

情報基礎科目におけるルーブリック自己評価のデータ分析

Data Analysis of Rubric Self-Assessment in Basic Information Courses

深井裕二*

Yuji Fukai

概要

本学の初年次情報基礎科目では、ルーブリックによる自己評価を導入し、情報スキルの習得状況を自ら把握することを学習の一環としている。ルーブリックを用いれば評価の観点と基準が明確化され、自己評価を通じて具体的な学習の振り返りが可能である。本研究では、ルーブリックを用いた情報スキルの客観的な自己評価を行い、自己評価および振り返りに対する主観評価としてアンケート調査を実施した。調査結果ではルーブリック評価による振り返りに対する肯定的傾向が見られ、因子分析を用いたデータ分析によって振り返りに対する有効感について検討した。

1. はじめに

ルーブリック評価方法は学習成果の評価手段として広く普及しており、ペーパー試験などでは測定が難しいとされるスキルや能力に対するパフォーマンス評価において客観的な評価ができる。これを自己評価に用いた場合、評価基準である記述語からは高評価すなわち高いスキルを得るためにどのようなことに着目し習得すべきかを読み取ることができる。つまりルーブリックによる自己評価では、学習者が学習や思考過程に関心を持ち、学習の不足点や取り組み方を認識することで学習意欲の向上へとつながり、学習の振り返り効果が得られる^(1,2)。

本学では学部・学科共通の初年次の基本教育科目として、情報基礎スキルを習得するための「情報処理法」を開講している。本科目ではルーブリック自己評価によって情報スキル習得に対する学習の振り返りを実施しており、これにはルーブリック自己評価システムを開発して導入している。今回、自己評価に対するアンケート調査を実施し、その結果についてデータ分析した。

2. ルーブリック自己評価とそのアンケート調査

本授業で実施する自己評価は情報スキル習得に対するルーブリック評価を行うものである。評価作業にはルーブリック評価用に開発したWeb型のレス

ポンスシステムを用いた。ルーブリックの評価作業は各評価レベルの記述語を読み該当するレベルを判断し選択するものである。この作業を情報システムで実施する場合、大学等で広く使われているMoodleのアンケート機能⁽³⁾など、択一式のアンケート調査ができるシステムであれば利用可能であろう。しかしルーブリック評価では比較的長文の選択肢が並ぶため、可読性が重視されていない画面レイアウトでは評価作業にストレスが生じる。そこでルーブリック自己評価向けのWebシステムを開発し導入した。図1は本システムの操作画面であり、Webブラウザにルーブリックの記述語が表示される。この画面では、Web画面上での長文選択肢の可読性を重視して文字フォントの種類・サイズや段落レイアウト等の画面設計が施されている。

本システムを用いた自己評価の実践として、評価作業を各クラスの終講時期に一斉実施した。評価項目は、(1)PCとWebの基本操作、(2)Word、(3)Excel、(4)PowerPointの4つの情報スキルに関するテーマで構成され、全30項目となる。各評価項目はルーブリック形式によって評価レベルを5段階(5:高~1:低)とする記述語で構成される。自己評価の終了時には各項目を(1)~(4)ごとにまとめたレーダーチャートを画面表示することで評価結果を受講者にフィードバックしている。自己評価システムの利

*北海道科学大学工学部情報工学科

用効果として、受講者はルーブリックを読み評価を判定する際、授業での学習内容を思い出すことで振り返りが行われる。さらにチャートにまとめられた自己スキルの把握から、自分の得意不得意や学習方法とその成果の認知が得られる。このようにして、

その後の学びに良い影響を与えるための自己省察機会を与えることが本実践の狙いである。

今回、自己評価の後で自己評価作業やシステムの操作性、およびルーブリックの妥当性などに関してアンケート調査を実施した。回答はリッカートスケ

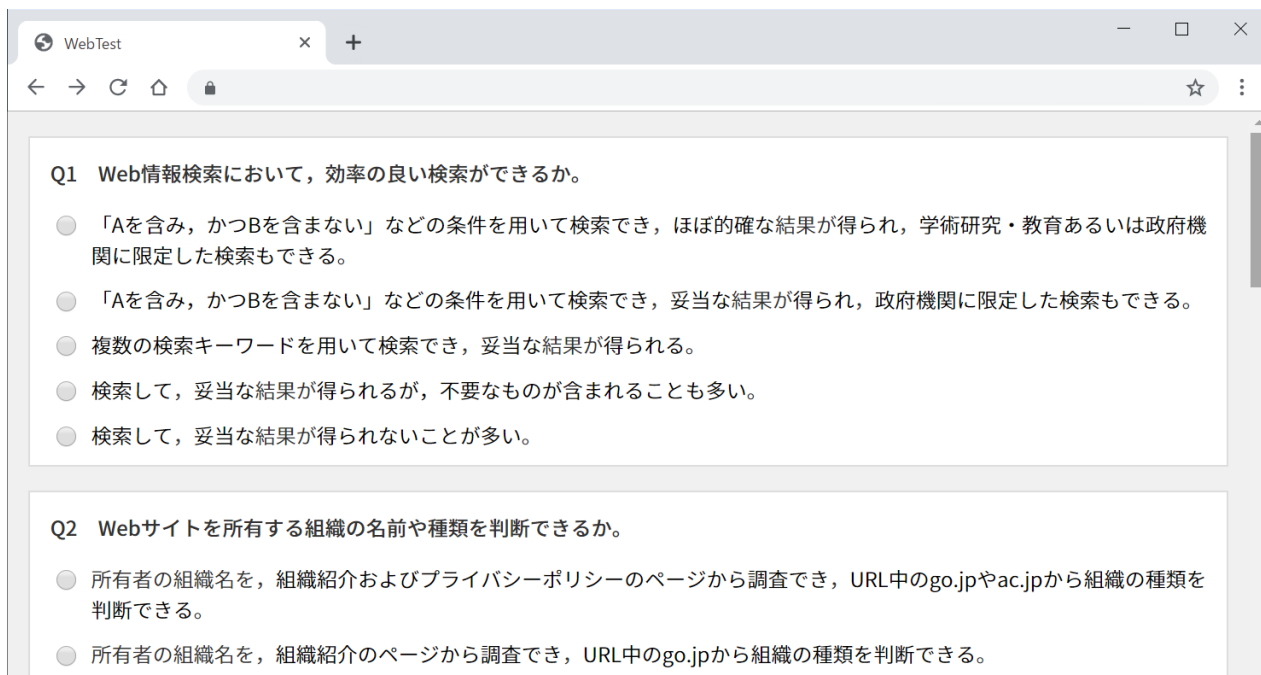


図1 ルーブリック自己評価システムの画面

表1 アンケート質問項目と平均回答値

	質問項目
Q01	選択肢の文章が長すぎた
Q02	選択肢の文章が短すぎた
Q03	選択肢の表現を区別・比較しやすかった
Q04	選択肢が理解しやすかった
Q05	選択肢が読みやすかった
Q06	自己評価で学習の振り返りができた
Q07	自己評価で自己のスキルが確認できた
Q08	自己評価では授業や課題の内容を思い出せた
Q09	自己評価はやりやすかった
Q10	自己評価は今後の学習にとって有意義である
Q11	Moodleのアンケートより見やすかった
Q12	Moodleのアンケートより操作しやすかった
Q13	ルーブリック方式は疲れない
Q14	ルーブリック方式は妥当・適切な評価ができる
Q15	ルーブリック方式は自己のスキル確認に適している
Q16	ルーブリック方式は学習の振り返りに適している
Q17	他の科目でもルーブリック自己評価をやってみたい

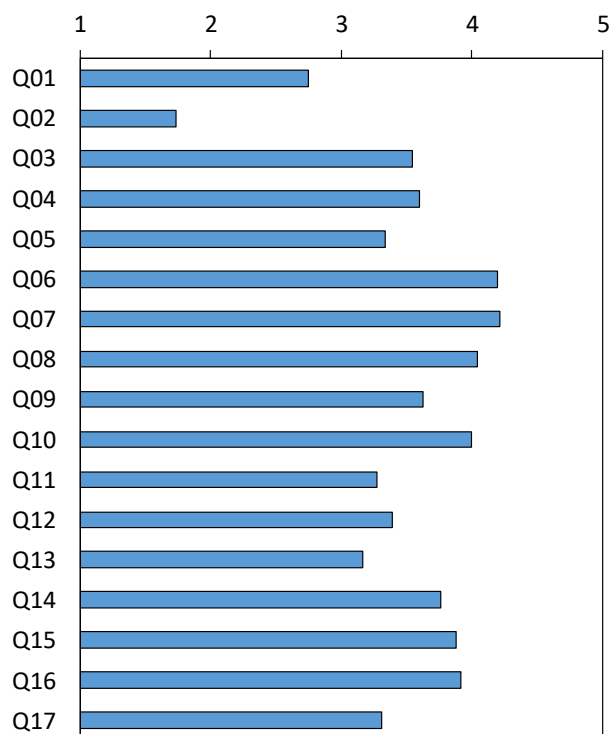
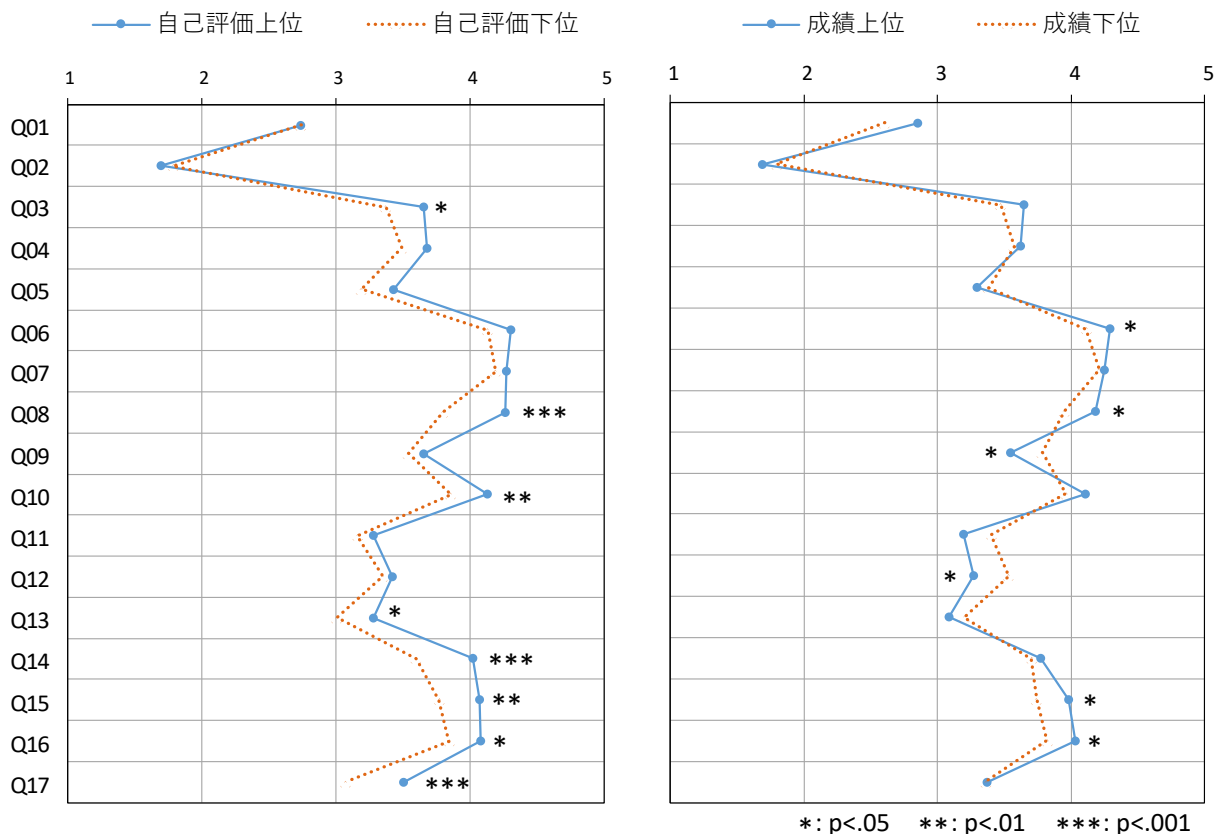


図2 アンケート平均回答値



(a) 自己評価の上位・下位群の比較

(b) 成績の上位・下位群の比較

図3 アンケート回答値の比較

ールの5件法(1:まったくそう思わない, 2:あまりそう思わない, 3:どちらともいえない, ややそう思う, 5:とてもそう思う)とした。各質問項目およびそれらの平均回答値を表1および図2に示す。図2の回答値を見ると、自己評価およびループリックの有効性に関する質問(Q06~10, Q14~16)で肯定的な傾向が強かった。特に自己評価によって振り返りや自己のスキル確認ができたと感じている者(Q06, Q07)は、平均回答値で約4.20と高い。なお、ループリックの文長や表現の適切さ、本システムの操作性などに関し否定的反応は弱いことから、それらが与える自己評価やループリックの有効性の感じ方に対する悪影響は小さいと判断した。

次に、こうした有効性の感じ方における傾向を分析するために、自己評価の平均得点の違いや本科目の成績による比較を試みた。方法として平均自己評価値の順に回答者を均等に3分割し、上位群1/3と下位群1/3とで比較をした。また科目成績についても同様の手順で上位群と下位群を比較した。これらの結果を図3に示す。上位群と下位群間でアンケー

ト質問項目の平均回答値に対し、t検定による有意差が見られたものは、図3(a)の自己評価の上位・下位群間ではQ03, Q08, Q10, Q13~17の主に自己評価およびループリックの有効性に関する項目であり、いずれも自己評価上位群が高かった。また、図3(b)の成績の上位・下位群間では成績上位群が高かったものはQ06, Q08, Q15, Q16と、自己評価上位群ほど多くないが自己評価およびループリックの有効性に関する項目で高かった。反対に成績上位群で低かったものはQ09, Q12であり、自己評価やシステム操作がやりにくかったと感じる様子がうかがえる。

3. 因子分析による傾向分析

さらに自己評価やループリックの有効性に関する傾向を明確化するために、概念が近い質問項目をまとめて全体的傾向の分析を試みた。

本アンケートの質問項目は似た意味のものがある。それらの相関行列を調べたところ、やや相関(0.7>r≥0.4)のあるものが9.6%、弱い相関(0.4

> $r \geq 0.2$)があるものが39.0%であった。そこで複数の説明変数に影響を与えている共通因子を抽出する分析手法である因子分析を行った。まず初期段階の因子分析の対象を質問Q01~Q17の17項目とし、それらの固有値ベクトルを求めた。因子数は固有値1.0以上かつスクリープロットの減衰状況を見て決定した。抽出法および回転法として最小残差法およびプロマックス回転を用い、因子負荷量0.4未満の項目を削除しながら因子分析を反復した。因子分析の最終段階において因子数は4、削除した項目はQ01, Q02, Q13, Q16, Q17であり、12項目が残った。

図4および表2に、因子分析の最終的なスクリープロットおよび因子負荷量を示す。図4をもとに固

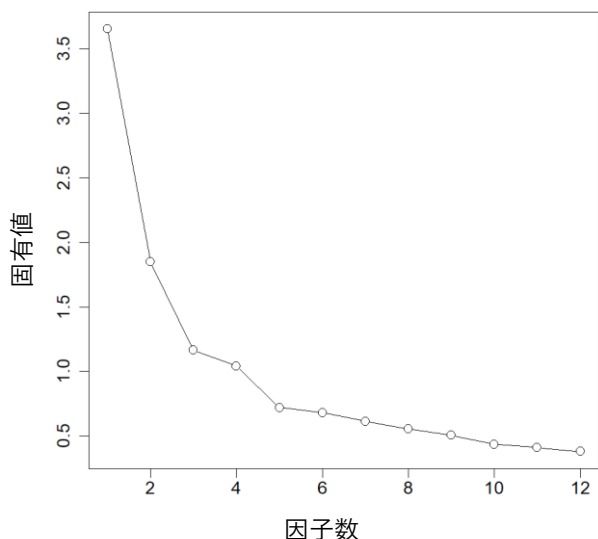
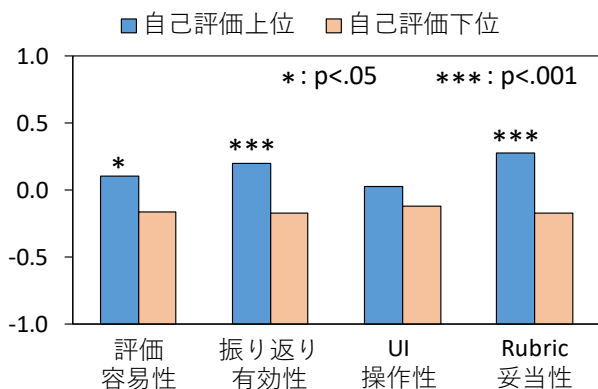


図4 スクリープロット

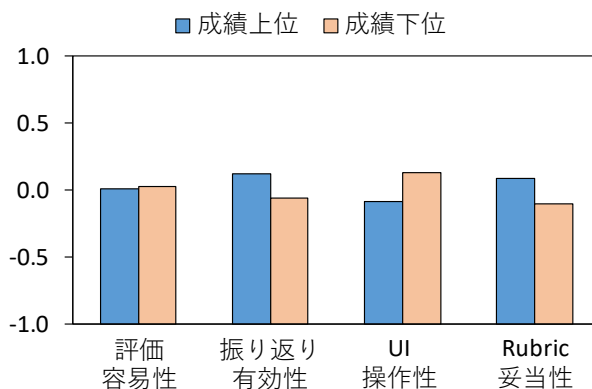
有値 ≥ 1.0 となるように因子数は4を採用した(累積寄与率=48.2%)。また表2において各因子に対し質問項目の内容から因子名を命名した。因子1は「選択肢が理解しやすい」「選択肢が読みやすい」「選択肢の表現を区別・比較しやすい」「自己評価はやりやすい」という質問に対する負荷量が高いことから「評価容易性」とした。因子2は「自己評価で学習の振り返りができた」「自己評価では授業や課題の内容を思い出せた」「自己評価で自己のスキルが確認できた」「自己評価は今後の学習にとって有意義

表2 因子負荷量

因子名	変数	因子1	因子2	因子3	因子4
因子1 評価容易性	Q04	0.807	-0.137	-0.111	0.108
	Q05	0.715	-0.024	-0.001	-0.021
	Q03	0.694	-0.003	0.100	-0.106
	Q09	0.541	0.203	0.060	-0.013
因子2 振り返り有効性	Q06	-0.062	0.785	-0.063	-0.066
	Q08	-0.036	0.562	0.044	-0.065
	Q07	-0.014	0.512	-0.001	0.079
	Q10	0.099	0.505	-0.034	0.116
因子3 UI操作性	Q11	-0.004	-0.071	0.912	0.028
	Q12	0.011	0.019	0.604	0.015
因子4 Rubric妥当性	Q14	-0.003	-0.035	-0.003	0.785
	Q15	-0.020	0.041	0.044	0.688
固有値		1.939	1.479	1.227	1.142
寄与率 (%)		16.2	12.3	10.2	9.5
累積寄与率 (%)		16.2	28.5	38.7	48.2



(a) 自己評価の上位・下位群の比較



(b) 成績の上位・下位群の比較

図5 因子得点の比較

である」という質問に対する負荷量が高いことから「振り返り有効性」とした。因子3は「Moodleのアンケートより見やすかった」「Moodleのアンケートより操作しやすかった」という質問に対する負荷量が高いことから「UI（ユーザインタフェース）操作性」とした。因子4は「ルーブリック方式は妥当・適切な評価ができる」「ルーブリック方式は自己のスキル確認に適している」という質問に対する負荷量が高いことから「Rubric 妥当性」とした。

得られた各因子において因子得点を求め、再び自己評価の上位群と下位群および成績の上位群と下位群を比較してみた。図5(a)は自己評価における、また図5(b)は成績における上位群と下位群の因子得点の平均を比較したものである。自己評価では「評価容易性」($p < .05$)、「振り返り有効性」($p < .001$)、「Rubric 妥当性」($p < .001$)にt検定による有意差が見られ、いずれも上位群の方が高いことが示唆された。よって自己評価の上位群は下位群に対し、因子1, 2および4の影響度が高いため、ルーブリック評価による振り返りに対する有効感が高いと考えられる。一方、成績の違いでは有意差が見られなかったため、他者評価による能力の違いは有効感にあまり影響しなかったと考えられる。

4. 考察と展望

以上の分析で得られた傾向から、全体的には自己評価やルーブリック評価に対して概ね肯定的であったが、因子分析の結果からはルーブリック評価による振り返りに対する有効感や自己評価の得点と関係性があることがわかった。この関係性により有効感に影響が出た原因について次の仮説を考えた。

- (1) ルーブリック評価と自覚する潜在的能力評価との観点の一致性が学習者によって異なった。
- (2) ルーブリックの各評価観点におけるレベルの基準設定が学習者によっては高難度だった。
- (3) ルーブリック評価には得意不得意があり、学習者によって評価の的確さに違いが生じることで評価結果の妥当性の感じ方が異なった。

仮説の(1)では、自覚する潜在的能力評価とは「自分はこういう事柄がこれくらいできる」と自覚している際、この事柄とルーブリック評価観点・基準が乖離していれば今回の評価観点・基準に当てはまりづらくなり自己評価値が低下する可能性がある。こうした状況が多ければ、ルーブリック評価による振り返りに対する有効感や自己評価は低下するものと思われる。

(2)では例えば学習者自身は中程度のレベルを想像していたが、実際に評価してみると中程度であるレベル3ではなく下位のレベル2であったというように、学習者にとってレベル設定が全体的に高難度であったり、隣接レベル間の難度的間隔が大きかったりなど、評価設定に対する不満や不的確な評価があった可能性が考えられる。(3)では文章に対する判読や理解が求められる状況で、ルーブリック評価の不慣れによる不的確な評価、あるいは学習者の無意識な過小自己評価傾向などは、評価が不的確となり総括的な自己評価結果を見たときに不満を感じ有効感が低下する可能性が考えられる。

(1)への対策として、ルーブリックを学習前に学習者に示すことで、学ぶべきスキルおよび評価基準を理解させ印象付けることで、ルーブリック評価と自覚する潜在的能力評価の一致性を高めることが挙げられる。(2)への対策として長期的実践を通じ、評価レベル設定を実態に応じて調整することが挙げられる。(3)に対しては、事前にルーブリック自己評価の体験機会を設けて慣れさせておくことが挙げられる。仮説のいずれの場合でも、ルーブリック評価による振り返りに対する有効感や自己評価の満足感が低いと、モチベーションの面などで振り返りの効果が低減する恐れがあるため、こうした問題をできるだけ解消することが望ましい。例えば、ルーブリック自己評価と自分の自由な評価基準による自己評価を併用し、複眼的に総合評価する方法が挙げられる。

自己評価の実践にあたり、過大および過小な自己評価に関する心理的傾向も考慮すべきであろう。Kruger & Dunningによると、能力等の自己評価では特に成績の低い人には顕著に過大評価する傾向にあるという認知バイアス(cognitive bias)があることが明らかにされており、Dunning=Kruger効果として知られている⁽⁴⁾。また中高生による英語能力の自己評価の研究では、成績上位群(上位25%)は自己評価で顕著に過小評価するという傾向が報告されている⁽⁵⁾。このような認知バイアスが自己評価による振り返り学習にどのように影響するののかも重要である。特に客観的評価に優れたルーブリック評価を用いた場合、こうした影響がどの程度抑えられるかを調べることで自己評価におけるルーブリックの有利性を検証できるものと思われる。

またルーブリック自己評価に関して、アカデミック・ライティング能力のルーブリックでは教員によ

る他者評価と学習者による自己評価には大きなズレがあることが明らかにされている⁽⁶⁾。さらにルーブリックにおける教員と学習者の評価のズレに関する振り返りの研究によると、ズレを学生に提示することで振り返りが促進されることが報告されている⁽⁷⁾。これらのようなルーブリック自己評価で生ずる他者評価や成績との乖離については、それを問題視する一方で振り返り学習の効果やルーブリック評価の判定力を高める学習材料としての活用法にも目を向けたい。今回得られた結果から、自己評価手法としてのルーブリックと振り返り効果に対し、さらなる試みと調査によってそれらの関係性を明らかにしていく必要がある。

5. まとめ

本研究では、情報スキル教育においてルーブリック自己評価を用いた振り返り学習を実践した。それに対するアンケート調査結果から、本実践における自己評価、ルーブリックおよび振り返りに対する有効性が肯定的に捉えられていることが確認できた。これは本実践が振り返り学習の効果を得る基盤として機能したことや、ルーブリック自己評価が学習に有効であるという実感を与えられる可能性があることを意味している。ルーブリック自己評価は大学4年間と Semester 制による学習機会の大きなボリュームを活用し、学びを次の学びに活かす自己省察の効果的手段として注目できる。また、教育者と学習者の両者に役立つアセスメント手法として有用であろう。今回の分析結果では自己評価点の上位群は特にルーブリック評価による振り返りに対する有効感が高かった。このような傾向に対し様々な実践方法によって調査し、ルーブリック自己評価による振り返り学習の効果をさらに高める方法について検討してみたい。

参考文献

- (1) 原田三千代：内省型ルーブリックによる対話的評価活動の分析，三重大学教育学部研究紀要，自然科学・人文科学・社会科学・教育科学・教育実践，Vol. 68，pp. 317-332，2017.
- (2) 大作光子，嶺坂尚：ルーブリックによる自己評価の分析を通じた探究学習の支援のあり方，日本図書館情報学会誌，Vol. 61，No. 4，pp. 232-251，2015.
- (3) Churchward, M.: Questionnaire module, http://docs.moodle.org/19/en/Questionnaire_module, 参照日：2019-9-11
- (4) Kruger, J., Dunning, D.: Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments, *Journal of Personality and Social Psychology*, Vol.77, pp.1121-1134, 1999.
- (5) 西村洋一：中高生は英語の能力についてどのくらい正確に自己評価をしているか？－学習動機づけと事後の成績との関連も踏まえた検討，北陸学院大学・北陸学院大学短期大学部研究紀要，Vol.11，pp. 89-100，2018.
- (6) 斎藤有吾，小野和宏，松下佳代：パフォーマンス評価における教員の評価と学生の自己評価・学生調査との関連，日本教育工学会論文誌 Vol. 40 (suppl), pp. 157-160, 2017.
- (7) 斎藤有吾，小野和宏，松下佳代：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，Vol. 39，No. 2，pp. 48-57，2017.