

学生の自律的なパフォーマンス改善を促す学生主体の評価活動

—初学者対象の大学授業にどのように取り入れるか—

岩田 貴帆

1. はじめに

大学教育において、知識・スキルを活用して取り組む課題を学生に与えることは馴染み深い。学問的分野であれ、職業的分野であれ、なんらかの専門的な知識・スキルを活用して課題に応えるようなレポートやプレゼンテーション、教職や医療の実技演習、卒業制作や卒業論文などは、古くから大学の授業に取り入れられてきた。これらは成績付けの機能を果たしてきた一方で、そういった課題に取り組むこと自体が、学生にとっては知識・スキルを活用する力を高める学習機会となる機能については、近年になって注目されはじめた。

1990年代頃から、真正の学力を評価する理論や枠組みとして「真正の評価」(Wiggins 1998)、パフォーマンス評価 (Hart 1994) といった理論・概念が登場した。そういった、知識・スキルを活用する力を要する複雑な課題の評価規準 (criteria) をめぐっては、評価規準を下位要素に分解するのではなく、どの程度達成しているかの水準 (standard) に準拠してパフォーマンスの全体的な質を評価する「スタンダード準拠評価」が提唱された (Sadler 1987)。加えて、教員だけが学習を評価する主体ではなく、学習者本人を評価主体として積極的に位置づける形成的評価論 (Sadler 1989, Earl 2003) も登場した。これらの研究に共通する目標は、学習者が、課題における自分のパフォーマンスを自己評価することで優れた点・改善すべき点を把握し、自律的にパフォーマンスを改善できるように促すことだといえる。「よりよいパフォーマンスを生み出すべく、学習者⁽¹⁾自身が評価規準を共有し、思考過程を自己調整すること」(石井 2015, p.46) によって、知識・スキルを活用する力を育むことが期待できる。

だが、自律的にパフォーマンスを改善できるようになるために必要な評価規準を理解することは、特に当該分野の課題に初めて取り組む初学者の学生にとって容易でない。評価規準 (パフォーマンスを評価する際に採用される観点や側面) は汎用的なものではなく、個々の大学授業の背景にある学問分野や職業分野において、何が価値があると認められているかによって異なるものである。初学者にとってその分野の評価規準は未知のものであるため、当該分野の専門家である大学教員がどのような観点でパフォーマンスを評価しようとしているのか、初学者が理解できるとは想定できない。例えば、心理統計学の実習で分析結果を報告するレポート課題において、「帰無仮説が棄却された」という分析結果をどのように報告し主張を導くことが適切なのかは、心理統計学の分野でどのような書き方が優れているとされるかに依存するだろう。加えて、その評価規準において質の低い水準から質の高い水準までどのような段階が存在する

か、という評価基準（どの程度優れているのか、という達成の度合いや水準）に関しても、初学者は全く知らない状態である。このため、課題を与えられるだけでは、どのようなパフォーマンスが優れたものとされるのか、そこに至るステップはどのようなものなのかかわからないまま闇雲に取り組むことになり、「ランダムで非効率な試行錯誤」(Sadler 1989, p.120) となってしまう。したがって、初学者の学生が課題に取り組む前に、評価規準や評価基準を学生が理解できるように指導することが重要とされている。

そのためにはどのような指導が必要なのだろうか。従来、形成的評価論では教員から学生へのフィードバックが重視されてきたが、近年、フィードバックのみに頼る指導の限界が指摘されている。その理由は次の2点にまとめることができる。1点目は、教員リソースには限界があるため、学生に個別のフィードバックを行うことが難しい場合がある点である (Boud & Molloy 2013)。2点目は、教員が丁寧にフィードバックを行ったとしても、学生が評価規準を理解していなければ、フィードバックの意味を理解することができない点である (Sadler 2010)。1点目は実際的な問題であるが、2点目は指導の方針を検討する上で本質的に重要である。教員からのフィードバックよりも前に実施する指導の必要性が浮かび上がる。

そこで多くの先行研究(e.g. Yucel et al. 2014; Reinholz 2016)が着目しているのが、Sadler (1989)の提唱した、学生が評価主体となる評価活動である。学生が評価者の立場となって何らかのパフォーマンスを評価する評価活動を授業に取り入れることによって、当該課題ではどのようなパフォーマンスが優れているのかを経験的に学ぶことができる。ただし、評価活動において、学生が何を評価するのかという評価対象は、先行研究において様々なバリエーションが存在する。自分のパフォーマンスを対象とする自己評価、ピアのパフォーマンスを対象とするピア評価に加えて、教員の用意したパフォーマンスの例を評価する活動も存在する。それぞれの評価活動に関する研究は多くみられるが、これらを横断的に検討し、どのように授業に取り入れることが適切かは明らかになっていない。

そこで本稿では、初めて当該分野の課題に取り組むような初学者を対象とした授業において自律的なパフォーマンス改善を促すには、どのように評価活動を取り入れることが適切かを、先行研究のレビューに基づいて検討することを目的とする。

その方法として、まず、こういった評価活動が存在するのかを特定するために、この領域を理論的・実証的に研究している豪州のディーキン大学 CRADLE センターの研究チームによるレビュー論文 (Tai et al. 2018) を参照する。Tai et al. (2018) では、学生の自律的なパフォーマンス改善を促しうる教育実践として、①自己評価、②ピア評価／ピアフィードバック／ピアレビュー⁽²⁾、③フィードバック、④ループリック、⑤典型事例 (exemplars) の5つを挙げている。2.では、これらのうち評価活動に該当するものを特定する。その際、自律的なパフォーマンス改善を促す学生主体の評価活動として適切な定義を検討する。3.では、それぞれの評価活動の利点と課題を先行研究に基づき整理する。そして4.では、異なる評価活動を組み合わせることによってそれぞれの課題を補い合う可能性について検討する。

なお本稿では、『教育評価重要用語辞典』の「評価規準と評価基準」(石田 2021a) を参考に、評価規準 (criteria) は「パフォーマンスを評価する際に採用される観点や側面」の意味で、評価基準 (standard) は「パフォーマンスがどの程度優れているのか、という達成の度合いや水準」

の意味で用いる。これら両方を指す場合には、「評価規準や評価基準」とする冗長さを避けて、「評価基準（規準）」と記す。

2. Tai et al. (2018) の 5 つの教育実践から特定する評価活動とその定義

2. では、Tai et al. (2018) が列挙した 5 つの教育実践のうち、学生主体の評価活動に該当するものを特定する。その条件は、「学生が評価する主体となること」、「自律的なパフォーマンス改善を促すために教員が授業に取り入れる活動であること」、そのために、「学生がパフォーマンスに再度取り組んだり、修正したりする機会があること（形成的評価であること）」である。これら全てにあてはまるものが評価活動であるとする。評価活動として捉えることができるよう、必要に応じて定義の検討も行う。

①自己評価

Tai et al. (2018) は自己評価を「学生が自分自身の作品を評価すること」としている。また、Tai et al. (2018) の共著者で CRADLE のセンター長である Boud (1995, p.12) は、自己評価とは「学生が自分の作品に適用すべき評価基準（規準）が何であるかを特定して、それらをどの程度満たしているかを判断すること」としている。これらを踏まえつつ、「自律的なパフォーマンス改善を促すために教員が授業に取り入れる活動」という条件にあてはまるように範囲を限定して捉える自己評価を本稿の議論の対象とする。すなわち、評価活動としての自己評価とは「教員の教授的意図をもった指示により学生が取り組む、評価基準（規準）に照らして自分自身のパフォーマンスを評価し、優れた点・改善点を判断する活動」である。

②ピア評価／ピアフィードバック／ピアレビュー

次に、ピア（ある学生個人にとって、自分と同じ授業を受講し、同じ課題に取り組んでいる他の受講生）のパフォーマンスを評価することについて、Tai et al. (2018) はピア評価／ピアフィードバック／ピアレビューという 3 つの呼称を併記している。このうちピアレビューは「学生がピアの作品について評価し、判断を下した上で、フィードバックコメントを書いて構築すること」(Nicol et al. 2014, p.103) とされることから、ピア評価とピアフィードバックの両方の要素を含む最も包括的な語といえ、本稿ではピアレビューを検討の対象とする。

Nicol et al. (2014) は書き言葉によるフィードバックに限定しているが、伝達的手段としては話し言葉もありうる。また、自己評価同様、大学授業の背景にある専門分野に由来する評価基準（規準）に照らしてピアレビューが行われることが重要である。これらも踏まえて、ピアレビューを「教員の教授的意図をもった指示により学生が取り組む活動であり、学生が評価基準（規準）に照らして、ピアのパフォーマンスの優れた点・改善を要する点を判断し、パフォーマンスをした本人に伝達する一方で、自分のパフォーマンスの優れた点・改善を要する点をピアから伝達される活動」とする。そしてこれを評価活動の 1 つとして捉えることにする。

③フィードバック

Tai et al. (2018) は、「教員」に類する語 (teacher/tutor/staff) を一度も使わずにフィードバックという教育実践を論じている (pp.474-476)。このフィードバックという教育実践が何を意味するのか読み取りづらいが、「対話としてのフィードバック」「学生へのコメント」という記述からは、教員が学生のパフォーマンスを評価した上で学生になんらかの関与があるものを想定

していることは確かである。そこで本稿では、③フィードバックは、「教員から学生へのフィードバック」であると明確化する。その定義は「教員が学生のパフォーマンスを評価し、優れた点・改善を要する点を判断した結果を、本人に伝達すること」とする。これは、教員が評価主体であるため、学生主体の評価活動には該当しない。

④ルーブリック

ルーブリックとは、「成功の度合いを示す数値的な尺度あるいは評語と、それぞれの数値や評語にみられる認識や行為の質的特徴を示した記述語からなる評価基準表のこと」（石井 2015, p.45）であり、本稿でもこれをルーブリックの定義とする。この定義からもわかるように、ルーブリック自体は評価活動ではなく、評価活動で用いられるツールである。

それでは、どのような目的で評価活動に用いるツールなのだろうか。これを明確にするために、評価規準と評価基準がそれぞれどのようにルーブリックに反映されるか確認していく。

まず、ルーブリックには成功の度合いが設けられているところに、評価者が評価基準を用いて評価しようとしていることが反映されている。また、1 観点のみの全体的ルーブリックの場合には、1 つの評価規準で評価していることになり、複数観点の分析的ルーブリックの場合には、複数の評価規準で評価していることになる。いずれの場合にも、評価者の考える評価規準が、ルーブリックの観点に反映されていることになる。そして、各観点の各水準には、該当するパフォーマンスの質的特徴が記述語として示されているが、これは評価者の考える評価規準と評価基準を、言語的手段によって表現したもの（Sadler (1987) のいう Verbal description）といえる。

このように、ルーブリックには評価基準（規準）が言語的手段によって反映されているため、「評価基準（規準）に照らして」評価する活動である自己評価やピアレビューの際に、評価基準（規準）を言語的手段によって確認するためのツールとして捉えることができる。他方で、言語的手段に頼らなくとも、評価基準（規準）を思い出したり、他の手段によって確認したりすることが可能なのであれば、ルーブリックを用いなくとも、評価基準（規準）に照らした評価活動は実施可能といえる。

⑤典型事例

典型事例（exemplars）とは、パフォーマンスの「質または能力がどの程度優れているかを定めた水準の典型的なものとなるように選ばれた重要な例」（Sadler 1987, p.200）であり、本稿でもこの定義を用いる。パフォーマンスの質がどの程度優れているかを定めた水準とは、「A/B/C/D/F」や「優/良/可/不可」といった評語や、その評語に相当するパフォーマンスの一般的特徴を説明した記述語を添えたものであるルーブリックが該当する。そういった水準に典型的なパフォーマンス、例えば、「A」という評語に該当するパフォーマンスが典型事例となる。

定義からわかるように、典型事例自体は評価活動ではない。典型事例を学生が評価することが評価活動にあたり、評価基準（規準）の理解を学生に促すとされる（Rust et al. 2003）。そういった評価活動を実施している先行研究では assessment workshop (Rust et al. 2003), marking class (Hendry & Anderson 2013), marking exercise (Yucel et al. 2014) と様々に呼称されている。本稿では、学生が課題の遂行や自己評価・ピアレビューの前に、あらかじめ評価者の立場にたって評価する練習を行う、という意義を踏まえ、評価練習と称することにする。

評価練習の前提として、学生に典型事例を提示する際には、教員による評価（得点やその根拠）は伝えない。また、重要なプロセスとして、典型事例を学生個人が評価するだけでなく、評価結果を踏まえた学生同士の議論、教員からの解説（guidance）や教員・学生間の対話が重視されている（To & Carless 2016）。以上の議論を踏まえ、評価練習を「パフォーマンスの典型事例を学生が評価した上で、評価結果とその根拠を学生同士で議論したり、教員から解説したりすることで、学生が評価基準（規準）を理解する活動」と定義する。典型事例は、評価練習における評価対象であると位置づくことになる。また、評価練習においてもルーブリックは、評価基準（規準）を言語的手段によって確認するためのツールとして利用可能である。

小括

以上から、自律的なパフォーマンス改善を促す評価活動は(A) 自己評価、(B) ピアレビュー、(C) 評価練習の3つであると特定された。Tai et al. (2018) の5つの教育実践との関係や、3つの評価活動の定義を表1に示す。それぞれの評価活動の評価対象は異なる一方で、3つの評価活動に共通する点は、学生が評価主体となること、評価基準（規準）を言語的手段によって確認するためのツールとしてルーブリックを用いることがあることである。

表1 Tai et al. (2018) の5つの教育実践から特定した評価活動とその定義

教育実践	①自己評価	②ピア評価/ピアフィードバック/ ピアレビュー	③フィードバック	④ルーブリック	⑤典型事例
			評価主体が 教員	ツール (表の最 下段へ)	
評価活動	(A) 自己評価	(B) ピアレビュー	(C) 評価練習		
定義	教員の教授的意図をもった指示により学生が取り組む、評価基準（規準）に照らして自分自身のパフォーマンスを評価し、優れた点・改善点を判断する活動	教員の教授的意図をもった指示により学生が取り組む活動であり、学生が評価基準（規準）に照らして、ピアのパフォーマンスの優れた点・改善を要する点を判断し、パフォーマンスをした本人に伝達する一方で、自分のパフォーマンスの優れた点・改善を要する点をピアから伝達される活動	パフォーマンスの典型事例を学生が評価した上で、評価結果とその根拠を学生同士で議論したり、教員から解説したりすることで、学生が評価基準（規準）を理解する活動		
評価対象	本人の パフォーマンス	ピアのパフォーマンス	教員の用意した典型事例		
評価主体	学生				
ルーブリック	評価基準（規準）を言語的手段によって確認するためのツールとして用いることがある				

3. 3つの評価活動の利点と課題

それでは、これら3つの評価活動を授業にどのように取り入れることが有効と考えられるだろうか。3. では、自律的なパフォーマンス改善を促すにあたってのそれぞれの利点と課題を、これら評価活動を実施した実践の報告を含む先行研究を踏まえて整理していく。

(A) 自己評価の利点と課題

下川原 (2017) は、1 週目の看護実習のあとに形成的評価として自己評価を取り入れたことで、ほとんどの学生が実習要項の「実習目標」や「実習評価」の内容を意識しながら2 週目の実習に取り組むことができていたことを報告している。また、幼稚園教諭養成課程における授業実践を報告した藤井 (2018) は、模擬保育後に実施した自己評価において、学生が自分自身の実習を振り返る経験を言語化して発表することで、自分の成長と課題を明確化することができたとしている。このように、自己評価の利点は、次のパフォーマンスにおいて改善点を修正しようとするきっかけとなる点にあるといえる。

一方で、自己評価の課題として、看護実習の実践後に自己評価を実施した丸茂・河部 (2009) は、「初学者である学生は看護観の形成途上にあることから実践時の感覚のままに再構成しやすく、それ故に自分の関心で看護場面の再構成が部分的にな」ってしまうため、自己教育につながる自己評価とするには、他者の助言を要するとしている。同様に、Reinholz (2016) は、「自分の作品について詳細に判断をくだすことは、しばしば他者の作品を判断することよりも難しい。なぜなら、自己評価における評価者は、作品と『近すぎる』し、距離をとった客観的な観点を欠いているからである」としている。これらのことから、自己評価のみでは、自分の視点だけに頼った主観的な自己評価に陥ってしまい、自分のパフォーマンスの優れた点・改善を要する点を適切に把握することが難しいという課題があるといえる。

(B) ピアレビューの利点と課題

ピアレビューは、評価するプロセスと評価を受けるプロセスの両面において利点がある。

まず、ピアを評価するプロセスについて McConlogue (2012) は、質の高いライティングがどのようなものかを学生がより深く理解できるようになったことを報告している。また、Nicol et al. (2014) は、ピアレビューに取り組んだ学生へのインタビュー調査から、学生は教員の示した評価規準だけでなく、自分自身のパフォーマンスも内的表象に置き、それと比較しながらピアを評価するため、自分自身のパフォーマンスの優れた点や改善すべき点の理解も進む、ということ进行を明らかにしている。これらをまとめると、ピアのパフォーマンスを評価するプロセスを通して、評価基準 (規準) をより深く理解したり、それを自分のパフォーマンスに適用したりするきっかけとなるという利点があるといえる。

また、ピアから評価されるプロセスについて Orsmond et al. (2002) は、ピア評価者はプロセスとは独立して成果物を客観的に見るので、本人よりも客観的な判断を下すことができる、としている。レポート課題でいえば、筆者の観点ではなく読者の観点を学生が持つことを促す (Cho et al. 2010) ことができる。これらのことから、ピアから評価を受けるプロセスによって、自分のパフォーマンスを客観的に見た情報を得られるという利点が考えられる。

一方で課題として、どのようにピアを評価し、フィードバックを提供するかを学ぶトレーニングを欠いたままピアレビューを実施した場合、ピアから有益なフィードバックが提供されないということが挙げられる (Sridharan & Boud 2019)。評価規準を明示されていたとしても、学生がその意味を理解していなければ、適切にピアを評価し、有益なフィードバックを提供することはできない (McMahon 2010)。ピアからの評価が誤っていた場合に起きる問題として、自分のレポートは改善の余地があると適切に認識できていたのに、「君のレポートはとてもよく

表 2 学生の自律的なパフォーマンス改善を促す上での 3 つの評価活動の利点・課題

評価活動 の名称	(A) 自己評価	(B) ピアレビュー	(C) 評価練習
利点	・ 次のパフォーマンスにおいて改善点を修正しようとするきっかけとなる	・ ピアを評価することで、評価基準（規準）をより深く理解したり、それを自分のパフォーマンスに適用したりするきっかけとなる ・ ピアから評価を受けることで、自分を客観的に見た情報を得られる	・ 教員の考える評価基準（規準）を学生が具体性を伴って理解できるようになる ・ 自分のパフォーマンスの質を高めることができる
課題	・ 自分の視点だけに頼った主観的な自己評価に陥ってしまい、自分のパフォーマンスの優れた点・改善を要する点を適切に把握することが難しい	・ ピアが評価基準（規準）の意味を理解していないと、有益なフィードバックにならない ・ ピアからのフィードバックが誤っていた場合、本人が無批判に信じてしまうとパフォーマンスの改善点を見失ってしまう	・ 課題で求められている評価基準（規準）を理解できても、それを自分のパフォーマンスにうまく適用することはできないとは限らない

できている」というピアからの誤ったフィードバックを信じてしまい、自分のレポートの改善は不要であると判断してしまった学生の例が報告されている（Yucel et al. 2014）．これらから、ピアレビューの課題として、ピアが評価基準（規準）の意味を理解していないと、有益なフィードバックにならないこと、ピアからのフィードバックが誤っていた場合、本人が無批判に信じてしまうとパフォーマンスの改善点を見失ってしまうことの 2 点が存在するといえる．

(C) 評価練習の利点と課題

評価練習の利点として Rust et al. (2003) は、学生が典型事例を評価し、教員や他の学生と対話することを通して、ルーブリックや教員による解説といった言語的説明で提供される明示的な知識を補完するような暗黙知を含めて評価基準（規準）を教員から学生に共有することができることを挙げている．その結果、対照群と比較して、評価練習に参加した学生はレポート課題の得点が統計的に有意に高かったことを報告している．同様に、学生が「小論文で何が要求されているかを明確に理解することができ」（Hendry & Anderson 2013）、自分が課題に取り組む際の有益なガイドとなった（Hendry et al. 2012）と調査結果をまとめている．これらより、教員の考える評価基準（規準）を学生が具体性を伴って理解できるようになること、それにより学生が自分のパフォーマンスの質を高めることができること、が評価練習の利点であるといえる．

一方、Rust et al. (2003) では、評価練習を実施した群と、実施していない群の自己評価の適切さを比較して、期待に反して、両群に違いがみられなかったとしている．その理由として自信過剰な学生が存在が考察されている．課題で求められている評価基準（規準）を理解できても、それを自分のパフォーマンスにうまく適用することができないとは限らない点が、評価練習の課題といえるだろう．

小括

以上に整理してきた 3 つの評価活動の利点と課題を表 2 にまとめた．このうち、自己評価は主観的になりがちであるという課題は、ある意味自明なことではあるが、当該研究領域におい

ては確認しておくべき重要な点と考えられる。なぜならば、先述の Rust et al. (2003) は評価練習によって自己評価が適切になると仮説立てていたように、評価基準（規準）を理解することと適切に自己評価できることを同一視した先行研究もみられるからである。Rust et al. (2003) の結果を踏まえると、適切に自己評価するという行為は、評価基準（規準）を理解するという側面だけでなく、自信などの主観的なバイアスを乗り越えて、自分のパフォーマンスを客観的に捉えるという側面も存在するといえる。したがって、自己評価が主観的になりがちであるという課題を踏まえて評価活動を授業に取り入れることは重要といえるだろう。

4. 評価活動の相補的な活用

3. で明らかになったように、単一の評価活動を実施するのみでは、それぞれに課題がある。そこで 4. では、初めて当該分野の課題に取り組むような初学者を対象とした授業を念頭に、評価活動を組み合わせることで、それぞれの課題を、それぞれの利点で補い合える可能性について検討する。

4.1. 自己評価の課題をピアレビューの利点が補う可能性

自己評価の課題は、自分の視点だけに頼った主観的な自己評価に陥ってしまい、自分のパフォーマンスの優れた点・改善を要する点を適切に把握することが難しいということであった。これに関連して、自己評価の成長プロセスを理論的に提示した安彦（1987）は、他者からの評価を参考にしつつそれに依存しすぎないように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」というプロセスを示している。「自己評価Ⅰ」は主観的で自分の都合に合わせたものであり、「自己評価Ⅱ」は、より正確で客観的な視野の広いものとされる。その移行には、〈他者評価〉を受けるだけでなく、その是非を自ら判断して取捨選択することが重要であるという。ピアレビューの利点の1つに、ピアから評価を受けることで、自分を客観的に見た情報を得られるという点があるため、ピアレビューは〈他者評価〉として利用可能性を有するものといえる。だが、注意を要するのは、ピアから評価を受けた学生が、その是非を自ら判断して取捨選択することが重要である点だ。ピアからの評価結果に依存してしまうと、自分のパフォーマンスの優れた点・改善を要する点を自分で把握する状態は実現できないであろう。したがって、ピアからの評価に依存しすぎず参考にすることで自己評価の客観性を高めることを促すことができる場合には、自己評価の課題をピアレビューが補う可能性があるといえる。

4.2. ピアレビューの課題を評価練習の利点が補う可能性

ピアレビューには、ピアが評価基準（規準）を理解していないと、有益なフィードバックにならない、という課題がある。Yucel et al. (2014) では、これに対処するために、あらかじめ評価練習を実施して評価基準（規準）の理解を学生に促した上でピアレビューを実施している。このように、評価練習の利点である、具体性を伴いながら、評価基準（規準）を理解できるということによって、ピアレビューの課題を乗り越えることが期待できる。

ピアレビューの利点には、ピアを評価することで、評価基準（規準）をより深く理解できるというものもあったが、もともと一定の理解がある学生が、より深く理解できるようになるという意味であることに留意を要する。そもそも評価基準（規準）を全く知らない初学者の場合

は、あらかじめ別の手段で指導しておく必要がある。ここで、教員のもつ評価基準（規準）を学生に共有することができるのは評価練習のみであるということに注目したい。これをもたらすのは、教員が典型事例をどう評価するかを解説したり学生と対話したりすることで、教員の考えを共有できるという典型事例に特有の構造である。自己評価やピアレビューでは、クラス内の一人ひとりの学生が評価する対象が異なるため、評価結果の誤りを教員が訂正する機会を設けることは難しい。評価練習では、クラス全体で共通のパフォーマンスを対象にして評価し、対話することができる。それゆえ、対話を通して、教員のもつ評価基準（規準）についての考えを共有できると考えられる。

4.3. 評価練習の課題を自己評価の利点が補う可能性

では、評価練習だけ実施すれば十分ではないか、という意見もありうる。3. で整理したように、評価練習によって評価基準（規準）を理解できたとしても、それを自分のパフォーマンスに適用することができるとは限らない。これを補うものとして、岩田（2020）では、評価練習を実施したあとに、自己評価を実施している。評価練習のみを実施したあとに学生が作成したレポートの平均点に対して、その後自己評価とピアレビューを実施して、学生が改善したレポートの平均点が統計的に有意（0.1%水準）に向上していたことが報告されている。このパフォーマンス改善は自己評価を実施したことだけに起因するわけではないが、少なくとも、評価練習だけでは、パフォーマンスの改善に十分ではないことが示唆される。3. で整理したように、自己評価の利点は、次のパフォーマンスにおいて改善点を修正しようとするきっかけを学生に与えることができることであった。学生の中には、課題遂行プロセスにおいて、評価基準（規準）を自ら意識して適用しようとはしない学生も一定数いるかもしれないが、岩田（2020）のように、授業において全員が取り組む学習活動として自己評価を実施すれば、多くの学生が評価基準（規準）を自分のパフォーマンスに適用することができるようになる可能性がある。

4.4. 評価活動の組み合わせと具体的な5つの手順の抽出

以上から、自己評価・ピアレビュー・評価練習の3つを組み合わせることで、初学者向けの授業で自律的なパフォーマンス改善を促すことができる可能性があるといえる。3 つすべてを取り入れた授業実践を報告している文献はかなり限定的だが、社会学に固有の思考法を活用した文章作成力の育成を目指す授業科目における、岩田（2020）の実践研究を挙げることができる。3 つの評価活動をどのように授業に取り入れるか、授業設計の参考となるより具体的な示唆を得るために、授業科目全体の実施手順を抽出してみよう。

- (1) 【授業準備】前年度の受講生が提出したレポートの中から、典型事例を選ぶ。
- (2) 【評価練習】教員評価は伏せた状態の典型事例を配付し、ルーブリックを使って評価するように指示する。学生は、個人で評価した結果を、グループで共有する。最後に、教員評価とその根拠を解説し、質疑応答を行う。
- (3) 【課題遂行】授業時間外に、学生がレポートの草稿を作成する。
- (4) 【自己評価とピアレビュー】学生が授業に持参した草稿を対象に、自己評価を事前に行った上でピアレビューを実施。自己評価とピア評価の結果を比較して、ズレた理由を両方で検討する協議ワークを行った上で、再度自己評価を実施し、改善すべき点を明確化することを促す。

(5) 【改善・提出】学生は、レポートを改善した上で最終的なパフォーマンスを提出する。

このような手順で評価活動を取り入れたところ、(5)における改善は、教員からみても適切な改善がなされており、教員評価得点は統計的に有意に向上したことが報告されている(岩田2020)。草稿に対して教員からの個別のフィードバックは行われていないという意味において、自律的なパフォーマンス改善が促されたといえるだろう。

4.5. 評価活動にルーブリックを用いることの負の側面

上述の岩田(2020)以外にも、複数の評価活動を授業に取り入れている実践研究では、複数の評価活動に共通のルーブリックを用いている(e.g. Yucel et al. 2014)。自己評価やピアレビューにおいて評価基準(規準)に照らして自分やピアのパフォーマンスを評価するにあたり、評価練習と共通のルーブリックを利用することで、評価練習で理解した評価基準(規準)を思い出すことが容易となると想像できる。加えて、Yucel et al. (2014)はルーブリックを用いた実践の背景として、大学というアカデミックな場における評価に大きな不安を感じている初年次学生が多いことを挙げている。本稿の対象としている初学者の学生にとっては特に、どのような特徴を有するパフォーマンスであればどのような評価を受けるのかが明確に示されているルーブリックは、不安軽減という点でも有益であると思われる。

しかし、1.で概説したように学生主体の評価活動の理論的土台の1つとして参照されるSadler(1987, 1989)は、近年の論文(Sadler 2009, 2014)においてルーブリックにラディカルな批判を展開している。評価活動を取り入れる際にルーブリックを用いるかどうかを授業者が検討するための参考となるよう、ルーブリックの負の側面を、Sadler(2009, 2014)の論に詳しい石田(2021b)に基づいて確認しておこう。

まず、批判の1点目は、事前に評価基準が複数設定された分析的ルーブリックでは、本来捉えるべきパフォーマンスの全体的な質を判断することができない場合があることである(Sadler 2009)。確かに、実際の授業における学生のパフォーマンスは、事前に設定した評価基準では捉えきれない特徴を有していることはありうる。自己評価やピアレビューの評価対象がそのようなパフォーマンスであった場合、ルーブリックを用いると、3.で整理したような利点は損なわれてしまうかもしれない。

批判の2点目は、そもそも記述語によって評価基準を定義することは本質的に不可能ということである(Sadler 2014)。Sadler(1987)ではこの問題を乗り越えるものとして典型事例を補完的に用いることを提案したが、Sadler(2014)では、ルーブリックと典型事例が1対1に対応することはないため、評価基準を純粋に表現することはしばしば不可能であると論じられている。確かに、例えば「独創的な表現」といった記述語に対応する典型事例を教員が用意して評価練習をしたとして、独創的な表現のパフォーマンスには様々なタイプや水準があるにもかかわらず、学生は独創的な表現とは何かについて、特定の典型事例のみに基づいた限定的な理解しか得られないかもしれない。そのような場合には、ルーブリックを用いて評価練習を行っても、3.で整理したような利点が損なわれるかもしれない。

このような負の側面が顕在化し、評価活動の利点を脅かすことが危惧される場合には、ルーブリックを使用することは適切でないだろう。特に、ルーブリックが実際の学生の多様なパフォーマンスに裏打ちされていないようなルーブリックの場合、すなわち妥当性の低いルーブ

リックである場合に、Sadler (2009, 2014) の指摘するような負の側面が顕在化するおそれがある。

ルーブリックの妥当性に関連して田中 (2008) は、各水準に対応するパフォーマンスの例と対応付けながら複数の者が合議によって作成することが重要としている。4.4. で引用した岩田 (2020) の場合も、田中 (2008) を踏まえて、TA や協力者の作成したレポートを参照しながらルーブリックを作成している。さらに、前年度の受講生が提出した多様なレポートの採点を通して、ルーブリックが妥当なものであると判断したという。その結果、TA と授業担当教員のルーブリックを用いた採点の一致率は 94.5% だったと報告されている (岩田 2020, p.119) ことから、一定の妥当性があるとみられる。Sadler (2009, 2014) の指摘する問題が完全に回避されているかを検証する手段はないが、少なくとも、このように田中 (2008) の論を参考にルーブリックの妥当性を高めることは、1 つの対応策となる可能性がある。

以上を踏まえ、事例に基づいて作成することで妥当なルーブリックを作成できる場合には、複数の評価活動において共通のルーブリックを使用することが適切といえるだろう。

ただし、事例を集めたとしても妥当なルーブリックを作ることが難しい場合もありうる。また、授業科目が設置された初年度であったり、課題の内容が大幅に改訂された初年度であったりして、事例を集めることが困難なケースもありうる。そのような場合には、無理してルーブリックを作成することは推奨されない。初年度は、記述語なしの評語のみの「A/B/C/D」を評価ツールとして用いて評価活動を実施し、学生が提出したパフォーマンスに基づいて、次年度からは記述語を記したルーブリックを用いるという方法も考えられる。

5. まとめと今後の課題

本稿では、評価練習・自己評価・ピアレビューという 3 つの学生主体の評価活動を組み合わせて授業に取り入れることで、学生の自律的なパフォーマンス改善を促せる可能性があることを先行研究のレビューから論じた。3 つを組み合わせている実践研究も存在するが、なぜ、このように評価活動を組み合わせることがよいのかという理由は十分には検討されていなかった。本稿では、利点と課題を整理したことによって、それぞれの課題を補うためにそれぞれの利点が活用できることを示すことができた。

一方、本稿では、それぞれの評価活動を具体的にどのように実施することが適切かまでは明らかにできなかった。例えば、評価練習において学生が評価する対象となる典型事例を、どのように作成したり選んだりすることが適切かは、重要な論点であり別途の検討を要する。

注

- (1) 石井 (2015) は「子ども」としているが、当該文献は教育段階を限定しない教育評価論を解説した文献といえるため、引用にあたって「学習者」とした。
- (2) Tai et al. (2018) では、2 つ目の教育実践について、Tabel 1 の見出しは Peer feedback /- review と併記されている一方、Abstract では peer assessment となっているほか、本文の該当箇所の節見出しは peer assessment and peer review となっている。3 つの用語は区別されておらず、Tai et al. (2018) は 3 つ全てを重視しているとみられるため、本稿では 3 つを併記した。

付記 本稿は、JSPS 科研費 JP21J15102 の助成を受けた研究によるものである。

引用文献

- 安彦忠彦 (1987) 『自己評価：「自己教育論」を超えて』 図書文化
- Boud, D. (1995) *Enhancing learning through self assessment*. Kogan Page
- Boud, D., & Molloy, E. (2013) Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712
- Cho, K., M. Cho, & D. J. Hacker (2010) Self-Monitoring support for learning to write. *Interactive Learning Environments*, 18 (2), 101–113
- Earl, L. (2003) *Assessment as learning: using classroom assessment to maximise student learning*. Corwin Press
- 藤井美津子 (2018) 教育実習における自己評価と園評価から見る教育実習指導の検討：客観的自己評価力の向上を目指して教職実践演習につなげる. 滋賀文教短期大学紀要, 20, 1–13
- Hart, D. (1994) *Authentic assessment: a handbook for educators*. Addison-Wesley (田中耕治監訳 (2012) 『パフォーマンス評価入門：「真正の評価」論からの提案』 ミネルヴァ書房)
- Hendry, G. D., Armstrong, S., & Bromberger, N. (2012) Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, 37(2), 149–161
- Hendry, G. D., & Anderson, J. (2013) Helping students understand the standards of work expected in an essay: Using exemplars in mathematics pre-service education classes. *Assessment & Evaluation in Higher Education*, 38(6), 754–768
- 石田智敬 (2021a) 評価規準と評価基準 (西岡加名恵・石井英真 (編著) 『教育評価重要用語事典』 明治図書, p.44)
- 石田智敬 (2021b) スタンダード準拠評価論の成立と新たな展開：ロイス・サドラーの所論に焦点を合わせて. カリキュラム研究, 30, 15–28
- 石井英真 (2015) 教育評価の立場. (西岡加名恵・石井英真・田中耕治 (編) 『新しい教育評価入門人を育てる評価のために』 有斐閣, pp.23–49)
- 岩田貴帆 (2020) 協議ワークを取り入れたピアレビューによる学生の自己評価力向上の効果検証. 大学教育学会誌, 42(1), 115–124
- 丸茂美智子・河部房子 (2009) 実習体験に対して看護学生が行った看護場面の自己評価に関する研究：自己教育の観点からの検討. 千葉看護学会会誌, 15, 18–26
- McMahon, T. (2010) Peer feedback in an undergraduate programme: Using action research to overcome students' reluctance to criticise. *Educational Action Research*, 18(2), 273–287
- McConlogue, T. (2012) But is it fair? Developing students' understanding of grading complex written work through peer assessment. *Assessment & Evaluation in Higher Education*, 37(1), 113–123
- Nicol, D., Thomson, A. & Breslin, C. (2014) Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122

- Orsmond, P., Merry, S., & Reiling, K. (2002) The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 27(4), 309–323
- Reinholz, D. (2016) The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301–315
- Rust, C., Price, M. & O'Donovan, B. (2003) Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147–164
- Sadler, D. R. (1987) Specifying and promulgating achievement standards. *Oxford review of education*, 13(2), 191–209
- Sadler, D. R. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144
- Sadler, D. R. (2009) Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179
- Sadler, D. R. (2010) Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550
- Sadler, D. R. (2014) The futility of attempting to codify academic achievement standards. *Higher Education*, 67(3), 273–288
- 下川原久子 (2017) 看護学実習における形成的評価の一考察：統合看護実習の学生による中間評価から。八戸学院短期大学研究紀要, 44, 21–28
- Sridharan, B. & Boud, D. (2019) The effects of peer judgements on teamwork and self-assessment ability in collaborative group work. *Assessment and Evaluation in Higher Education*, 44(6), 894–909
- Tai, J., Ajjawi, R., Boud, D., Dawson, P. & Panadero, E. (2018) Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481
- 田中耕治 (2008) 『教育評価』, 岩波書店.
- To, J. & Carless, D. (2016) Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education*, 40(6), 746–764
- Yucel, R., Bird, F. L., Young, J., & Blanksby, T. (2014) The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assessment & Evaluation in Higher Education*, 39(8), 971–986
- Wiggins, G. (1998) *Educative assessment: designing assessments to inform and improve student performance*. Jossey-Bass.

(日本学術振興会特別研究員 高等教育学コース 博士後期課程3回生)
(受稿 2022 年 8 月 31 日、改稿 2022 年 11 月 7 日、受理 2022 年 12 月 15 日)

学生の自律的なパフォーマンス改善を促す学生主体の評価活動

—初学者対象の大学授業にどのように取り入れるか—

岩田 貴帆

本稿では、近年注目されている、学生の自律的なパフォーマンス改善を促すための学生主体の評価活動について、初学者を対象とする大学の授業にどのように取り入れることが適切か検討した。まず、Tai et al. (2018) の列挙する5つの教育実践のうち、学生主体の評価活動に該当するものを特定した結果、(A) 自己評価、(B) ピアレビュー、(C) 評価練習の3つとなった。この3つの評価活動の定義および利点・課題を先行研究に基づき整理した上で、3つを組み合わせて授業に取り入れることで、それぞれの課題を補い合うためにそれぞれの利点が活用できる可能性について検討した。これに該当する実践を報告している研究から、授業科目全体の実施手順を抽出し、授業設計の参考となるより具体的な示唆を得ることを目指した。

Student-Centered Assessment Activities Implemented to Encourage Students to Improve Their Performance Autonomously: How to Implement These Activities in a University Class for Beginners

IWATA Takaho

This paper examines how university instructors who teach beginning students should incorporate student-centered assessment activities into their classes to encourage students to improve their performance autonomously. First, student-centered assessment activities were identified among the five educational practices enumerated by Tai et al. (2018): (A) self-assessment, (B) peer review, and (C) assessment exercises using exemplars. The definitions, benefits, and shortcomings of each assessment activity were identified based on previous research. It was considered that the benefits of (C) could be used to make up for the shortcomings of (B), the benefits of (A) could be used to make up for the shortcomings of (C), and the benefits of (B) could be used to make up for the shortcomings of (A). Therefore, it is suggested that teachers should implement all three assessment activities in their classes for beginning students. The process of the entire course, including the three assessment activities, is discussed to provide useful suggestions for course design.

キーワード：自己評価、ピアレビュー、評価練習、ルーブリック、大学の授業

Keywords: Self-Assessment, Peer Review, Rubrics, Assessment Exercise Using Exemplars, Class in a University