

スピーキング活動「やり取り」の指導と評価方法の研究

—自己、相互、教員評価から—

英語科 川上佳則

英語の「発信力」のみならず、そのタイトルから「論理」を強く意識させる新課程の論理・表現 I の授業において、スピーキング活動「やり取り」に焦点を当てた指導と評価方法を研究実践した。特に指導と評価の一体化を意識し、定期的な教員評価に日常的な生徒自己評価と相互評価を加味することで、持続可能な形で評価機会の充実を図る。また、生徒自身が評価者として成長することで、客観的に評価観点の理解を深めながら、自己の取組を主体的に改善する動機づけにしたい。

1 年分の各評価データから、今後参考とすべき評価方法に関する有用な教示を得た。また、生徒の意識アンケート分析からは、彼らのやり取り活動に対する前向きな変化が見られた。

キーワード：スピーキング「やり取り」 生徒相互評価 教員評価 評価者の育成

1. 研究の背景と目的

平成 28 年 12 月の中央教育審議会答申を踏まえ、平成 30 年より学習指導要領の改訂が行われた。2022 年から年次進行で段階的に適用されるその内容の中でも、今回はスピーキング活動の中の「やり取り」に焦点を当てた試みを実施することにした。統合的に育成すべき言語活動として「聞くこと」、「読むこと」、「話すこと [やり取り]」、「話すこと [発表]」、「書くこと」の五つの領域が示されたとき、特に「話すこと [やり取り]」についてその指導と評価の難しさを直感した教員は私たちだけではないだろう。一方、既存のやり方だけではいわゆる「主体的・対話的で深い学び」に裏付けられたより実践的コミュニケーション活動を推進することが困難であることもまた、同時に感じつつあったことのように思える。

本実践は「話すこと」、「書くこと」を中心に発信力の強化を図る科目として設定された「論理・表現 I」において、思考・判断・表現を「論理の構成や展開を工夫して」繰り返す経験ができるような仕掛けを指導と評価のパッケージで複数回設定し、簡潔かつ持続可能な方法で授業計画に組み込むことを試みる。

また、生徒の活動頻度の増加に比例して増す教員評価の負担感が実践の妨げにならないよう、生徒自身による評価活動方法を事前に指導・促進した場合、生徒評価が教員評価だけに頼らない妥当な評価として扱えるかどうか検証する。

2. 研究対象

本校で論理・表現 I（2 単位）を受講しているのは高校一年生 120 名（30 名×4 クラス）であるが、当該学年は論理・表現 I を 15 名ずつの分割授業として実施しており、そのうち研究者が担当する 3 クラス計 45 名（男子 20 名・女子 25 名）を研究対象とした。

3. 指導実践期間

今回の指導を実践する期間は、2022 年 4 月初旬から 2023 年 3 月である。なお、本指導研究は翌年度の論理・表現Ⅱ以降も継続して実施する予定である。

4. 先行研究

現在、国内にも多種多様なスピーキング指導と評価に関する研究実践（深澤, 2010, Nakamura, 2002, 高井. et. al, 2019）がある一方、スピーキングの「やり取り」活動、その指導と評価の一体化、またその妥当性に焦点を当てたものはそれほど多くない。Kusumayanthi (2022) は、教員の評価は必要とする一方、生徒の相互フィードバック後に、多くの生徒がスピーチのグレードを上げたと報告する。また、妥当性自体、信頼性を含む複雑なもの（Bachmann, 1996）であり、評価方法自体の妥当性を検討するとき、Messick (1996) が分類する 6 種の側面全てについて検証することは難しいが、少なくとも「内容的」、「実質的」、「構造的」側面については、十分意識しつつ教員、生徒相互評価ともに研究を進める必要がある。

5. 仮説提示

- (1) 生徒にスピーキング「やり取り」の活動に関するルーブリックを明示し、評価のポイントを指導した上で活動を三人一組で行い、生徒がその取組を相互評価と自己評価を中心に継続して複数回振り返ることで、教員評価と相まって活動への意欲と自信は高まる。
- (2) 平均化された生徒の自己評価や相互評価に一定の妥当性があれば、独立して最終的な評価に反映する割合を増やすことができ、学期に複数回評価活動を実施しても、教員の評価負担は軽減される。

6. 実践方法

6.1 実践の手順

授業はテキストを進めながら、あらかじめ計画した課（2 課毎に 1 度を基本とする）の終わりに、課内で学ぶ内容を元にしたスピーキング「やり取り」活動を設定する。生徒は三人一組で相互にパフォーマンスをチェックし合う形式（以下パフォーマンスチェック）で活動する。課毎に作成するルーブリックと評価シートはパフォーマンスチェックシート（図 1）に全て記載し、各課の最初に生徒に配付して、指示説明する。生徒は順に活動と相互評価を行い、最後に自己評価とコメントを記入し提出する。やり取り活動は全て時間制限を設けて実施し、生徒の活動中は担当教員がタイムキーパーとして活動を支える。

<p>Lesson 9 Performance Check 相互評価シート</p> <p>今回の Speaking(やりとり)Theme: ..聞いたことを確認で繰り返し、さらに聞く</p> <p>※ 今回は、質問者 A を評価します。a healthy life についてアドバイスを求めましょう。 また、質問者はもらったアドバイスは確認するように繰り返すこと。詳しくは以下のダイアログを参照。</p> <p>A: What is your advice for a healthy life? B: →問いに対する B の助言 A: →B の助言を確認するように1度簡単に繰り返し。*評価のポイント…評価者はここをしっかりと☑ Is there anything you need to do to stay healthy? B: →問いに対する B の助言 A: →A の感想 ex.) I see. Sounds great! / That's good. Thanks. / Good idea. I'll try it.</p>

図 1 パフォーマンスチェックシート抜粋

6.2 指導上の工夫

6.2.1 三人一組活動（トリオワーク）の実施

本実践では相互評価の頻度を上げるために、活動を三人一組（以下トリオワーク）で実践する。生徒Aと生徒Bが指定されたやり取りを行い、生徒Cがその場で評価する。なお、やり取りで生徒Cが評価するのは、生徒Aのみである。やり取りは1セッションあたり1～2分間程度で、これを三人ローテーションで3セット行う。基本、毎回授業の最初から最後まで、やり取り活動以外でも常に三人が机を合わせて協働的に学ぶことになる。三人での実施は考え得る中で最もミニマムな規模であり、授業中に評価活動を実施する時間的な負担をできるだけ少なくすることで、年間の授業計画を逼迫しないよう設計した。

6.2.2 評価基準の構成

課末に行うやり取り活動では、生徒の自己評価と相互評価結果をワークシートに記録する。生徒が判断に迷うことなく評価し合えるように、Volume、Eye Contact、Logic の3項目について Excellent-Good-Poor の3段階で評価するよう設定した。また、具体的な評価基準を明示したものをパフォーマンスチェック毎に配付した。また、プリントによる提示に加えて、口頭でも具体的な評価のポイント（特に文章だけでは伝わり切らない内容）について指導・助言を行った。なお、Logic 項目のループリックのみ毎回作成し、それ以外の項目は当初に作成した内容を継続して示した。

6.2.3 論理性

今回特に留意したのは、評価活動を通して生徒が論理性を表す Logic の観点を理解して、評価者として成長できるかという点である。課毎で異なるやり取り活動を行う以上、それぞれ異なる Logic の観点を設定する必要がある、これが本実践の要となる部分であることから、毎週の科会で観点の妥当性を検討した上で決定した。論理の評価基準は、恣意的な判断を避けられるように、なるべく抽象性を抑えた具体的な基準になるよう工夫したものを提示し、担当者間で共有した。（図2）

9. 項目別評価基準			
①Volume.			
Excellent	3 points	:一貫して力強く、とても明瞭に聞こえる。	
Good	2 points	:聴き取りづらいところもあるが、内容は分かる。	
Poor	1 points	:何を言っているのか聞き取れないことが多い。	
②Eye Contact.			
Excellent	3 points	:聞き手の理解を図るように、視線でも意思疎通できる。	
Good	2 points	:常に聞き手を見ようと意識して話そうとしている。	
Poor	1 points	:全く聞き手の方を見ていない。	
③Logic *この記述をベースに、テーマによって多少足し引きし直して提示する。			
Excellent	3 points	:内容を深める問いかけができ、対話の展開を工夫する様子がよく分かる。	
Good	2 points	:簡単な問いかけができ、対話の展開を意識する様子が分かる。	
Poor	1 points	:内容が単調で、相手への問いかけが少なく、対話の展開がほとんどない。	

図2 担当者間で共有したガイダンス

6.2.4 教員評価

なお、生徒がトリオワークで行う評価活動以外に、複数の教員が1人の生徒を評価する機会（パフォーマンスチェック with teachers）を1、2学期に1回ずつ設けた。評価の観点と段階は生徒とほぼ同じものを利用した。教員評価は平均化した数値を生徒が自己評価したパフォーマンスチェックシートの上に赤で転記して、生徒にフィードバックする。この教員評価を、生徒の自己評価・相互評価と比較し、生徒主体の評価活動の妥当性を検証する。

7. データ収集と分析の方法

7.1 パフォーマンスチェックシートと教員評価

全6回のパフォーマンスチェック毎に、生徒が相互評価と自己評価を記録したものを回収してPDF化する。自己評価に合わせて振り返りとして書かせたコメントも生徒意識の変化を探るデータとして検証する。前述したように教員評価は個別に評価し平均化したものを記録し、生徒評価と比較検証する。

7.2 生徒評価の妥当性について

生徒が行う相互評価の妥当性については、十分な妥当性があるとする先行研究（深澤, 2010, Nakamura, 2002）を参考に本実践を進める。今回はその検証方法として、パフォーマンスチェックで行った生徒の自己評価・相互評価と教員評価とデータの相関を見る。先行研究との違いは、異なるやり取り活動を生徒、教員それぞれが評価した内容を比較検証する点である。生徒評価を平均化して教員評価との相関が高ければ、生徒自身が行う評価に一定の妥当性があると考えてよいだろう。なお、相関係数の高低だけでなく、生徒の評価経験の積み上げによる評価者としての成長につながるような変化が見られるか併せてチェックする。

7.3 生徒意識アンケート

2022年の4月、7月、12月の3回、第1学年生徒全員に論理・表現Ⅰの授業に関する意識アンケートを実施した。年度当初は中学時代に感じた英語の難しさや活動経験等を問う項目を付け加えたが、アンケートで継続して問い続けた内容は以下の3点である。（図3）なお、アンケートは全て本校で利用しているICT支援システム「Classi」を利用し、オンラインで配信・集計した。

- | | |
|----|--|
| 問1 | あなたはしっかりと準備ができれば、英語で「日常的なこと」について、情報や考え、気持ちなどを英語で話して伝え合ったり、やり取りを通して必要な情報を得たりすることができますか。 |
| 問2 | あなたはしっかりと準備ができれば、英語で「社会的な話題」について、聞いたり読んだりしたことを活用しながら、意見や主張などを論理の構成や展開を工夫して話して伝え合うことができますか。 |
| 問3 | あなたは英語でやりとりをするときは論理の構成や展開を工夫しましょうと言われた時、その意味を理解し、自ら工夫することができますか。 |

図3 生徒意識アンケート内容抜粋

8. 結果

8.1 生徒評価と教員評価の相関

生徒が行うやり取り活動とその評価活動は、計7回実施した。その内、課末に行った生徒間活動が5回、教員相手に行ったのが2回（内1回はALT含む）である。今回集約したデータから検証できた相関は、一部の自己評価において一定の相関（0.5以上）が見られたものの、相互評価と教員評価の相関はどのクラスにおいても総じて弱く、先行研究で多く見られたような強い相関は見られなかった。（表1）

表1 クラス別に見た教員評価との相関係数

評価\組	A組(n=15)	B組(n=15)	C組(n=15)
生徒相互評価	0.39	0.37	0.28
生徒自己評価	0.12	0.66	0.46

（注）n= 人数

8.2 評価者としての変化

当初から、生徒自身が評価者として評価する力量を伸ばすことを一つのねらいとして進めたこともあり、生徒の評価者としての成長を、評価回数を経て教員評価との相関がどのように変化したのかという観点で検証する。クラス毎にまとめたものが以下の表である。（表2）

表2 クラス別、回数毎に見た教員評価との相関係数

A組\回数	第1回	第2回	第3回	第4回	第5回
生徒相互評価	0.12	0.19	0.50	-0.10	0.34
生徒自己評価	0.06	0.17	0.10	-0.21	-0.25

B組\回数	第1回	第2回	第3回	第4回	第5回
生徒相互評価	0.04	0.38	-0.20	0.49	-0.20
生徒自己評価	0.06	0.47	0.19	0.41	-0.04

C組\回数	第1回	第2回	第3回	第4回	第5回
生徒相互評価	-0.09	-0.25	0.50	0.36	0.28
生徒自己評価	-0.07	0.00	0.61	0.37	0.36

どのクラスにおいても負の相関から正の相関までその変化は一定でなく、クラス、回数どちらについても何とも判断し難い拡散した結果となった。

8.3 生徒意識アンケート結果

4月から12月まで3度実施した生徒意識アンケートは以下の結果となった。結論から言えば、およそ9ヶ月間の活動経験を経て、生徒たちの意識に論理・表現Ⅰの目標として設定されている「日常的なこと」と「社会的な話題」に関するやり取りについて、また「論理の構成や展開の工夫」への理解と実践に関しても肯定的な意識の変化が見られた。

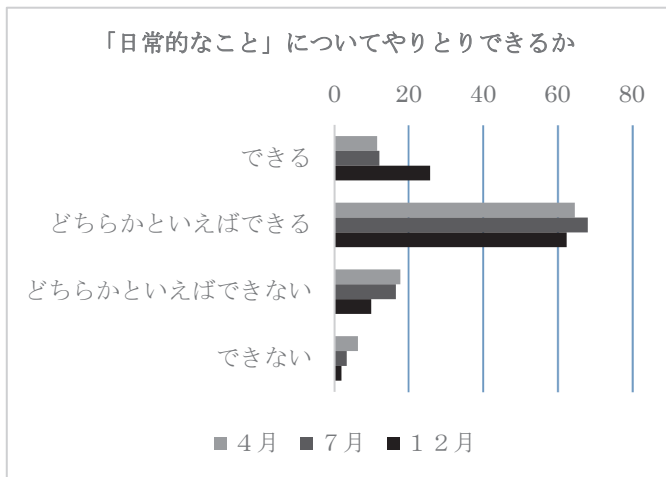


図4 日常的なことに関するアンケート結果

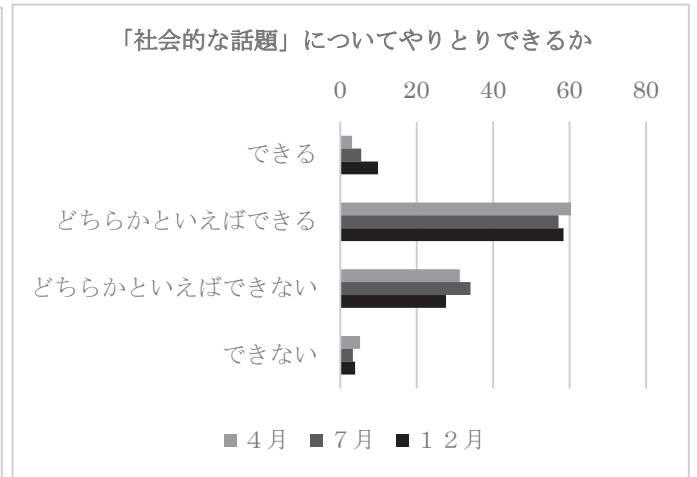


図5 社会的な話題に関するアンケート結果

まず、日常的なことについて（図4）、「できる」あるいは「どちらかといえばできる」と答えた生徒の割合は、76.1%から 88.1%に増加した。特に「できる」と答えた生徒は4月から2倍以上に増え、全体の1／4を占めた。また、社会的な話題については、7月時点でわずかに落ち込んだが、最終的に約5%増えた。一方、社会的な話題について（図5）、「どちらかといえばできない」あるいは「できない」と感じる生徒の割合が、3割を下回ることにはなかった。

最後に、やり取りをする時「論理の構成や展開を工夫できるか」との問いに関して、「できる」の数はわずかに減少したものの、「どちらかといえばできる」と併せて見ると、ポジティブな意識を持つ生徒は10%弱増加している。なお、「どちらかといえばできない」と「できない」を合わせると10%減少した。（図6）

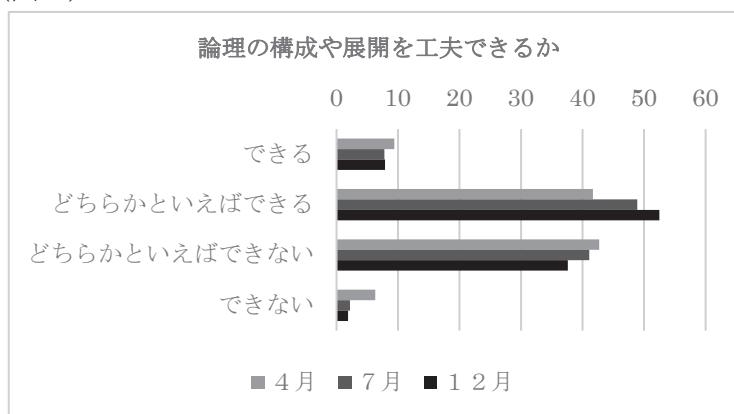


図6 論理の構成や展開の工夫に関するアンケート結果

この他にも、「英語の授業で苦手、あるいは上手いかわからないこと」を4技能5領域から複数回答可で継続してたずねたところ、「ペアで英会話を続けること（スピーキング「やり取り」）」だけが段階的に改善（13.3%→12.2%→9.9%）されていた。

9. 考察

9.1 意識アンケートから見る評価活動と活動意欲の関係

前述したように、生徒アンケートからは、仮説（１）で提示したポジティブな内容に近い結論となった。元々４月当初から、主要な質問項目に５～６割の生徒が一定の効力感を示していたが、特に「日常的なこと」について「できない」、「どちらかといえばできない」層の減少がそのまま「できる・どちらかといえばできる」層の増加に移行したことは、トリオワークでの身近な話題の相互評価活動が苦手意識の軽減に影響を与えたと考えてよいだろう。一方、課題となる社会的な話題の習熟については、日常に立脚した実感を伴う活動を心がけると同時に、これまで以上に「私」を離れて「社会」に向ける視点も養う必要性を感じた。

なお、自由回答欄には、毎回２～３件ずつ「文法の指導・解説」を求める声が見られた。本学年は、細かい文法については自走型の生徒を育てるべく参考書を利用した家庭学習で実施するよう指導・助言をしているとはいえ、学びの需要に応えるべく課題のフィードバックや質問の機会を充実させることも検討しなければならない。

9.2 関連しない生徒評価と教員評価

生徒評価には妥当性があり、教員評価と一定の相関があるという先行研究を見込みの高い土台としてスタートした本実践だが、実践と検証方法の独自性が強かったことで、仮説（２）で期待した相関とは大きく異なる結果となった。

まず問題にしたいのは、複数回の活動を平均化したとは言え、それぞれ異なる活動を対象として評価した点である。生徒のスピーキング能力はタイトルやテーマによって多少変動したとしても、複数回実施し、平均化すれば大差ないとする仮定は、元々の根拠が弱く、拙速な考えだった。本来であれば、同一内容の評価し、比較すべきであった。

次に、材料とした生徒評価、教員評価データともに、総合評価を用いたことである。当初から論理に特化した指導を展開してきたのだから、比較検証するループリック項目を **Logic** に絞ればまた違った結果になった可能性もある。

また、教員評価を基準にその妥当性を測るため、比較する教員評価の回数は数回あれば良しとし、５回分の生徒評価のデータに対し、２回分の教員評価のデータを平均化して比較したことから、比較検証するデータ数として適正さを損なっていたと感じている。

そして最後に、ループリックに関する理解と判断の習熟が十分で無かった可能性を考えたい。当初から教員間で共同して基準づくりを行った上で、生徒にはプリントと口頭で周知してきたループリックだったが、特に生徒間評価を集約する度、シビアさに欠ける評価を散見することが少なからずあった。川上（2021）で検証した通り、生徒は「相互評価は高く、自己評価は低く付ける傾向」がここでも見られた。評価基準に関する指導をもっと定期的にすべきだったと感じる一方、毎回評価基準が変わる **Logic** についてどのようにしたら生徒の基準理解が進むかは今後の難しい課題となった。ただ、成績評価としての妥当性が十分確認できなかったとしても、相互評価活動自体は学習活動を充実させる手立てとして、小泉(2018)が言う「学習支援の形成的な評価」活動としての価値は十分にあると考えている。

10. まとめと今後の課題

新指導要領に従って生徒の統合的な英語運用能力を育成するには、現行のアンバランスな指導領域とその評価の方法を検証し、説得力があり、誰でも実践可能なパッケージを設計する必要がある。

検定試験等を除けば、未だ受験に直結するイメージが無いスピーキング活動「やり取り」の位置付けが今後どのようなものになるにせよ、トリオワークに参加し続ける生徒の多くにとってやり取り活動が日常化しつつあり、彼らは活動を通して自分の中の表現欲求を見出し、さらなる成長の場面を求めている。また、活動のアウトラインをさらに明示すべく、Galaczi&Taylor(2018)が言及する「会話を主導する力」をはじめとした、現実的なコミュニケーションに根ざした力の習熟を目指すより実践的な指導方法を模索していく必要があるだろう。

パフォーマンス評価を進めることは「学習者の学習促進と教師の授業改善」が目的であると高井(2019)が示唆するように、生徒の主体的な英語学習意欲の向上に大きく資するだけでなく、指導者側にも原案作成から評価基準の設定まで日常的に協働しながら授業設計していくことを求めている。全てのプロセスで教員間の相互理解に基づいた協働が必須になることは、本実践を通して科全体で共有できた価値ある知見だと考える。

今後、継続して生徒評価の調査研究を行う場合は、評価基準の落とし込みに十分な時間を費やす必要があるだろう。また、比較検証する際は、回数と比較項目の精査を行った上でより精密な検証として実施し、研究方法の妥当性と信頼性を高められるよう、日々試行錯誤していきたい。

11. 参考文献

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. U.K.: Oxford University Press.

川上佳則. (2022). 「続・自由英作文指導実践-継続実践と過年度比較検証から-」 愛知教育大学附属高校紀要, 2022, 49, p. 133-141.

小泉利恵. (2018). 「自己評価・相互評価を取り入れたスピーキング指導」 数研出版, *CHART NETWORK*, 87 号, pp. 9-12.

深澤真. (2010). 「スピーチにおける生徒相互評価の妥当性」 全国英語教育学会紀要, 2010, 21 巻, p. 181-190.

Galaczi, E., & Tylor, L. (2018). International competence: Conceptualisations, operationalisation, and outstanding questions. *Language Assessment Quarterly*, 15, 219-236.

Kusuymayanthi, S. (2022). The effect of peer feedbacks on students' speech: *English Journal Literacy UTama*, vol.7, No.1, September, pp. 603-610.

高井一雄. et.al. (2019). 「ルーブリックを活用した授業実践とパフォーマンス評価-学習者の自信形成と教師の協働を目指して-」 富山大学人間発達科学部, 第 14 巻第 1 号:p. 63-71.

Nakamura, Y. (2002). Teacher Assessment and peer assessment in practice. *Educational Studies*, 44, p. 203-215.

Messick, Samuel. (1996). Validity and washback in language testing. *LANGUAGE TESTING*, vol.13, no. 3, p. 241-256