通り評価の客観性を担保することが難しいスピーキングの評価においては、繰り返し確認することができるように、受験者のパフォーマンスを録画することが多い。また、オンデマンドの方法をとることで、実験協力者はそれぞれ十分な時間を確保して、納得がいくまで繰り返し対象のビデオを見るのが可能になり、より正確なデータを得ることが可能になると考えられる。上記7種類のパフォーマンスビデオをTeams上で共有し、実験協力者に対して以下のような指示を与えた。

Teamsの「ファイル」に7人の自己紹介のビデオがアップされています。小学生が「Unit 1 自己紹介をしよう」の単元末に実施したパフォーマンステストを録画したビデオだという想定で評価して下さい。

その上で、実験協力者にまずそれぞれ7種類のパフォーマンスを4段階で総合的に評価し、自由記述欄に評価した理由を記述するよう指示をした。その後、7種類のパフォーマンスを再度見て、6つの観点別に(文法の正確さ、発音の正確さ、流暢さ、ジェスチャー、アイコンタクト、声の大きさ)4段階で評価するように指示をした。最後に評価をやってみて、気がついたことや難しかったことなどを記入するように指示をした。

4.4 データ分析

データ分析に当たっては、実験協力者から得たデータをcsvファイルに変換し、JASP (version 0.17.1, JASP Team, 2023)を用いて、総合評価を従属変数、6つの観点を独立変数として重回帰分析を行った。その結果をもとに、文法の正確さ、発音の正確さ、流暢さ、ジェスチャー、アイコンタクト、声の大きさの6項目について、実験協力者が何を重視して評価を行っているのかについて、比較・考察を行った。また、6項目以外に被験者の採点に与える要因について考察を行った。その後、テキストマイニング用のソフトウェ

アである KH Coder (version3. Bata04, 樋口, 2020) を使用し, 被験者の自由記述欄のデータを 4 段階の総合評価を外部変数として対応分析を行うことで, 実験協力者が何を基準に総合評価をしているのか分析を行った。

5 結果と考察

7 種類のスピーキングのパフォーマンスに対する 4 段階の評価の平均値, 標準偏差, 最小値, 最大値, 95%信頼区間は以下の表 1 のとおりである。すべての観点を満たしたパフォーマンス (PF) には実験協力者全員が 4 の評価をつけていることが分かる。一方で他の 6 種類のパフォーマンスについては, 2 から 3 の評価を平均に採点者によって大きなばらつきがあり, 4 の評価をした採点者から 1 の評価をつけた採点者まで大きな幅があることが分かる。また, Cronbach's coefficient $\alpha = .59$ であり, 採点の信頼性の低さを示す結果となった。

表 1
記述統計量

	M	SD	Min	Max	95% CI
PF	4.00	0.00	4.00	4.00	[4.00, 4.00]
GA	2.24	0.73	1.00	4.00	[2.39, 2.08]
PA	2.76	0.57	2.00	4.00	[2.87, 2.63]
FL	2.47	0.55	1.00	4.00	[2.58, 2.35]
GT	2.97	0.54	1.00	4.00	[3.07, 2.85]
EY	2.39	0.64	1.00	4.00	[2.52, 2.25]
VL	3.16	0.65	1.00	4.00	[3.29, 3.02]

注: M = 平均値, SD = 標準偏差, Min = 最小値, Max = 最大値, CI = confidence interval

5.2 総合評価と文法の正確さ, 発音の正確さ, 流暢さ, ジェスチャー, アイコンタクト, 声の大きさの関連

総合評価に対する 6 つの観点 (文法の正確さ, 発音の正確さ, 流暢さ, ジェスチャー, アイコンタクト, 声の大きさ) の影響を検討するために, 重回帰分析を行った。具体的には総合評価を従属変数, 6 つの観点を独立変数にして重回帰分析を行った。その結果は以下の表 2 に示すとおりである。

独立変数である 6 つの観点 (文法の正確さ, 発音の正確さ, 流暢さ, ジェスチャー, アイコンタクト, 声の大きさ) はいずれも有意であり, その標準化偏回帰係数は, 文法の正確さ ($\beta = 0.43$), アイコンタクト ($\beta = 0.38$), 流暢さ ($\beta = 0.23$), ジェスチャー ($\beta =$

0.18), 発音の正確さ ($\beta = 0.16$), 声の大きさ ($\beta = 0.15$) の順に高い結果となった。このことから, 実験協力者は主に文法的な正確さとアイコンタクト, 流暢さを基準にスピーキングの評価をしていることが示された。

決定係数 $R^2 = .53$ であり, 分散分析による有意性検定の結果, 有意であることが示された ($F(6, 601) = 114.34, p < .001$)。このことから, 上記 6 つの独立変数により, 従属変数である総合評価の 53% を説明できることが明らかになった。

それでは, 実験協力者は 6 つの観点 (文法の正確さ, 発音の正確さ, 流暢さ, ジェスチャー, アイコンタクト, 声の大きさ) の他にどのようなことを根拠にスピーキングの評価を行っているのだろうか。それを検証するために, なぜその評価をしたのかに対する記述文の分析を行った。

表 2
総合評価を従属変数にした重回帰分析の結果

変数	B	SEB	β
文法の正確さ	0.34	0.02	0.43 **
発音の正確さ	0.14	0.03	0.16 **
流暢さ	0.17	0.03	0.23 **
ジェスチャー	0.13	0.02	1.18 **
アイコンタクト	0.28	0.02	0.38 **
声の大きさ	0.13	0.03	0.15 **
切片	-1.07	0.02	

注: $R^2 = .53, **p < .001$, B = 偏回帰係数, SEB = 標準誤差, β = 標準化偏回帰係数

5.3 採点者がどのような根拠でスピーキングのパフォーマンスを評価しているか

KH Coder (version3. Bata04, 樋口, 2020) を使用し, 実験協力者の自由記述欄のデータから, 被験者が何を根拠に採点をしているのか考察を行った。使用した分析方法は対応分析であり, これによりクロス集計に基づいて分析された結果を可視化することができる。以下の図 1 は, KH Coder によって出力された対応分析の結果である。

外部変数は被験者による 4 段階のスピーチの評価 (4 ~ 1) であり, 四角形で示されている。対応分析では, それぞれの外部変数に関連が強いものほど原点から外側に, どの外部変数にも共通するものは原点付近 (0.0) に配置される。

具体的には外部変数 4 の外側には, 「明るい」, 「表情」, 「自分」, 「素晴らしい」, 「笑顔」, 「抑揚」, 「豊

いう意識があるかどうかをきちんと判断できるようにしたことだ。流暢さについては、正直なところ限界があると思っている。小学校の外国語で、完璧な発音、正確な言葉で、流暢にというよりは、しっかり相手と向き合って話ができるというところに主眼を置いて評価した。

発音や、質問に対する回答などの、英語を話すことの技術面に偏った評価をせず、英語が苦手な子供でも高く評価されることができるよう、表情や目線、ジェスチャーや、声のトーンにも注目して見ました。

実験協力者が上記のような「笑顔」、「気持ち」、「表情」、「明るい」、「感情」、「豊か」のような英語の運用能力としては直接的に関係がない要因をもとに評価することの理由のひとつは、英語が苦手な学習者への配慮や、できるだけ学習者の良いところを認め、評価したいという思いがあるからであると考えられる。

6 まとめと教育的示唆

本研究は、小学校外国語科における話すこと（やり取り）のパフォーマンステストにおいて、採点者が重視する評価の観点や根拠を明らかにするものである。その結果、以下のことが明らかになった。まず、本校で実施した実験では事前に具体的な評価基準を伝えなかったため、採点者間の評価に大きな差が生まれた（Cronbach's coefficient $\alpha = .59$ ）。また、重回帰分析の結果、それぞれの要因の標準化偏回帰係数の値から、「文法の正確さ」、「アイコンタクト」、「流暢さ」、「ジェスチャー」、「発音の正確さ」、「声の大きさ」の順で調査協力者の評価に影響していることが示された。調査協力者の評価の理由に関する記述をテキストマイニングで分析した結果からも本研究の実験協力者は「アイコンタクト」や「表情」、「明るさ」などの非言語的要素を想像以上に重視していたことが分かった。今回の実験協力者の自由記述文からは、このような非言語的要素を評価の重要な要因として考える理由として、英語が苦手な学習者への配慮や、できるだけ学習者の良いところを認め、評価したいという思いがあることが推測された。たとえ文法的な誤りや間違いがあったり、発音が不十分であったりしても、アイコンタクトやジェスチャーなどの非言語的要素を活用した相手に伝えようとするその態度によって総合評価が上がっているケースが多く見受けられた。評価が学習者の情

意面・動機づけの面に与える影響へ配慮することは重要であり、実際の言語使用場面において、意思伝達を図るために非言語的要素が果たす役割は大きいことが事実であるが、英語の熟達度を測ることを目的とした試験で、「笑顔」、「表情」は言うまでもなく「アイコンタクト」や「ジェスチャー」など非言語的要素がどの程度まで重視されるべきであるかについては、評価の妥当性の観点から十分に検討がなされるべきであろう。特に、本研究の結果が示すように、非言語的要素がCAFのような言語的要素よりも強く評価に影響することは避けるべきであろう。さらに、本研究の結果から、このような非言語的要素を評価するにあたっては、信頼性の面でも大きな問題があることが示された。

スピーキングテストが高等学校や大学の入学試験にも活用され、より評価の妥当性・信頼性が重視されるようになった今、学校での試験と入試との一貫性は大きな課題となりそうである。英語教育に関わる教員は、スピーキングの評価において言語的要素と非言語的要素とのバランス、妥当性・信頼性の高い適切な評価方法について、多角的に検討することが必要であると考える。

参考文献

- 稲毛逸郎・ローン悦子・ローンマリー（2006）.「英語スピーキングテストにおける 英語母語話者と日本人の評価傾向に関する研究：面接試験と英語教育等に関するアンケートの分析をもとに」『長崎大学教育学部紀要・教科教育学』47 129-143.
- 今井裕之・吉田達弘（編）（2007）.『HOPE：中高生のための英語スピーキングテスト』教育出版
- 岡崎 浩幸・清水 義彦・押田 正子（2020）.「小学校外国語科スピーキング・パフォーマンス評価に関する実践研究」『富山大学人間発達科学部紀要』15（2）, 117-127.
- 片桐一彦（1999）.「スピーキング評価：日本人英語教師 vs. AET」『全国英語教育学会紀要』10, 93-102.
- 田村 岳充（2022）.「教科としての英語の「評価」における小学校教員の意識調査」『中部地区英語教育学会紀要』51, 101-108.
- 樋口耕一（2020）.「社会調査のための計量テキスト分析—内容分析の継承と発展を目指して—第2版」ナカニシヤ出版.
- 横山千聖（2019）.「日本語学習者の Academic Speaking の評価基準の構築と妥当性の検証」『広島大学大学院教育学研究科紀要』2（68）, 205-214.
- 萬谷隆一（1982）.「Speaking の能力とその評価について：Fluency の観点から」『中国地区英語教育学会研究紀要』, 12, 15-18. https://doi.org/10.18983/casele.12.0_15.

Housen, A., & Kuiken, F. I. (2009) . Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30 (4) , 461–473.

<https://doi.org/10.1093/applin/amp048>

JASP Team (2023) . JASP (Version 0.17.1) [Computer software].

Najmi, K., Siyyari, M., Abassian, G. (2020) . The Relationship Between Three Components of CAF in Learners' Oral

Performance at Different Proficiency Levels. *Journal of English Language Pedagogy and Practice*, 13 (26) , 254–284. doi: 10.30495/jal.2020.683555

Foster, P., & Skehan, P. (1996) . The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, 18 (3) , 299–323. doi:10.1017/S0272263100015047

(2023 . 8 . 25 受理)