英語スピーキング・パフォーマンスにおけるピア評価の匿名化

岡田　靖子

Abstract:

Anonymous Peer Assessment: Speaking Performances in English as a Foreign Language

In a peer assessment process, students evaluate the performance of other students or peers based on the norms set by teachers. While peer assessment encourages the active involvement of students in collaborative learning, questions have been raised regarding its reliability and validity. To increase the reliability of peer assessment, students can judge their peers' performance anonymously. This empirical study investigated the perception of Japanese students on the importance of anonymity of the rater in peer assessment. The assessment was conducted on speaking performances in the subject of English as a Foreign Language (EFL). The participants were 25 Japanese first year university students who had been enrolled in an English writing course. The students were divided into two groups. They read out aloud from a prepared script and assessed their own and peers' performances in the classroom. Qualitative and quantitative analyses were conducted on the ratings of peer assessment and responses to open-ended questions on judging peers with and without the rater's name. The researcher discusses the pedagogical implications for peer evaluation, focusing on the importance of preserving anonymity while giving and receiving feedback from students.

要 旨：

ピア評価は、教師が作成した評価基準を用いて学習者同士でパフォーマンスを評価する、あるいはフィードバックを与える学習プロセスである。ピア評価を用いると、協働学習に対する学習者の積極的なかかわりを促進させることにつながるが、信頼性や妥当性についての問題も指摘されている。ピア評価の信頼性を高める手段として、学習者が匿名で他の学習者のパフォーマンスを評価する方法があげられる。本研究では、英語学習者のスピーキング・パフォーマンスを匿名または記名によるピア評価を実施し、その効果の相違を検証した。研究対象者は英語ライティングのクラスを履修している日本人の大学生 1 年生 25 名である。このクラスの学生を 2 グループに分けて事前に準備した原稿を発表させ、発表のビデオ映像を観ながら自己評価とピア評価を実施した。自己評価とピア評価の得点を量的分析し、匿名と記名によるピア評価に関する自由記述質問紙の回答は質的分析を行った。評価者の匿名性を保持しながら学習者間でフィードバックを交換することの重要性に着目し、ピア評価の教育的示唆を検討する。

キーワード：

ピア評価　匿名　スピーキング　英語学習　量的・質的分析

1. はじめに

ピア評価とは、学習者間で相手のパフォーマンスに対して改善の参考になるようなフィードバックを与える相互評価プロセスのことである。ピア評価は学習者の協働学習を支援する教育方法の一つであり（Orsmond, Merry & Reiling, 1996）、外国語教育の現場でもコミュニケーション活動の一環としてスピーキングやライティングの能力を向上させるために取り入れている。英語学習者のオーラルパフォーマンスにピア評価を用いた研究（Fukazawa, 2010; Okada, Sawaumi, & Ito, 2014; Peng, 2009; 岡田・いとう, 2014）が行われているが、その評価の信頼性を高めるために、評価者の匿名性を検証した研究（e.g., Hirai, Ito & O'ki, 2011）はあまり実施されていない。ソーシャルメディアでも匿名による投稿者の意見について責任の所在が問題視される昨今、社会においても個人を特定化することの是非が問われている。

そこで本研究では、外国語の教育現場で学習者が匿名と記名の両方でピア評価をすることによって、評価者と被評価者の立場を経験する学習者の

意識にどのように影響するのかについて検証する。両方の立場に置かれた学習者の評価者に対する意識を理解することによって、ピア評価を活用しながら効率的に学習者の言語能力やスピーキングスキルを育成できると考えられる。

2. 先行研究

2.1 スピーキング能力の評価

スピーキングは、評価することが難しい能力として指摘されている。その理由として、多くの要因が聞き手の印象に影響を与えること（Luoma, 2004）や、日本の場合、英語授業が大人数で実施されていること（Fukazawa, 2010）などが挙げられる。そこで、スピーキング能力を評価しなければならない教師の負担を軽減するために、学習者同士によるピア評価を導入する取り組みが行われている（笠巻, 2016；Fukazawa, 2010）。このピア評価を効率的に実施するための方法として、ピア評価のための訓練の重要性（Patri, 2002; Peng, 2009; Saito, 2008）が指摘されている。Saito（2008）は、教師と学習者による評価の間に相関があること、ピア評価の訓練はメタ認知活動、つまり、学習者のスキル面に対する注意を高めることにつながることを利点として挙げている。また、Peng（2009）は、学習者のパフォーマンスについて、教師は「成果」（product）として評価するが、学習者は「プロセス」（process）として評価するという違いを報告している。

一方、ピア評価の欠点として、評価が教師より緩くなり、極端な評価を避ける傾向がみられる（笠巻, 2016）ことから、ピア評価の信頼性は低いと指摘されている。同様に、Hirai et al.（2011）は、大学生を対象として、学習者のオーラルパフォーマンスに対する教師と学習者の評価の違いを比較した結果、文法と語彙の項目で学習者より教師のほうが厳しく評価することを示した。ピア評価では、友人のパフォーマンスに対して低い評価を与えにくいというフレンドシップ効果（friendship effect）が生じる（Falchikov, 2005; Peng, 2009）ことを踏まえると、学習者が置かれている学習環境が評価の得点に反映される可能性があることを考慮すべきだろう。

2.2 学習者間評価の匿名化

学習教育では、オンラインを用いたピア評価の匿名化の研究（庄・長登・稲垣・隅谷・中村, 2015 ; 庄・三浦, 2012）が実施されている。庄・三浦（2012）はオンライン環境におけるレポートのピア評価について検証したところ、学習者の人間関係や効力感[1]と学習効果の関係は示されず、匿名による効果は認められなかった。また、ピア評価を匿名群と記名群に分けて実施した研究（庄ほか, 2015）では、ピア評価はいずれのグループにもよい方向に影響している一方で、ピア評価を実施するグループによって匿名を好む場合と、記名を好む場合がある可能性を示唆した。さらに、英語学習者のオーラルパフォーマンスにおけるピア評価を検証した研究（Hirai et al., 2011）では、個人またはペアでピア評価を匿名で実施した 2 グループを比較した。その結果、個人で匿名のピア評価を行ったほうが、教師による評価との高い相関関係が認められ、ピア評価の信頼性が高まることが示された。また、学習者が評価する側とされる側になることで、共有責任（shared responsibility）の増大が報告される一方で、学習者の環境や情況が評価内容に与える影響が大きいことを踏まえると（Hirai et al., 2011）、匿名によるピア評価の信頼性についてさらなる検討が必要とされる。

3.　研究目的

本研究の目的は、学生のスピーキング・パフォーマンスのピア評価における匿名と記名の効果の相違を検討することである。

4.　調査方法

4.1 研究対象者

東京都内の私立大学において経済学を専攻する 1 年生 1 クラスが、2016 年度後期に必修科目である英語ライティングの授業の一環として研究に参加した。履修登録者は 38 名であったが、研究データの使用に関する同意書を求めたところ、7 名[2]が研究参加を希望しなかった。また、長期欠席者な

1 自己効力感（自分が目標を達成するための行動を遂行できるかどうか）や集団効力感（グループで目標を達成するために行動を遂行できるかどうか）が含まれる。

2 研究参加に同意しなかった学生からも他の学生と同様、評価シートや自由記述質問紙を回収したが、データ分析には使用しなかった。

どを除外した結果、最終的に研究対象者とされた学生は 25 名（男性 20 名、女性 5 名）であった。入学と同時に TOEIC Bridge テストを受験し、このスコアによってクラスが編成されていた。研究対象者の平均は、149.6 点であった。クラスの全体的な印象としては、遅刻する学生が複数いたので、クラス全体のモチベーションを上げるのが難しいときもあったが、基礎的な英語力は身に付けているので、授業運営に関しては特に問題はなかったことがあげられる。

4.2 授業形態

この授業は、90 分の授業が週 2 回、全 28 回実施された。英語によるライティングスキルの向上を目指していたため、学生自身のことやインタビューしたクラスメートのことをまとめるタスクが多く含まれていた。アウトプットについては、スピーキングよりライティングが中心であった。この授業は筆者が担当していた。授業ではスピーチを 2 回実施し、ピア評価における評価者の有無による効果を比較した。1 回の授業で全ての学生のスピーチを実施することは時間的な制約があるため、学生を名簿順に 2 つのグループ（A グループは 13 名、B グループは 12 名）に分け、授業時間内にスピーチやピア評価が終了するように配慮した。TOEIC Bridge テストの平均点は、それぞれ 149.5 点と 149.7 点であった。

4.3 指導手順

スピーチのトピックは、1 回目は「パートナーの特徴」、2 回目は「パートナーの夢」についてであった。180－200 単語で原稿を作成させ、初稿は 3－4 人のグループでピアレビューを実施し、修正させた。再校、再々校については教師が内容や言語についてのコメントを与え、書き直しをさせた。

まず 1 回目のスピーチは、A グループが初めに発表し、B グループは聞くことに集中した。発表の様子は、三脚にマウントされたビデオカメラで全て撮影された。学生はビデオ映像を観ながらピア評価（記名）・自己評価を行った。評価終了後、ピア評価シートは発表者に渡され、発表者は評価の得点やコメントを確認した。

次に、B グループが 1 回目のスピーチを発表し、A グループは聞くことに

専念した。A グループと同様にビデオ映像を使ってピア評価（匿名）・自己評価が実施された。評価後、教師がピア評価シートを全て回収し、発表者に配布し、評価得点やコメントを確認させた。

2 回目のスピーチも 1 回目と同様、まず A グループが発表し、B グループは聞き手となった。ビデオ映像を用いたピア評価（匿名）・自己評価が行われた。授業時間に余裕がなかったため、評価が終了したのちに教師が評価シートを回収した。

最後に、B グループが発表し、A グループは聞くことに集中した。ビデオ映像を観ながらピア評価（記名）・自己評価が実施された。両グループの評価シートが発表者に配布され、各学生は評価の得点やコメントを確認した。図 1 にスピーチの発表から評価までの指導手順を示した。

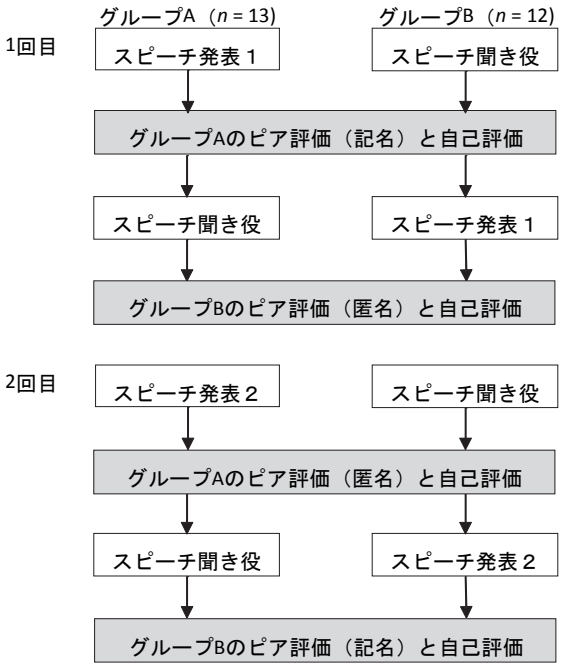

図 1　指導手順の概要

4.4 調査材料

本研究では、ピア評価における匿名の影響を検討するために、自己評価

とピア評価の評価得点を算出した。教育現場における自己評価は心理的に働く自己防衛によってその効果は限定的であるが、ピア評価は他者を評価することを学び、またピア評価自体が自己へのフィードバックにつながることが指摘されている（藤原・大西・加藤, 2007）。したがって、教育現場においてピア評価の有用性を検証することは、学習意欲の向上につながると考えられる。

4.4.1 評価シート

評価項目は、先行研究（Yamashiro & Johnson, 1997）で使われた評価表を参考に作成された。本研究では、（1）声の大きさ、（2）話すスピード、（3）イントネーション、（4）話し方、（5）姿勢、（6）手足の位置、（7）アイコンタクト、（8）顔の表情、（9）トピック、（10）言語使用、（11）語彙の 11 項目[3] について評価を求めた。4 件法（1 : あまり良くない、4 : とても良い）で得点を記入させた。さらに、全体的なコメントも自由に記入できるよう欄を設けた。また、評価の参考となるよう、各項目の評価ポイント（例: 姿勢の場合、「背筋はまっすぐ伸ばす」）を示した。

4.4.2 自由記述質問紙

すべてのスピーチとピア・自己評価が終った際に、評価者の匿名性について質問紙を用いて回答させた。設問は以下の２つである。

- 自分が他の学生を評価する際に、評価者（自分）の名前を記入せずに評価を与えることについてどのように感じますか。
- 他の学生に自分が評価をされる際に、評価者の名前が記入されずに評価を受けることについてどのように感じますか。

各設問では、その理由についても記入してもらった。さらに、「ピア評価に

[3] 先行研究（Yamashiro & Johnson, 1997）では、評価項目に「序論」「本論」「結論」などの内容に関する項目を含んだ 14 項目を評価対象としていたが、本研究では学習者が原稿を作成している段階で、内容に関するフィードバックを教師がすでに与えていることから、この 3 項目は概ね達成しているとみなし、ピア評価には含めなかった。

おいて評価者が記入されたほうがよい」について、評価する立場と評価される立場の両方から 5 尺度（１：そう思わない、５：そう思う）で回答してもらった。

4.5 研究デザイン

本研究は授業の一環として実施され、研究参加者の無作為配置が行われていないことから、準実験デザインを取り入れた。匿名・記名で実施するピア評価の時期を両グループでずらして交代制にし、匿名・記名の両方でピア評価を経験できるようにすることで、倫理上の問題に配慮した。

4.6 分析方法

本研究では，自己評価とピア評価の得点は IBM SPSS Statistics 23 を用いて量的分析を実施した。自己評価とピア評価の結果を分析するにあたり，評価シートで使用された 11 項目から「音声コントロール」(項目 1〜4)，「ボディランゲージ」（項目 5〜8)，「有効性」（項目 9〜11）という 3 つの下位尺度を作成した。合成変数を作成するにあたって、内的信頼性を確認するためにクロンバックのアルファ係数を算出したところ、1 回目は .85、.84、.95 (自己評価)、.78、.77、.94 (ピア評価)、2 回目は .86、.84、.98 (自己評価)、.75、.66、.97 (ピア評価)であり、概ね十分な信頼性が示された。量的分析の結果をさらに検討するために、自由記述質問紙の回答を質的に分析した。

5.　結果

5.1 自己評価とピア評価の分析結果

自己評価の得点の平均値と標準偏差は、表 1 に示した通りである。時期（1 回目と 2 回目）を参加者内要因、グループを参加者間要因とした混合モデルの分散分析を行った。

表 1　自己評価の平均値と標準偏差

	1 回目		2 回目	
	M	*SD*	*M*	*SD*
グループ A（*n* = 12）				
ボイスコントロール	2.77	0.74	2.92	0.69
ボディランゲージ	2.71	0.86	3.27	0.67
有効性	3.39	0.69	3.31	0.97
グループ B（*n* = 10）				
ボイスコントロール	2.63	0.82	2.98	0.70
ボディランゲージ	2.73	0.80	2.75	0.66
有効性	2.90	0.75	3.13	0.82

まず、自己評価のボイスコントロールでは、時期の主効果が有意であったが[$F(1, 20) = 4.60, p = .04, \eta p^2 = .19$]、グループの主効果は有意ではなかった[$F(1, 20) = .02, p = .88, \eta p^2 = .001$]。また、それらの交互作用は有意ではなかった[$F(1, 20) = 0.78, p = .39, \eta p^2 = .04$]。

自己評価のボディランゲージでは、グループの主効果は有意ではなかった[$F(1, 20) = 0.69, p = .42, \eta p^2 = .03$]。時期の主効果[$F(1, 20) = 7.42, p = .01, \eta p^2 = .27$]および時期とグループの交互作用[$F(1, 20) = 6.21, p = .02, \eta p^2 = .24$]は有意であった。単純主効果の検定を実施した結果、自己評価のボディランゲージにおいて、A グループは 1 回目より 2 回目を有意に高く評価していることが示された[$p = .001$]（図 2 参照）。

有効性では時期の主効果[$F(1, 20) = 0.31, p = .58, \eta p^2 = .02$]、グループの主効果[$F(1, 20) = 1.04, p = .32, \eta p^2 = .05$]、時期とグループの交互作用[$F(1, 20) = 1.40, p = .25, \eta p^2 = .07$]ともに有意ではなかった。

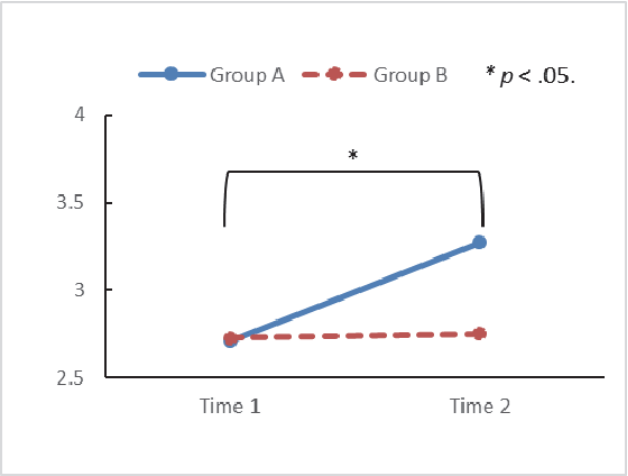

図 2　ボディランゲージの自己評価

次に、ピア評価の平均値と標準偏差を表 2 に示した。ピア評価のボイスコントロールでは、時期の主効果[$F(1, 21) = 3.77, p = .07, \eta p^2 = .15$]は有意でなかったが、グループの主効果[$F(1, 21) = 4.55, p = .05, \eta p^2 = .18$]が有意であり、B グループは A グループよりも評定値が高かった。それらの交互作用効果[$F(1, 21) = 1.49, p = .24, \eta p^2 = .07$]は有意ではなかった。

表 2　ピア評価の平均値と標準偏差

	1 回目		2 回目	
	M	*SD*	*M*	*SD*
グループ A（*n* = 12）				
ボイスコントロール	3.40	0.28	3.52	0.18
ボディランゲージ	3.43	0.29	3.63	0.17
有効性	3.73	0.12	3.76	0.27
グループ B（*n* = 11）				
ボイスコントロール	3.61	0.18	3.64	0.13
ボディランゲージ	3.64	0.22	3.62	0.11
有効性	3.90	0.05	3.73	0.09

ボディランゲージについては、時期の主効果[$F(1, 21) = 5.71, p = .03, \eta p^2 = .21$]は有意であったが、グループの主効果[$F(1, 21) = 1.61, p = .22, \eta p^2 = .07$]は有意ではなかった。それらの交互作用効果[$F(1, 21) = 8.44, p = .01, \eta p^2 = .29$]が有意であったため、単純主効果の検定を行った。その結果、A グループはピア評価のボディランゲージにおいて、1 回目より 2 回目のほうが有意に高いことが示された[$p = .001$]（図 3 参照）。つまり、ピア評価のボディランゲージは、評価者を記入して実施した 1 回目のピア評価より匿名で実施した 2 回目のピア評価方法が A グループで有意に高いことが示された。

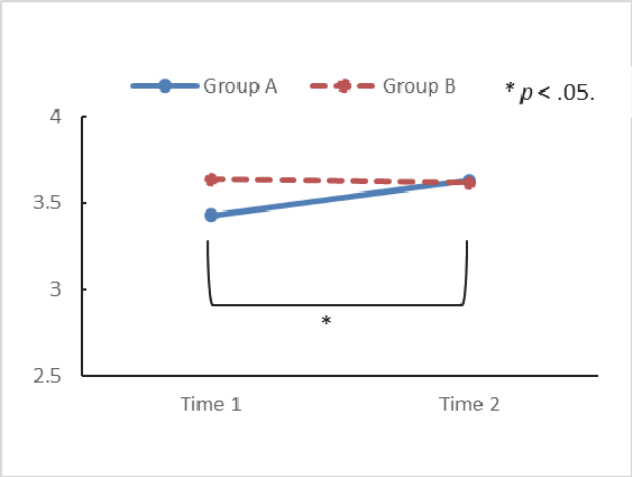

図 3　ボディランゲージのピア評価

有効性については、時期の主効果[$F(1, 21) = 2.72, p = .11, \eta p^2 = .11$]およびグループの主効果[$F(1, 21) = 1.63, p = .22, \eta p^2 = .07$]は有意ではなかったが、時期とグループの交互作用効果が有意であった[$F(1, 21) = 6.30, p = .02, \eta p^2 = .23$]。そこで、単純主効果の検定を行った結果、B グループの有効性では、2 回目は 1 回目よりも有意にピア評価が低いことが示された[$p = .009$]（図 4 参照）。つまり、ピア評価の有効性では、1 回目の匿名のピア評価より 2 回目の記名で実施したピア評価のほうが B グループで有意に低いことが示された。

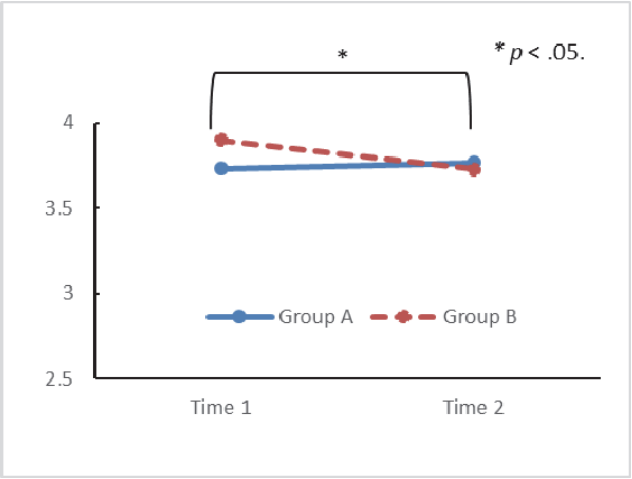

図 4　有効性のピア評価

5.2 自由記述質問紙回答の分析結果

まず、グループ間で「自分が評価者であるとき、名前を書いた方がよい」（評価者）と「自分が評価されるとき、評価者の名前が書かれてある方がよい」（被評価者）の得点を検討した結果を表 3 に示した。A グループでは評価者より被評価者の評定値のほうが高いが、この差は有意ではなかった[t =2.14, df =11, p = .06]。

表 3　グループ別の平均値と標準偏差および t 検定の結果

	グループ A（n = 12）		グループ B（n = 11）		
	M	*SD*	*M*	*SD*	*t* 値
評価者	3.08	1.56	3.09	1.30	3.09
被評価者	3.83	1.16	3.09	1.30	3.48

次に、自由記述質問紙の回答を検討したところ、学生の回答は 3 種類（匿名が良い・記名が良い・どちらでもよい）に分類された。

学生が評価する立場にいる場合、匿名で評価したい理由について、「素直に思ったことが書きやすい」「相手にどう思われるかなどきにせずに書けていい」「自分の意見を反映しやすいのは無記名のほうでした」とあるように、

評価者を記入しないほうがコメントを書きやすいことを指摘している。また、「名前を書くと評価は甘くつけてしまうと思う。名前を書かないほうがより厳しく評価できると思う」という意見から、緩い評価を与えにくい環境を作ることが可能になると考えられる。

一方、学生が記名で評価したい理由として、匿名だと「責任感が薄れる。実際の発表を真剣に見て相手の利益になるような意見を書くことを怠ってしまう」とあるように、評価者の記入はコメントの責任の所在を明確に示す役割を果たすと考えられる。

どちらでも良いと回答した学生は、「言葉の使い方を誤らなければよいと思う…正直、どちらでも良い」「酷評しやすい、素直な意見が書ける反面、適当に書く人もいると思う。なぜならば、誰が書いたか分からないから」と記述しているように、「誰が評価するのか」ではなく「どのように評価するのか」のほうが重要ではないかと考えられる。

次に、学習者が被評価者である場合、匿名で評価された方が良い理由として、「誰がどのように思っているのかがビシビシ伝わってくる」「評価者の思ったことや意見などが無記名の方が正確に反映されているのではないか」と挙げており、匿名で実施した評価のほうを高く信頼する傾向があることも示された。さらに、「名前がわからなければ、少しきついことを言われても、誰かのこと嫌になったりしないし、問題ないと思う」「特定の誰かからの評価だと思わずに気楽に見ることができる」とあるように、評価の内容を個人的に捉えてしまい、結果的に人間関係にも影響する可能性が示された。

一方で、記名で評価された方が良い理由として、「誰が書いたのかが気になり、夜眠れなくなる人がいるのかなと思った」「誰が書いてくれたのか少し気になるので、名前があってほしい」とあるように、評価の内容によっては評価者を特定することができるようにしたほうが良い場合もある。「匿名でいい加減な評価をされると嫌悪感を抱くので、名前が書いてあった方がいいと感じた」「発表内容と無関係の意見が見られたので、相手のことを考えてコメントするという姿勢が薄れているように感じた。ネットでの誹謗中傷はこの感覚に近い」などと述べられていることから、評価者を特定したほうが、評価の信頼性が高まると考えられる。

最後に、「誰がどのように思ってくれていたのかを知りたいので、正直、名前はあった方がよいが、書く人のことを考えると、名前ありだとあまり厳しい評価がつけられない」という記述に見られるように、学習者が評価者・被評価者のどちらの立場にいるかによって考え方が異なることも示された。

6.　考察

6.1 量的・質的分析結果の検討

ビデオ映像を用いたスピーキング・パフォーマンスの自己評価において、A グループのボディランゲージ（姿勢・手足の位置・アイコンタクト・顔の表情）において、有意に高く評価されていることが示された。A グループと B グループは英語力も同等であることが示されており、普段は同じ教室で授業を受けていた。B グループより A グループのほうで出席率が若干高いことを考慮すると、A グループの方がまじめに授業に取り組み、自己評価ではビデオ映像を観ながら正当な評価を与えていたのではないかと考えられる。

次に、A グループのピア評価のボディランゲージでは 1 回目の評価者を記入した評価より、2 回目の匿名による評価の得点が有意に高くなり、評価者を記入すると評価が甘くなるという学生の回答と矛盾することが示された。しかし、A グループの自己評価でも、ボディランゲージは 1 回目より 2 回目のほうが高く評価されていることから、実際に学生のボディランゲージは 1 回目より 2 回目のほうが上達していた可能性が考えられる。ボイスコントロールや有効性は音声的な判断を求められるのに対して、ボディランゲージは視覚的な判断が必要とされることから、日本人英語学習者にとって音声的より視覚的な情報を用いて評価するほうが、より適正に評価しやすかった可能性もありうる。評価者を記入しても匿名であっても、ボディランゲージの場合、実際のパフォーマンスが優先されていたと考えられるだろう。

B グループの有効性では、1 回目（M = 3.90, SD = 0.05）の匿名によるピア評価より 2 回目（M = 3.73, SD = 0.09）の評価者を記入したピア評価で、有意に得点が低くなることが示された。1 回目のピア評価では、B グループ

の有効性の平均値が高いことから、天井効果の影響があったことが考えられる。そのため、1 回目より 2 回目のほうが実際の有効性は向上していたのにもかかわらず、低い評価になってしまった可能性もありうる。

ピア評価のボディランゲージと同様、有効性でも評価者を記入すると緩い評価になるという学生の回答とは一致しなかった。評価者を記入したピア評価では、匿名の場合より低い評価がつけられた理由として、ボディランゲージのように視覚的な情報を評価するのではなく、音声情報を理解しながら評価する必要があることから、リスニングが得意でなかったり、たまたま聞き逃してしまったりした学生の場合、有効性を正しく評価することは容易でなかったかもしれない。あるいは、実際のパフォーマンスにおいても、この有効性に関する項目が劣っており、評価者を記入したピア評価であったにもかかわらず、適正な評価が与えられたのではないかと推測される。

6.2 教育的示唆

本研究では、学習者のスピーキング・パフォーマンスにおけるピア評価の匿名と記名の両方で行うことによって、匿名より記名で評価をするほうが厳しい評価をつける傾向が示された。先行研究（Falchikov, 2005; Peng, 2009; 笠巻, 2016）では、ピア評価は緩くなりがちであることを指摘しているが、本研究の結果から、評価者を記入するかしないかによっても評価得点に差が生じることが明らかになった。自由記述の回答から、匿名によるピア評価はコメントを記入しやすいことや評価を厳しくしやすいなどの長所がある一方で、コメントの責任の所在が不明確になることが短所として挙げられる。庄ほか（2012）の研究では匿名または記名のいずれかを好むことを示唆しているが、本研究では 1 つのクラスを 2 グループに分けて比較したため、表 3 が示すようにグループ間の差異は見られなかった。それゆえ、匿名と記名のどちらがよりピア評価に有効であるかについての明確な結果は出なかった。

しかしながらピア評価で重要なのは、誰が書いたか（who）ではなく、どのようなこと（what）をどのように（how）書いたかではないかと考えられる。評価者と被評価者の人間関係などが評価に反映されるとピア評価の信

頼性の低下につながる（藤原ほか, 2007）。したがって、ピア評価を教育現場で導入するにあたって注意しなければならないのは、他の学習者に対して責任のある評価を与えることであると考えられる。

ソーシャルメディアなどの匿名による投稿とは異なり、教室という限られた学習環境で実施される匿名のピア評価では評価者を特定することができないが、同じ教室で学んでいる学生がそこに含まれていることは間違いない。評価者が匿名であっても、評価者として責任を持ちながら評価をつけるべきであり、評価者が特定されないことを理由にして言いたいことを遠慮なく書いてもよいわけではない。評価者が与えた評価に被評価者が納得することでピア評価の教育的効果が期待できる（藤原ほか, 2007）ことから、ピア評価に関わる様々な要因についても検証していくことが求められると考えられる。

本研究の対象者は、必修科目としてこの授業を履修していたことから、英語学習に対する関心がとりわけ高い学生の集団ではなかったが、ピア評価の匿名・記名に対して率直なコメントを様々な角度から寄せてくれた。ピア評価を匿名にするか記名にするかの判断については、先行研究（庄ほか, 2015）で指摘されているように、学習者の好みを事前に知ったうえで、ピア評価を実施することによって、ピア評価に対する責任感を高めることにつながると考える。また、実際に使用された評価シートを学生に例として示しながら、どのように評価されたら効果的であるか考えさせることも必要だろう。結果的に、ピア評価のコメントの質が上がると、学習者のパフォーマンスに良い影響を与えることになるのではないか。

7.　結論

本研究は大学での授業の一部として行われた準実験であるため、研究参加者の人数も十分に確保することができなかった。スピーキング・パフォーマンスのピア評価を匿名で実施することの長所と短所が明らかにされたとはいえ、自己評価とピア評価だけでなく、教師による評価を分析に加えることによって、匿名と記名の違いによるピア評価の信頼性をより厳密に検証できただろう。今後は、本研究とは異なる英語力の学生を対象としたピア評価の匿名性に関する研究を実施することで、ピア評価の教育的効果

を示すことが必要だと思われる。

謝辞

本研究は科研費(C)15K02730 の助成を受けたものである。また、清泉女子大学言語研究所の客員所員として活動した研究成果である。本研究をまとめるにあたり、澤海崇文氏（流通経済大学）から様々なご助言をいただいた。この場を借りて感謝の意を表する。

参考文献

Falchikov, N. (2005). *Improving Assessment Through Student Involvement.* London: RoutledgeFalmer.

Fukazawa, M. (2010). Validity of peer assessment of speech performance. *Annual Review of English Language Education in Japan, 21*, 181-190. http://www.jstage.jst.go.jp/article/arele/21/0/21_KJ00007108621/_pdf

Hirai, A., Ito, N., & O'ki, T. (2011). Applicability of peer assessment for classroom oral performance. *Japan Language Testing Association Journal, 14*, 41-59.
http://www.jstage.jst.go.jp/article/jltajournal/14/0/14_KJ00008421105/_pdf

Luoma, S. (2004). *Assessing Speaking.* Cambridge University Press.

Okada, Y., Sawaumi, T., & Ito, T. (2014). Different effects of sample performance observation between high and low proficiency English learners. *The 6th CLS International Conference Proceedings*, 394-413. http://www.fas.nus.edu.sg/cls/CLaSIC/clasic2014/proceeding.htm

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*(3), 239-250.

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing, 19*, 109-131.
http://www.drronmartinez.com/uploads/4/4/8/2/44820161/the_influence_of_peer_feedback_on_self-assessment_language_testing-2002-patri-109-31.pdf

Peng, J. F. (2009) *Peer Assessment of Oral Presentation in an EFL Context.*

Dissertation Abstracts International, 70-12 (UMI No. 3380148).

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing, 25*, 553-581.

Yamashiro D. A., & Johnson, J. (1997). Public speaking in EFL: Elements for course design. *The Language Teacher, 21*(4), 13-17. http://jalt-publications.org/tlt/articles/2124-public-speaking-efl-elements-course-design

岡田靖子・いとうたけひこ（2014）「自己評価・ピア評価からみた学習者のビデオ映像活用の効果」『日本大学経済学部研究紀要』*76,* 47-55. http://www.eco.nihon-u.ac.jp/about/magazine/kiyo/pdf/76/76-47-55.pdf

笠巻知子（2016）「スピーキング・パフォーマンスの評価―学生相互評価の信頼性」『言語と文化』*10*, 1-10. https://www.kufs.ac.jp/graduate/publications/pdf/kiyou10.pdf

藤原康宏・大西仁・加藤浩（2007）「学習者間の相互評価に関する研究の動向と課題」『メディア教育研究』*4*(1), 77-85.

庄ゆかり・三浦郷史（2012）「オンラインピアレビューを用いた学習における匿名性と効力感」『情報処理学会研究報告』*2012-CE-114*(2), 1-8.

庄ゆかり・長登康・稲垣知宏・隅谷孝洋・中村純（2015）「ピアレビューの匿名化が学生の評価行動に与える影響」『システム情報学会誌』*32*(1), 135-139. http://www.jsise.org/journal/journal_jp/pdf/vol_032/3201135.pdf