

生徒間評価活動の実践研究

—学びに向かう力につなげる評価者の育成—

英語科 川上 佳則

論理・表現 I の授業において、スピーキング（やり取り）領域にフォーカスした指導と評価を一体化する研究実践に取り組む。評価基準となるルーブリック指導を徹底した上で、やり取り活動の生徒間評価を指導の最前面におくことで、教員評価に十分相関する生徒間相互評価・自己評価の実現が可能か追求する。合わせて、評価自体を学び、実践し合うことで期待する評価者としての成長が、やり取り活動の取組そのものに与える影響を考察する。

生徒評価データと生徒意識アンケート結果の分析から、生徒内で評価基準が確立していく様子や、相互評価活動を通してやり取り活動が活性化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） 生徒間評価 自己評価・相互評価 評価者の育成 トリオ

1. 研究の背景及び目的

本校英語科は、「論理・表現 I」が始まった令和 4 年度から、スピーキング活動における「やり取り」にフォーカスした研究活動を始めた。他の技能・領域に比べて、その実践と評価の難しさを際立って感じられたことがその理由の一つである。指導と評価の一体化が求められる中で、授業においては、効率よく協働しながらより主体的に学び合える仕組み作りを、評価においては教師・生徒共に持続可能な方法を模索してきた。昨年度からの取組では、常態的なトリオワークの導入と、第三者の視点を育成しながら生徒間評価の意識と精度を高めていくことで、一定の成果も見られるようになってきている。

本実践を通して期待するのは、生徒が教師による評価だけに依らず、生徒間の評価活動でも自らを正当に評価できる力を身につけることである。評価基準の認識と判断の精度を高めるために、大学生が作成した動画視聴による基準確認をしながら、規範となる教員評価と、生徒自己評価と生徒相互評価を含めた生徒間評価（以降、生徒間評価）の相関を見てその変容と効果をみる。やり取り活動における相互評価活動を充実させることで、さらなる学びへと主体的に繋げるマインドセットを身につけ、論理的表現の育成と英語発信力の向上を目指すことが主たる目的である。

2. 研究対象

本実践の対象生徒は、高校 1 年生で論理・表現 I（2 単位）を受講する 1 学級 15 名（男子 3 名・女子 12 名）である。なお、昨年度から論理・表現 I は、30 名 1 クラスを 2 分割して実施している。当該クラス生徒の英語の学力は、下位者の数が少なく、中位層に厚い分布となっており、本学年の中でみても中程度と言ってよい。明るく元気の良い集団で、新しい取組に対する意欲・関心は比較的高いと感じる。

3. 指導実践期間

今回の指導を実践する期間は、2023年4月から2023年12月までの予定である。なお、本調査・実践は2022年度からの継続実践である。

4. 先行研究

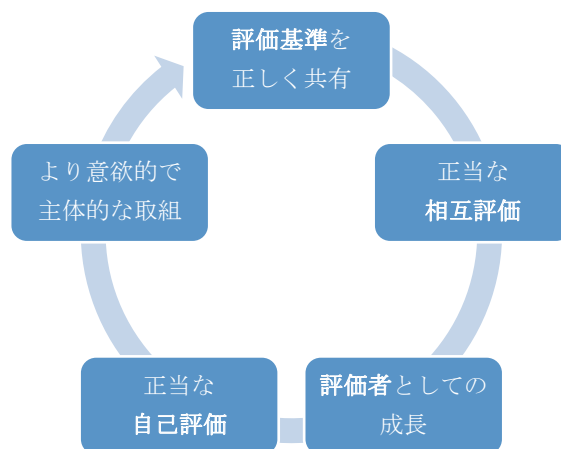
現在、国内にも多種多様なスピーキング指導と評価に関する研究実践（深澤, 2010, Nakamura, 2002, 高井. et.al, 2019）がある一方、スピーキングの「やり取り」活動、その指導と評価の一体化、またその妥当性に焦点を当てたものはそれほど多くない。Kusumayanthi(2022)は、教員の評価は必要とする一方、生徒の相互フィードバック後に、多くの生徒がスピーチのグレードを上げたと報告する。一方、評価そのものにフォーカスしすぎると、その妥当性をもつ性格は、信頼性を含む複雑なもの（Bachmann, 1996）であることから、例えば Messick(1996)が分類する多面的な内容全てについて検証するようなことは大変困難を伴った作業になり、本実践の本来の目的を見失うおそれがある。生徒間評価をそのまま成績に反映させることはできないが、評価活動自体に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」活動の価値が十分にあると考えている。

5. 仮説提示

昨年度の第1学年を対象とした同一の調査・実践は、教員評価と生徒間評価の相関を見るデータのバランスが不均衡であったことから、信頼に欠ける結果となった。今回は活動をビデオ録画し、教員評価と生徒間評価の直接的な相関と、指導による生徒意識の変容をみる。仮説提示は以下のようになる。

1. あらかじめルーブリック等で評価に関する基準確認の指導を行えば、生徒間評価は教員評価と正の相関をし、その後、成績に取り入れ得る信頼性を担保できる。
2. トリオワークを通して、他者を評価する評価者の視点を育成することで、生徒は自らの活動も正しく客観的に評価し、一層意欲的に「やり取り」活動に取り組むことができる。

<仮説における生徒に期待する変容のイメージ>



6. 実践方法

(1) 実践の手順

本研究では、研究対象のクラスの論理・表現 I の授業において、内容はすべてトリオワーク（3人組）で行う。また、各課の終わりに、テーマに沿ったやり取り活動（パフォーマンスチェック）を実施、そこで相互評価と自己評価、記述による活動の振り返りを行う。課毎で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。課末で生徒間活動を行い、最後に振り返り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。

Lesson 8

今回の Speaking (やりとり) Theme: 自分の意見を出す／聞き返す

🗣️ 以下のダイアログで B さんを評価します。B さんは①から③について表現してみよう。1ラウンド: 30秒

A: For the next school festival, what would you like to do together as a class?

B: I'm interested in ①興味があり、クラスでトライしてみたいこと

A: Bの内容を聞き返す? That sounds (fun/great/interesting/etc.)
Exactly what would you like to do?

B: ②クラスでトライしたいことの中で、自分が特にやってみたい活動・作業・パートを伝える文
③それをやりたい理由・楽しさを伝える文

A: Good for you!/That's a great idea!/Perfect!/

図1 パフォーマンスチェックシートの抜粋

(2) 指導上の工夫

1) 評価トレーニングのためのトリオワーク

前述したように本実践における生徒の活動はすべてトリオワーク（3人／グループ）で行う。最初のスモールトークから教科書内容理解の相談まで、活動の基本単位をトリオワークとし、個人活動から協働して学ぶ場へシームレスで移行できるようにしている。どの活動もトリオの中でペア活動をする間、3人目は第3の視点でメモをとったり、評価したり、あるいは観察して得た情報を再構築して他者に伝えたりと、工夫次第でペア活動よりも活動負荷を高めることもできる。

2) 生徒間の互惠関係を避けるローテーション

昨年度の実践（川上, 2022）と生徒アンケートから、生徒は同一グループ内の他者評価には一定程度「手心を加えてしまう」ことが分かってきた。グループ内の決まったメンバーで評価し合う関係性が、いわゆる付度に近い状況をつくり、それが評価そのものの精度を落とす要因になると推測できる。今年度実施した生徒の初期アンケートからも、やはり同様の傾向が見受けられる。（図2）

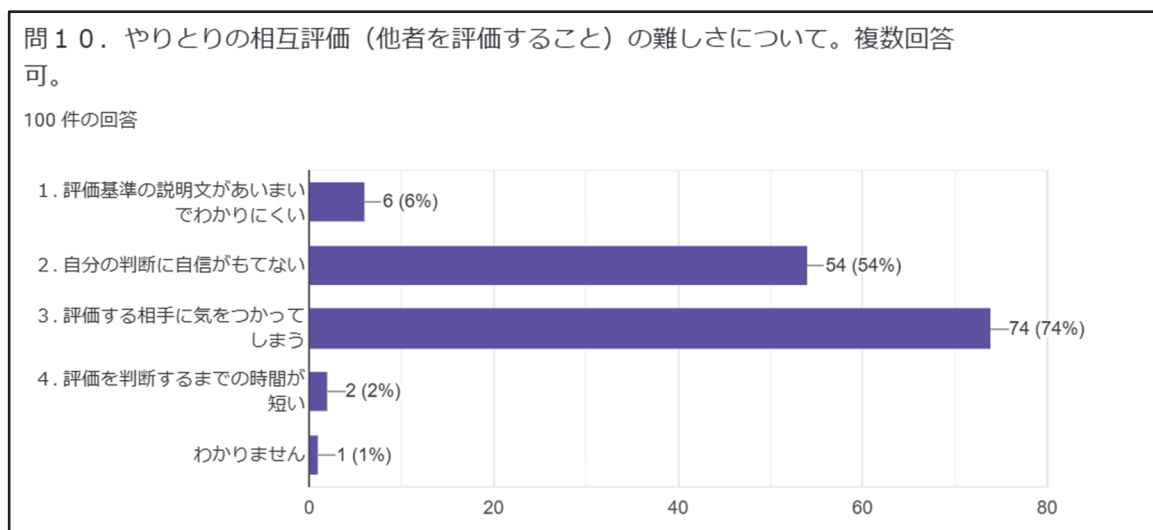


図2 生徒意識アンケート（6月）より抜粋

そこで、あらかじめ生徒を A、B、C に配し、トリオワークを一回実施する毎に、A が半時計回り、B が時計回りでグループ間を移動するローテーションを導入した。(表 1) C のみが固定された状態で A と B が毎回入れ替わることで、評価し合う状況を防ぐことができる。ローテーションを導入した当初は一部で困惑した様子も見られたが、現在は回数を経て十分適応している。また、毎回 ABC の担当を変えることで、対話する相手やローテーションのパターンが変化する新鮮さも演出できる。

表 1 5グループ・3ラウンド実施のローテーションパターン *A1: グループ1のAさん

Round 1			Round 2			Round 3		
評価される	相手をする	評価する	評価される	相手をする	評価する	評価される	相手をする	評価する
A1	B1	C1	B5	C1	A2	C1	A3	B4
A2	B2	C2	B1	C2	A3	C2	A4	B5
A3	B3	C3	B2	C3	A4	C3	A5	B1
A4	B4	C4	B3	C4	A5	C4	A1	B2
A5	B5	C5	B4	C5	A1	C5	A2	B3

3) 明瞭な達成基準を用いたループリック

設定した評価の観点は、Volume、Eye contact と Logic の 3 つである。「excellent: 3/good: 2/poor: 1」の 3 段階 (決めがたい場合は 2.5 や 1.5 も可としたことで実質 5 段階評価) で、文章化された各項目について評価する。(図 3) 今年度は、愛知教育大学の学生の協力により、評価基準別にクリップ動画を準備することができた。(図 4) これにより、判断が主観によって曖昧になりがちな Volume と Eye contact の基準を、クラス全員で客観的に共有できる。各評価段階に相当する内容に合わせて作成した複数のクリップから、教師があらかじめ選別したものを、教室で視聴した。さらに、比較可能なクリップ (例: Volume の評価 3 の例を複数) を作成してくれたことで、一つのモデルを見せてから他のクリップを見せ、その場で評価基準の理解度をテストすることもできた。

今回の評価基準		
①Volume		
Excellent	3 points	:一貫して力強く、とても明瞭に聞こえる
Good	2 points	:聞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		
Excellent	3 points	:プリントを見ずに助言①ができ、かつ助言②は3文以上できる
Good	2 points	:プリントを見ずに助言①ができ、かつ助言②は1~2文できる
Poor	1 points	:プリントを見て助言①と助言②はできるが、助言②は1文以下である

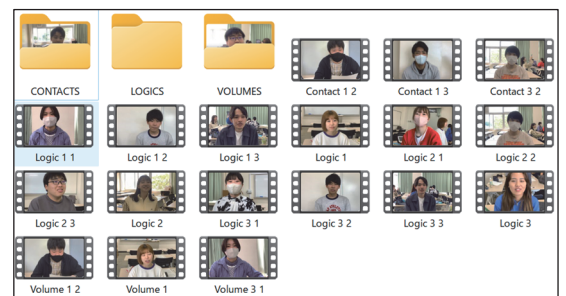


図 3 評価基準の例

図 4 大学生が作成した動画

また、Logic についても、「～について 2 文」や、「～の説明が 3 文以上」等、なるべく曖昧さを排除した形で提示することで、教師、生徒ともに判断しやすい形にしている。

7. データ収集と分析の方法

(1) パフォーマンスチェックシート

6 の (1) で示したように、生徒はパフォーマンスチェックシートを用いて生徒間評価、振り返りコメントを記録する。パフォーマンスチェックは 2 課に 1 度の頻度で実施しており、12 月時点で 8 課分のデータが分析対象になる予定である。録画したやり取りを後日教師が評価した教員評価、生徒による自己

評価、そして相互評価をそれぞれの相関関係から分析する。

(2) 生徒意識アンケート

6月と11月に同一の内容の意識アンケートを行う。アンケートは、CAN-DO リスト、パフォーマンスチェック、やり取りの評価活動の3点について調査する。質問の概要は以下の通りである。

- 問1～問4 : CAN-DO リストを活用した自己評価活動に関する質問
問5、問6 : パフォーマンスチェック（トリオワーク）に関する質問
問7～問12 : やり取り活動に関する質問
問13 : やり取り活動に関する感想や意見を自由記述式に回答

8. 結果

(1) 教員評価と生徒間評価の相関

該当クラスは合計8回のパフォーマンスチェックを行い、全てにおいて生徒は自己評価・相互評価を実施した。そのうち、抽出した生徒（延べ48名分）について教員が評価したデータと比較した。教員評価と相互評価、教員評価と自己評価、そして、相互評価と自己評価の3つの観点から相関関係を調査した。（表2）

表2 各相関の状況

	教員評価—相互評価 (n=48)	教員評価—自己評価 (n=48)	相互評価—自己評価 (n=48)
Volume	r=0.34	r=0.33	r=0.67
Eye contact	r=0.29	r=0.24	r=0.64
Logic	r=0.51	r=0.47	r=0.72

Volume と Eye Contact に関する教員評価と生徒間評価の相関は、いずれも0.3前後で、相関は非常に弱い結果となった。調査の中間発表時には0.5程度あった数値が、後半の活動で散らばっていったことになる。Logic については0.5前後とやや相関が見られる。なお、わずかではあるが、すべての項目について自己評価より相互評価の方が、教員評価に沿った数値になっている。一方、生徒間評価内の比較（生徒相互評価と生徒自己評価）を見ると、いずれの項目についても、相関があり、Logic については強い相関を示す結果となった。

(2) 生徒意識アンケート回答の推移

まず、CAN-DO リストを用いた自己評価活動に関して、「自信を持って回答している」、「どちらかといえば自信を持って回答している」と答えた割合は85%から91%へと増加した。また、CAN-DO リストを事前に見ることで「やる気になる」、「どちらかといえばやる気になる」割合も72%から76%と高い水準で推移している。続いて、パフォーマンスチェックにおけるトリオワークについて、「良い」、「どちらかといえば良い」とする割合は、2回とも97%と高止まりしており、生徒に強く支持された活動であることがわかる。一方、他者を評価すること（相互評価）について、「自信を持ってできる」、「どちらかといえば自信を持ってできる」と回答した割合は67%で横ばいであった。なお、その理由の内訳を見ると、「ループリックがあるから」と答えた割合は10%減少し、「何回か経験した」ことを理由に挙げた割合は6%増加している。

今回のアンケート結果を比較する中で注目したのは、やり取りの自己評価・相互評価の難しさに関する項目である。やり取りにおける「相互評価」で難しいのは、「相手に気を遣ってしまう」からと答えた割合は、11%減少した。やり取りにおける「自己評価」で難しいのは、「自信が持てず、自己評価を低くしてしまう」からと答えた数は80名から64名と減少した一方で、「自信が持てず、自己評価を高くしてしまう」と答えた数は18名から26名と増加した。

最後の自由記述欄は、6月、11月のいずれも10件程度の回答があった。総じて自分の課題や、本活動の意義や大切さをよく理解したポジティブな意見であった。(図5)

やり取り活動が多く、英語の、対話の練習がたくさんできるのが良いと思いました。また、やりとりが多いので楽しく授業を受けている気がします。やりとり活動では実際に話した時どう返すかを意識して取り組んでいます。

自分で作った英文や、本文でやりとりすることは、学校でしかできないことから、英語で会話する力が伸びると思うので、大切だと思います。私はやりとりの活動で、英語で会話する力を伸ばしたいです。

やりとり活動では、相手と対話する事で相手の事を知れて、更に自分の英語の改善をする事が出来るので、素晴らしい活動だと思う。これからも他者とやりとりを重ねて楽しく英語を学んでいきたい。

図5 生徒自由記述欄からの抜粋

9. 考察

今回、教員評価と生徒間評価を比較検証した結果、Logic以外の相関は総じて極めて弱いことが分かった。ルーブリック指導に加えて、動画を用いて評価観点に関する相互理解を深められるよう指導してきたことが、上手く結果に繋がっていない。今回のデータが、仮説1として期待した生徒評価と教員評価の「正の相関」を示したとは言い難い。ただ、教科書が進み、やり取りで求められる内容がより複雑になるにつれて、評価が難しくなっていったLogicについて、一定の相関(0.5前後)が見られたことは、ルーブリック指導の一定の効果があったと考えて良いのではないかと。また、VolumeやEye Contactなど、どちらかと言えば理解し易い評価基準であるにも関わらず、教員評価とかけ離れてしまう理由については、あるいは、録画による音量や映像の角度に問題があり、画面越しに教員が受ける印象が本来の活動と異なるものであった可能性もある。実際、生徒の自己評価と相互評価の相関は高く(0.7前後)、彼らの中には一貫した基準が身につけていると考えられる。仮説2で求めたように、少なくとも生徒のレベルにおいては、「他者を評価する評価者の視点を育成」しながら、自己評価もある程度適正にできるようになってきていると見て良いだろう。実際、8の2)の意識アンケートの結果から、本実践期間中、活動に関する意欲が極めて高い水準で維持されていることが明らかになっている。引き続きの課題となった教員評価との相関については、より正確に評価が反映されるよう、撮影したものと実際の活動にどの程度格差があるのか検証する必要がある。また、取りも直さず教員である自分自身が、自分の生徒評価が適正であるかどうか、過信せず、適宜同僚の教員とすり合わせをするなどの不断の努力が求められると強く感じた。

仮説についてまとめると、仮説1は不十分であったが、仮説2について期待された変容が一定程度見られたと言える。また、評価の「正しさや正当性」については課題を残しつつも、変容のサイクルは概ねイメージ通り機能しつつあると考えて良いだろう。アンケート結果や授業参加の観察から、サイクルのゴールである「より意欲的で主体的な取組」は、目に見える形で確認できるようになった。

当初から繰り返すが、評価者の育成にこだわりすぎた結果、本来高めたいやり取り活動への意欲を削ぐようなことになっては本末転倒になる。今後も絶妙なかじ取りが求められる中、形成的評価にとどまらず、さらに信頼性を高めることで、総括的評価に繋げられる活動まで追究していくのか、それとも、

評価自体の信頼性まではこだわらず、評価活動を通してやり取り活動の意欲的な状態を維持することを目的にして、現行のフレームワークで続けていくのか、本実践の目的を再考、再検討する必要がある。

10. まとめと今後の展望

2023年12月現在、アンケートの結果が示すように、生徒の活動への取り組み状況は良好である。考查等の学習成果からも、授業への取組が望ましい形で成績に反映されているように感じる。一方で、どうしても活動になじめない一部生徒への個別対応も必要になってくる。「相互評価システム」を通して、積極的に授業に参加する雰囲気を作りだしていけば、生徒の「自発的なスキル向上を促すことが可能である」と河野（2014）が言うように、この場合も、教師がすぐに介入するのではなく、トリオワークでフォーカスする「協働する力」が生徒間で上手く機能することを期待しつつ、バランス感覚を持った声かけや支援を心掛けていく。また、活動の軸となる評価基準づくりにおいては、特にLogicの観点づくりに加えて、改めて教員による「より適正な評価方法と基準確認」を、英語科全体で共有していきたい。

およそ2年をかけてイメージが付きにくいスピーキング（やり取り）の評価活動を日常的に、しかも生徒主体で実践できる仕組みづくりに取り組んできた。評価そのものの指導と理解に関しては課題を残したが、評価者としての責任と自覚を促しながら、課毎のテーマに求めるポイントを組み込んだルーブリックについて丁寧な説明を行い、生徒のやり取り活動を活性化することができた実感は確かにある。現在、統合的な英語力に求められる領域は多岐にわたる。スピーキング（やり取り）活動を通して生徒間に創出されるのは、双方向性が大前提になる極めて実際の対話活動だ。実践的なツールとしてのこれからの英語（あるいは英語学習）を考えたとき、とりわけやり取り活動で育まれる「対話を積み上げるリズム感覚」や「瞬発的な感情の発露を表現する実感」が、子どもたちのより深いところから立ち上がる主体的な学びに向かおうとする力を引き出すきっかけになるはずだ。今後も実践し易く、十分な効果を見込めるやり取り活動の充実を目指し、継続して「論理・表現Ⅰ・Ⅱ」の授業実践研究に取り組んでいきたい。

11. 参考文献

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

川上佳則.(2022).「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—」愛知教育大学附属高校紀要, 2023, 50, p.23-30.

河野昭彦 et. al (2014). 「学生の相互評価によるアクティブラーニング型授業」工学教育(J. of JSEE), 62巻6号, p. 6_62-6_67

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