

パフォーマンス課題における自己評価力を高めるための 協議ワークを取り入れた相互評価活動の開発[†]

岩田貴帆^{*1}・田口真奈^{*2}

京都大学大学院教育学研究科^{*1}・京都大学高等教育研究開発推進センター^{*2}

パフォーマンス課題において教員の評価負担を増やさずに実施可能な形成的評価である相互評価活動は、自己評価力を高める利点もあると主張されている。しかし、ピアからの評価に依存してしまうと自己評価力は高まらない。そこで、他者の視点を取り入れて自己評価を見直す機会となるよう、自己評価とピアからの評価の得点やその根拠が異なった箇所について異なった理由を協議するワークを取り入れた相互評価活動を開発した。実際の大学授業をパイロットケースとして学習活動を実施し、活動の前後で教員評価と自己評価のルーブリック得点の差が小さくなるかどうかを統計分析したところ、部分的ではあるが有意に小さくなった。

キーワード: 自己評価力, 相互評価活動, 高等教育, レポート課題, ルーブリック

1. はじめに

パフォーマンス課題を通した学習において形成的評価は重要であるが、特に高等教育の講義科目等で教員が複数回評価を実施することは、教員負担の観点から実行可能性が低い。教員による形成的評価の代替として学生同士の相互評価活動が実施されることがあるが、学生が必ずしも適切に他の受講生（以下、ピアと称する）を評価できるわけではないため、パフォーマンスの改善に寄与しない場合があるという課題が存在する。

一方で、相互評価活動を実施することで自己評価力が高まるという積極的な利点も主張されている。NICOL *et al.* (2014) は、ピアから評価を受ける経験だけでなく、ピアを評価する経験の重要性に着目してい

る。ピアのパフォーマンスを評価する際に、評価基準の理解の深まりや、自分のパフォーマンスとの比較が促されることが学生へのインタビュー調査から明らかになり、これが自己評価力の向上につながるとされている。

相互評価活動を教員の形成的評価の代替手段と位置づけるのではなく、自己評価力を高めるための学習活動と位置づけ、自己評価力が高まった結果としてパフォーマンスを自ら改善できうる、という点に着目するのが本研究の着眼点である。

なお、自己評価力の定義は、SADLER (1989) や TAI *et al.* (2018) といった学生の評価力に関する研究を参考に、「評価基準を理解し、それを自分のパフォーマンスに適用することで、目標とのギャップを適切に把握する力」とする。

2. 問題と目的

2.1. ピアからの評価に依存する危険性

相互評価活動を実施すれば必ず自己評価力が向上するわけではない。相互評価活動の前後に自己評価の機会を設けない場合、ピアからの評価を無批判に受け入れてしまうことによって、むしろ自己評価力は高まらない可能性も考えられる。

安彦 (1987) は、自己評価の成長プロセスにおいて、他者評価を参考にしつつそれに依存しすぎないように

2019年4月3日受理

[†] Takaho IWATA^{*1} and Mana TAGUCHI^{*2} :
Development of Peer Assessment Activity
Incorporating Calibration Work for Improvement of
Student's Self-Assessment Ability in Performance
Tasks

^{*1} Graduate School of Education, Kyoto University,
Yoshida Honmachi, Sakyo-ku, Kyoto City,
606-8501 Japan

^{*2} Center for the Promotion of Excellence in Higher
Education, Kyoto University, Yoshida
Nihonmatsucho, Sakyo-ku, Kyoto City, 606-8501
Japan

することが重要であるとし、「自己評価Ⅰ」⇒「他者評価」⇒「自己評価Ⅱ」という自己評価の成長モデルを示している。このモデルから自己評価力向上のための相互評価活動を検討すれば、事前に自己評価をした上で相互評価活動を行い、他者の視点を踏まえて自己評価を見直すための参考としてピアからの評価を位置づけるような活動のデザインが必要と考えられる。

2.2. 先行研究の限界点

自己評価力を重視した相互評価活動の実践研究の中で、相互評価活動の前に自己評価を実施し、それらの評価結果を関連づける学習活動を行っている実践研究として、WILLEY and GARDNER(2009)や高橋ほか(2016)、REINHOLZ (2015) が存在する。

WILLEY and GARDNER (2009) や高橋ほか (2016) は、自己評価とピアからの評価の得点をコンピュータの画面上で学生が比較することができるツールを開発している。しかし、得点を比較するだけでは、ピアからの評価を無批判に受け入れてしまう危険性がある。ピアからの評価を絶対視せずに批判的に捉え、ピアの視点を踏まえて自己評価を見直すためには、得点のみならず、その根拠も比較する必要がある。

REINHOLZ (2015) は、自己評価とピアからの評価のあとに学生同士のディスカッションを行っているが、自己評価とピアからの評価の得点と根拠を比較することに焦点を当てた活動は行っていない。「自己評価Ⅱ」への移行を確実に促すためには、自己評価とピアからの評価について言語化し比較することを通して「自己評価Ⅰ」を相対化することを促すような学習活動をデザインする必要がある。また、自己評価とピアからの評価の得点や根拠を比較するためには、ループリックなどの評価基準を用いて自己評価やピアからの評価を実施する必要がある。

2.3. 本研究の目的

本研究では、自己評価力を高めるにあたり、他者の視点を踏まえて自己評価を見直すことができる相互評価活動を開発し、実際の授業での実施を通して自己評価力向上の効果を確認することを目的とする。具体的には、「事前に行った自己評価（自己評価Ⅰ）よりも、ピアからの評価を踏まえた上で再度行った自己評価（自己評価Ⅱ）の方が妥当に評価できている」という仮説を検証する。

3. 協議ワークを取り入れた相互評価活動の開発

安彦（1987）の述べるところの「自己評価Ⅰ」が、

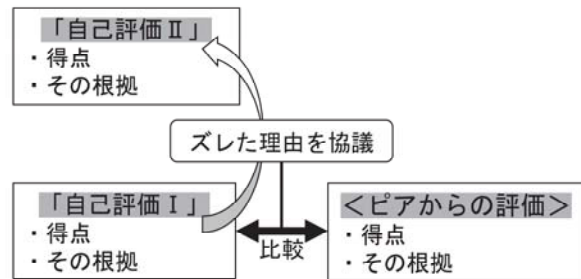


図1 協議ワークを取り入れた相互評価活動

＜他者評価＞を参考にしながら「自己評価Ⅱ」へと移行するために、相互評価活動に以下のような協議ワークと呼ぶ活動を取り入れる。協議ワークとはそれぞれ独立して実施された「自己評価Ⅰ」と＜ピアからの評価＞を比較して、その得点や根拠が異なっている箇所について、「なぜ異なっているのか」の理由を両者で話し合うワークである。その後、再度個人ワークとして自己評価を行うことにより、ピアからの評価に依存しすぎることなく「自己評価Ⅱ」への移行を促す(図1)。

また、実施にあたってのポイントとして、一般的な相互評価活動で行われる、ピアからの改善のアドバイスを設けないことが重要である。トライアルとして実施した授業実践（岩田・田口 2018）では、ピアからの改善のアドバイスの時間を設けたところ、この相互評価活動全体の目的（ピアからの評価を参考に自己評価を見直すこと）がぼやけてしまい、協議ワークの指示をした時間においても多くの学生がピアに対する改善のアドバイスをしていたことが録音データから明らかとなった。協議ワークが機能するためには、ピアからのアドバイスの時間は設けずに、ピアからの評価は自己評価を見直すためのきっかけであり、改善点を検討するのはあくまで本人であるという位置づけを学生に明確に伝えることが重要と考えられる。

4. 方 法

4.1. 大学授業「社会学Ⅱ」の文脈

開発のパイロットケースとして、ある国立大学の一般教養科目「社会学Ⅱ」（開講期間：2018年10月～2019年1月）において、学期末レポート課題の草稿を対象とした相互評価活動を実施した。

第一筆者はTAとして、授業担当者と協議しながら、授業運営・レポート課題とループリック作成・採点補助に関わった。「社会学Ⅱ」の学習到達目標は、「社会的思考法」を実践的に活用できることである。その

観点	観点1：社会現象の分析	観点2：幸せのための行動の分析
レベル4	レベル3の水準に加えて、報道や日常会話などにおける一般的な説明よりも深い説明によって、読んだ人に新たな認識枠組みを提供するような説明ができています。	レベル3の水準に加え、そのコミュニケーションがもたらしている負の側面についても考察できている。
レベル3	社会現象における人々のコミュニケーションを規定する構造を書けていて、その構造がどのようなコミュニケーションから形成されるか、説得力のある説明ができています。	自分のコミュニケーションが、構造にどのような影響を与えて、「幸せ」にどのようにつながるのかについて、実際の事例やデータなどを使って客観的な説明ができています。
レベル2	社会現象における人々のコミュニケーションを規定する構造を書けていて、その構造がどのようなコミュニケーションから形成されるかを説明しているが、説得力が低い説明である。	自分のコミュニケーションが、構造にどのような影響を与えて、「幸せ」にどのようにつながるのかについて検討しているが、主観的な説明にとどまっている。
レベル1	社会現象における人々のコミュニケーションを規定する構造を書けていない、もしくは、構造がどのようなコミュニケーションから形成されるかを説明できていない。	自分のコミュニケーションが、構造にどのような影響を与えて、「幸せ」にどのようにつながるのかについて検討していない。
レベル0	構造が説明されていない。	「幸せ」のために自分がとるべきコミュニケーションを書けていない。

図2 「社会学的思考法」を実践的に活用する力を評価するルーブリック

達成度を評価するパフォーマンス課題として、社会現象における自分自身の幸福について分析することを学期末レポート課題（約2000字以内）に設定した。ルーブリック（図2）は2観点5水準（レベル0～4）であり、観点1は社会現象の分析に関わる評価基準、観点2は自身の幸福の分析に関わる評価基準である。これら課題・ルーブリックの作成に際しては、レポートのサンプルを作成し、第一筆者と授業担当で独立に採点したのち協議することを複数回繰り返して修正を行った。

自己評価力に含まれる「評価基準の理解」を促すため、8週目、10週目、11週目の授業においてサンプルレポートを学生がルーブリックで評価したのち教員評価を解説するという活動を実施した。第12週目の授業で実施した相互評価活動の流れを表1に示す。③が協議ワークに該当する。④は自己評価力に含まれる「目標とのギャップの把握」を促すワークである。

学期末レポートは第15週目が提出期限であり、あわせて最終自己評価を提出するよう求めた。学生には、

草稿の内容や自己評価・ピアからの評価の結果は一切成績には含まれず、最終的な期末レポートのみが成績に反映されることを伝えた。各学生は、草稿に対する自己評価を2回（相互評価活動の前後）、完成稿に対する自己評価を1回実施した。

4.2. 変数化と統計分析

自己評価力（評価基準を理解し、それを自分のパフォーマンスに適用することで、目標とのギャップを適切に把握する力）を定量的に捉えるにあたり、以下のように変数化を行う。まず、教員は評価基準を理解し、学生のパフォーマンスに適用して目標（本実践ではルーブリックのレベル4）とのギャップを適切に把握する力を有していると考えられるため、教員評価を妥当な評価の基準値とする。一般に、教員間でも評価には多少のバラつきが生じることから教員評価を絶対視するべきではないが、教員評価と学生評価のズレに比べれば教員間のズレの方が小さいことから、自己評価と教員評価の差分が小さいほど自己評価力が高いと見なすことには一定の妥当性がある（斎藤ほか 2017）。

そこで、「自己評価Ⅰよりも、ピアからの評価を踏まえた上で再度行った自己評価Ⅱの方が妥当に評価できる」という仮説を検証するにあたり、ルーブリックの各観点における「自己評価と教員評価の差の絶対値」＝「自己評価の不適切さ」と変数化する。協議ワークを取り入れた相互評価活動を実施した「社会学Ⅱ」の受講者のうちデータ欠損のない25名を対象に「自己評価の不適切さ」を従属変数、「評価基準の観点（観点1～2）」「タイミング（1回目～3回目）」をそれぞれ対応あり要因とした二要因分散分析を実施する。仮説に対応するタイミングは1回目（自己評価Ⅰ）と2回目（自己評価Ⅱ）だが、評価対象が草稿ではなく改善後の完成稿に変わっても自己評価力が維持されるかを調

表1 相互評価活動①～④と完成稿提出までの流れ

①【自己評価Ⅰ】持参した草稿に対して、ルーブリックの各観点の自己評価Ⅰの得点と根拠をワークシートに記入（5分）
②【ピアからの評価】草稿を隣の学生と交換してルーブリックを用いて評価し、各観点の得点と根拠をワークシートに記入（13分）
③【協議ワーク】自己評価Ⅰとピアからの評価を突き合わせ、得点や根拠が異なっていた観点について異なった理由を2人で協議し、その結果をワークシートに記入（6分×2人）
④【自己評価Ⅱ】個人ワークとして、再度ルーブリックの各観点について自己評価Ⅱの得点とその根拠、学期末レポート提出までの改善点を記入（5分）
⑤草稿を改善し、完成稿を作成（授業外）
⑥完成稿に対する最終自己評価の得点とその根拠（3回目）をワークシートに記入（授業外）

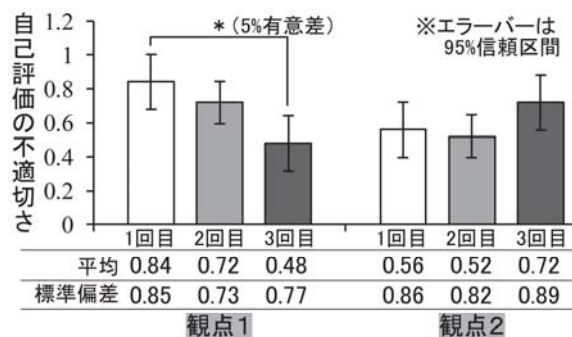


図3 自己評価の不適切さ

べるため、3回目の自己評価のデータも分析に用いる。

なお、上記のように研究のために草稿についても教員評価を実施したが、実践としては草稿の教員評価を実施する必要はない。本実践においても草稿の教員評価は学生にフィードバックしていない。

5. 結果と考察

「社会学Ⅱ」の受講者のうちデータ欠損のない25名の各観点・各タイミングの平均値の「自己評価の不適切さ」のグラフを図3に示す。二要因分散分析の結果、交互作用が有意 ($F(2,48)=4.076, p=.023, \eta_p^2=.145$) となった。そこで単純主効果検定を行ったところ、観点1については3回のタイミングで多重比較 (LSD 法) したところ、1回目と3回目で有意差があった ($p=.036$)。観点2についてはいずれのタイミングの組み合わせでも有意差がなかった ($p>.05$)。

有意差がみられた観点1について、相互評価活動の直後では有意には差が見られなかったが、完成稿においては有意差があったことから、書き直しのプロセスを経て自己評価力が定着していく可能性が考えられる。

6. まとめと今後の課題

本研究では、自己評価力を高める協議ワークを取り入れた相互評価活動を開発した。また、仮説検証の結果、部分的にはあるが自己評価は教員評価に近づいた。本研究の仮説検証は、あくまで全体的な傾向を捉えるものであった。今後の課題として、観点1と観点2で結果に違いが生じた理由を分析することや、授業の他の学習活動との関連を検討することから、なぜ・どのようにワークが機能し、自己評価力の向上に繋がるのかを明らかにする必要がある。さらに、今回確認できた効果は単一の課題における自己評価力の向上であり、他の場面でも維持されうるかは未確認である。

より多くの実践を通して一般化可能性を高めることや自己評価力を実証的に捉える方法の検討も必要である。

謝 辞

本研究にご協力賜りました関係者の皆様に深く御礼申し上げます。

参 考 文 献

- 安彦忠彦 (1987) 自己評価：「自己教育論」を超えて。図書文化，東京
- 岩田貴帆，田口真奈 (2018) 相互評価活動を取り入れた授業実践における学生の自己評価の変容：自己評価とピア評価のズレに着目して。日本教育工学会研究報告集，18(4)：9-16
- NICOL, D., THOMSON, A. and BRESLIN, C. (2014) Rethinking Feedback Practices in Higher Education: a Peer Review Perspective. *Assessment & Evaluation in Higher Education*, 39(1)：102-122
- REINHOLZ, D. L. (2015) Peer-Assisted Reflection: A Design-based Intervention for Improving Success in Calculus. *International Journal of Research in Undergraduate Mathematics Education*, 1(2)：234-267.
- SADLER, D. R. (1989) Formative Assessment and the Design of Instructional Systems. *Instructional Science*, 18(2)：119-144
- 斎藤有吾，小野和宏，松下佳代 (2017) ルーブリックを活用した学生と教員の評価のズレに関する学生の振り返りの分析：PBL のパフォーマンス評価における学生の自己評価の変容に焦点を当てて。大学教育学会誌，39(2)：48-57
- TAI, J., AJJAWI, R., BOUD, D., DAWSON, P. and PANADERO, E. (2018) Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work. *Higher Education*, 76(3)：467-481
- 高橋暁子，金西計英，松浦健二，吉田博，和田卓人 (2016) 自己評価と相互評価の差異を可視化するための携帯端末用ルーブリック評価ツールの開発と試用。教育システム情報学会誌，33(2)：120-125
- WILLEY, K. and GARDNER, A. (2009). Improving Self- and Peer Assessment Processes with Technology. *Campus-Wide Information Systems*, 26(5)：379-399

(Received April 3, 2019)