

ルーブリックの法学部教育への展開Ⅱ

—研究発表大会における2回目の実践報告—

喜 入 暁
山 本 健 人
大 島 一 悟

要旨

本稿は実践報告として、大阪経済法科大学で行われた2021年度法学部研究発表大会におけるルーブリック評価の結果を報告した。このルーブリックは、石上・喜入・大島（2020）によって開発されたものであり、法学部のみならず一般的な研究発表の評価基準として応用可能であると考えられるものである。しかし、使用した結果、教員評価者間の一致率は低く、ルーブリックによらない単一評価との相関も低かった。また、学生評価者は、一般的に適切とは言えない程度に高い評価をつけていた。今後、日常的な発表機会において継続的にルーブリックを使用し、改善点や問題点を修正したうえで、本質的な効果検証を行う必要があると考えられる。

はじめに

高等教育の評価方法が見直される中で、ルーブリックが注目を集めている（山口，2013）。これまでに多くのルーブリックが開発され、それらのデータベース化もなされている（日本高等教育開発協会，2022）。山口（2013）の定義に基づけば、ルーブリックとは、「いくつかの段階に分けて教育上の達成度の目安を記述して、学習者の達成度を判断する基準を示したもの」とされる。また、その目的の一つは、これまで評価の問題点とされてきた判断基準の非公表や評価者と被評価者間の共通認識の欠如へ対処し（田中，

2004),「成績評価の厳格化」を促進するために、評価基準を明確化することである(田宮,2014)。ルーブリックによる評価が有効である可能性は繰り返し議論されている(たとえば,山口,2013;田宮,2014)。それは、ルーブリックにより評価基準を明確化・具体化することで、教育者と学習者の学修内容や学習段階の共通認識を促し学習者の改善点を具体化することで学習の方向性を明確にすることが可能となるからである(田中,2008)。

これらは特に、これまでテスト法では定量的な評価が困難だった「思考・判断」、「関心・意欲・態度」、「技能・表現」などのパフォーマンスの評価、具体的には、作品や活動、レポートなどの評価に有効であるとされる(沖,2014)。パフォーマンスを定量的に測定し評価を可能にすることは、よりシステマティックな教育プログラムに基づく教育の質向上や教育効果を高めることに寄与すると考えられる。さらに、中等教育においてルーブリックの活用が積極的な学修姿勢の向上に寄与したり(加藤・藤田,2019;鈴木,2011a,b)、高等教育において学生自身がルーブリックを作成することで学習意欲が高められる(遠海・岸磨・久保田,2012;碓山・木村,2017)といった知見が報告されている。また、試験による到達度の測定が比較的容易な数学科目においても用いられ、学修意欲の継続に寄与していることが示唆されている(高木・林・二上,2019)。

本稿では、法学教育におけるルーブリックの実践例を報告する。なお、ここで使用するルーブリックは著者らが作成したものである(石上・喜入・大島,2021;以降、「第一報」と表記)。法学教育の領域におけるルーブリックの実践報告は、第一報の他にもなされているが(山下,2018;吉永,2019)、いずれも法学教育、特に「法解釈」に特化している。一方、第一報で報告したルーブリックの使用目的は法学部研究発表大会のための評価基準である。同大会では、「法解釈」に限定されない法学に関する自由な方法論に依拠した発表がなされるだけでなく、政治学や社会学など多岐にわたる領域にかかわる社会問題を扱う発表がなされる。そのため、第一報で作成されたルーブリックは、特定の学修領域の評価ではなく、より一般的な研究発表の評価基準であることを念頭におきつつ、法学部としてのエッセンスを加えたものである。つまり、評価基準そのものは他の学問領域の研究

発表の評価にも応用可能であると考えられる。

第一報と同様、本稿で報告されるのは、法学部研究発表大会におけるルーブリックによる評価の実践例である。第一報と同様に、ルーブリック評価は研究発表大会でのみ用いられており、日常的な演習などでは用いられていない。研究発表大会は2021年11月26日に行われ、今回の研究発表大会に使用するルーブリックの確定版が公開されたのは2021年10月13日であった。また、Microsoft Formsを用いた学生による評価フォーマットが公開されたのは同年11月23日であった。なお、暫定版のルーブリックの公開は、発表エントリー登録開始日である2021年6月15日であったが、確定版のルーブリックと同様の内容であった。したがって、学生が確定版のルーブリックを確認した期間は最大で45日間であった。ただし、熟読や訓練をするような要請はしていない。

方 法

デー タ

研究発表大会は、5会場で行われ、各会場では同一の3名の教員が評価にあたった。各会場では5発表（ただし1会場のみ6発表）が割り当てられた。したがって、各発表は3名の教員によって評価されており、3名の教員は各々5発表（1会場のみ6発表）を評価した。そのため、各評価がレベル1変数、評価者と発表がレベル2変数の階層構造のデータが得られた。ただし、分析では評価者の効果をプールし、発表のみレベル2変数として扱った。

ルーブリックの各項目およびその合計得点を各発表の評価得点として扱った。ただし、減点項目は各発表における3名の評価者の中で最も高かった減点を各評価得点から一律に減じた。

なお、学生評価では、各学生が任意の発表にオーディエンスとして参加し評価した。複数の発表に参加し評価することは可能であったが、すべての学生は単一の発表のみを評価した。評価した学生は43名であったが、後述するトラップ項目に適切に回答した学生は18名だった。

測定変数

ルーブリックの評価項目は、問題提起（4点満点；うち問題の設定が1点満点、背景が3点満点）、調査～インプット（6点満点；うち情報源が3点満点、視点が3点満点）、分析・主張～アウトプット（10点満点；うち主張の有無が1点満点、論理性が3点満点、批判的検証が3点満点、残された課題が3点満点）、発表技術（6点満点；うちプレゼンスキルが3点満点、プレゼン資料が3点満点）に加え、減点項目（0-3点）であった。また、ルーブリックに依存しない評定との関連性を検証するために、同時に総合点（1-5点）の評価も行った。なお、評価に用いたルーブリックは表1の通りであった。なお、学生の評価では、評価告目を理解せずにチェックをつける可能性がある（努力の最小限化；三浦・小林，2015参照）。そこで、評価項目を理解したうえで評価しているかどうかを検出するために、「残された課題」と「プレゼンスキル」の間に「【聞き手】この項目では何も考えずに「その他」を選択してください」というトラップ項目を設定し、その選択肢を「発表者は話を十分に聞いているという資料が示されている（問題文をきちんと読んでいる）」、「発表者は話を十分に聞いているという資料が示されているものの、視点が定まっていない。【※どちらかといえば、十分】」、「発表者は話を十分に聞いているという資料が示されているものの、視点が定まっていない。【※どちらかといえば、不十分】」、「その他」とした。

手続き

各発表の発表時間は10分であり、質疑応答が5分であった。教員の評価はルーブリック評価表に発表毎にチェックをつけるやり方で実施した。発表中のいずれの段階で行うかは決められていなかったため、発表を通して評価が行われ、質疑応答を終え次の発表が始まるまでには評価を完了した。学生の評価は、Microsoft Formsを用いて評価項目を選択する形で実施した。回答期限を設けなかったため、必ずしも発表後即時の回答ではなかった。なお、本稿の分析には当日中の回答のみをデータとして用いた。

表 1 ループリックの内容

		満点	減点 (2～1点)	0点
問題提起 (4)	問題提起 (1)	問題提起が曖昧ではなく明確に <u>なされている</u> 。		問題提起がなされていない、または曖昧である。
	背景 (3)	問題提起に至った背景が、主観的または曖昧でなく、問題設定と乖離していない (問題提起が必然的に導かれる)。	問題提起に至った背景が、主観的または曖昧でないが、問題提起との乖離がある。	問題提起に至った背景が、主観的である、または曖昧である (一貫性がない場合も含む)。
調査～インプット (6)	情報源 (3)	発表者の主張以外の情報 (判例・学説・事実・事例) のすべてに、情報源 (参考文献など) が明示されている。 ※ Wikipedia や個人ブログなど、信ぴょう性に乏しい情報は評価対象としない (情報源が示されていないものとして扱う)。	発表者の主張以外の情報 (判例・学説・事実・事例) のうち一部に、情報源 (参考文献など) が示されていない。	発表者の主張以外の情報 (判例・学説・事実・事例) のすべてに、情報源 (参考文献など) が示されていない。
	視点 (3)	複数の視点から引用した情報のすべてが、論の構成に寄与している (情報同士のつながりが明確である / いずれの情報も欠けても不十分な議論になり得る)。	情報が複数の視点から引用されているが、不要情報が含まれている、または情報のつながりが曖昧である。	情報が単一の視点にのみ基づいている。または複数の視点から引用されているが、選別にとどまっている。
分析・主張～アウトプット (10)	主張 (1)	問題提起と対応した主張 / 結論が、 <u>曖昧ではなく明確に示されている</u> 。		主張 / 結論が示されていない、または曖昧である。 主張 / 結論が問題提起と対応していない。
	論理性 (3)	主張 / 結論が、論理の飛躍や強引な結論導出がなく、 <u>明らかな根拠 (事実 / 事例を含む) を基に論理的に導かれている</u> 。	主張 / 結論が、根拠 (事実 / 事例を含む) を伴った形で示されているが、論理の飛躍や強引な結論導出が見られる (根拠が弱いことも含む)。	主張 / 結論の根拠 (事実 / 事例を含む) が示されていない。
	批判的検証 (3)	異論 / 反論を的確に取り込んでおり、それらに対して説得力のある応答がなされている。	異論 / 反論に触れられているが、正しくとらえていない (異論 / 反論として不適切である場合も含む)。 異論 / 反論に対する応答が説得力を欠く。	異論 / 反論に触れられていない。
	残された課題 (3)	残された課題が的確に挙げられている (今回の研究で、何ができ、何ができなかったのかを、明示できている)。	残された課題に触れられているが、正しくとらえていない (的外れである場合も含む)。	残された課題に触れられていない。
	プレゼンスキル (3)	聴衆を意識した、聞き取りやすいプレゼンである。	声の大きさとスピードのいずれかに難がある。 原稿を読み上げることに集中しすぎ、聴衆に意識を向けられていない。	声の大きさとスピードの双方に難がある。
	発表技術 (6)	プレゼン資料 (3)	パワーポイントの特性を活かした資料になっている。(文字数は必要最低限、図・写真など視覚的理解を促す工夫) (多すぎる文字数、アニメーションの多用など)	パワーポイントの特性を活かす工夫が見られない。(図がない、文章をそのまま貼り付け、字が小さいなど)
	その他 (減点事由)	発表時間が 14 分 30 秒 - 15 分 30 秒でない。(± 30 秒につき 1 点減点、ただし 16 分以上は強制終了し 2 点減点) 誤字脱字 (4 個以上の場合、マイナス 1 点)		
総合点 (1～5 点で回答)		質疑応答を踏まえ、総合的に優れた研究と評価できるか※以下は判断基準の例です。各観点 ^① を踏まえた上で、総合的に評価してください。 ・国民主権、権力分立、基本的人権の尊重など基本的事項を踏まえているか ・様々な価値観や法治主義、適正手続、罪刑法定主義、所有権の保護、契約自由等の法原理を理解しているか ・法的判断の持つ特殊性を踏まえ、「説得」の重要性を理解しているか ・様々な考え方の分析、コンセンサスを得るための調整がなされているか ・グローバル化に伴う法的問題と文化の多様性を理解しているか		

結果

記述統計量

表2にループリック評価項目の平均値を評価者（教員・学生）ごとに示した。なお、学生評価は43名の評価者のうち、トラップ項目に適切に回答した18名（学生評価者の41.9%）の評価のみを用いた。全体的に教員評価に比べて学生評価の方が高い。ただし、調査～インプットにかかわる項目とプレゼンスキルにかかわる項目には統計的な差は示されず、分析・主張～アウトプットに教員と学生の乖離が示された。なお、問題提起と主張、すなわち、発表としての体面において必然的に必要となる項目は、教員評価、学生評価いずれにおいても満たしていると判断される発表が大部分を占めた。

表2 ループリック評価項目の評価者（教員・学生）別の平均値と標準偏差

評価観点	教員 (N=75) ^a		学生 (N=18)		t (Welch)	d
	平均値	標準偏差	平均値	標準偏差		
問題提起 ^b	68	90.7%	15	83.3%	0.26	.053
背景	2.17	0.98	2.61	0.78	-2.03	-0.46
情報源	2.25	0.79	2.39	0.85	-0.62	-0.17
視点	2.11	0.86	2.67	0.69	-2.95 ^{**}	-0.67
主張 ^b	58	77.3%	17	94.4%	0.31	.058
論理性	1.89	0.81	2.50	0.92	-2.56 [*]	-0.72
批判的検証	1.39	0.96	2.33	0.97	-3.73 ^{**}	-0.98
残された課題	1.32	1.15	2.39	0.85	-4.44 ^{**}	-0.96
プレゼンスキル	2.29	0.77	2.67	0.69	-2.02	-0.49
プレゼン資料	2.33	0.72	2.61	0.70	-1.51	-0.38
ループリック評価得点	16.56	4.95	21.94	5.41	-3.85 ^{**}	-1.06
総合点	2.75	1.00	3.67	0.97	-3.59 ^{**}	-0.92

^a 延べ人数である。本文に記載通り、一人の教員が複数の発表を評価したが、ここでは各々独立した評価者として扱った。

^b 得点レンジが0～1であったため、平均値の列には1と評価された発表数を、標準偏差の列には1と評価された割合を記載した。また、検定統計量(t)の列には χ^2 乗値を、効果量(d)の列にはクラメルVを記載した。なお、その他の評価観点の得点レンジは0～3であり、そのトータル得点であるループリック評価得点の得点レンジは0～26であった。また、総合点の得点レンジは1～5であった。

* $p < .05$, ** $p < .01$

教員評価の一致率とループリック評価得点と総合得点との相関

ループリックの各評価観点に対する教員同士の一致率を検証するために、各発表における教員評価の級内相関係数（Intra-rater correlations: ICCs）、教員評価者における項目間の相関係数、学生評価者における項目間の相関係数を算出した（表3）。ICCは評価の一致率として用いられる指標であり、 $-1 \sim 1$ の値を取る。0の場合は完全な不一致、1の場合には完全に一致していると解釈する。なお、負の値を取る場合は一致の方向性が逆方向であるような場合（たとえば、一方が最高得点、他方が最低得点をつけるような場合）である。「一致」と解釈する基準は明確ではないが、おおむね0.7以上の場合に一致した評価であると判断するようである（たとえば、村澤, 2007; 対馬, 2010）。今回、ICCはすべての項目で0.7を下回り、総合点を除いて0.3にすら至っていない。したがって、今回のループリックの評価による一致率は極めて低いといえる。

表3 ループリックの評価項目間の相関係数と各発表に対する評価の一致率（ICC）

	1	2	3	4	5	6	7	8	9	10	11	12
1. 問題提起 ^a	.07	.84	.41	.61	.54	.41	.60	.41	.61	.58	.69	.74
2. 背景	.39	-.10	.33	.40	.36	.45	.49	.24	.51	.79	.60	.68
3. 情報源	.16	.26	.08	.84	.57	.79	.76	.76	.24	.27	.59	.82
4. 視点	.04	.33	.51	.09	.67	.74	.80	.74	.50	.20	.62	.87
5. 主張 ^a	.15	.23	.13	.33	.10	.58	.63	.57	.22	.18	.32	.68
6. 論理性	.07	.33	.29	.57	.36	.06	.79	.79	.28	.50	.59	.88
7. 批判的検証	.28	.38	.24	.34	.39	.47	.25	.90	.44	.29	.69	.91
8. 残された課題	.17	.33	.30	.29	.23	.32	.38	.17	.24	.17	.59	.81
9. プレゼンスキル	.06	.29	.19	.18	.04	.14	.03	.23	.30	.33	.53	.55
10. プレゼン資料	-.11	.26	.30	.14	-.01	.15	.16	.10	.58	.14	.49	.56
11. ループリック評価得点	.10	.11	.30	.28	.25	.33	.29	.22	.43	.32	.29	.76
12. 総合点	.30	.64	.59	.68	.42	.68	.63	.63	.48	.47	.47	.50

^a 2値変数である。

Note. 対角行列は級内相関係数（ICC）、上三角行列は学生評価者（N = 18）における評価項目間の相関係数、下三角行列は教員評価者（N = 75; ただし述ベサンプルサイズ）における評価項目間の相関係数を示す。なお、教員評価は階層データであるが、ここでは階層をプールし各々独立したデータとして扱った。相関係数の有意性検定はしなかった。

ループリック評価得点が総合得点とどの程度一致するのかどうかを検証するために、相関係数を算出した。また、教員評価と学生評価の散布図を

る必要があるだろう。さらに、評価項目数も多かった。これらのことは、即時的な評価の困難さにつながる要因である。逆に言えば、前大会や本大会での評価の一致率が低かった原因が、評価項目の不適切さではなく単に即時的な評価を強いられたことによる弊害であった可能性が考えられる。したがって、今後の日常的な発表機会において積極的に本大会でのループリック用い、評価に慣れた上で再度検証する必要があると考えられる。

第 2 に、ループリック評価得点と総合点の相関は、一般的には中程度であると考えられるものの（水本・竹内，2008）、この 2 つの変数間には、理論的には同一の測定をしていると考えられるため強い相関が予測されるものである。その予測を踏まえると、ループリック評価得点と総合点の相関関係は想定される関連性よりも弱いといえる。この点について、先に述べた即時的なループリック評価の困難さが原因の一つとして挙げられるだろう。しかし、その他の原因として、ループリック評価の基準として設定されておらず、かつ、評価に重要な要素が影響した可能性が考えられる。すなわち、本来測定すべき要素を、ループリックがとらえきれていない可能性が考えられる。たとえば、個別法専門家のディシプリンによる評価のばらつきが生じたかもしれない（ただし、今回のループリックは使用の一般化が目的であったため、慎重に考慮する必要があるだろう）。そのため、そのような要素を明らかにし、ループリック評価項目に含める必要があるかもしれない。ただし、現状、その要素は不明である。したがって、今後の日常的な発表機会を通してこのループリックを使用し、ループリックそのものを批判的に扱い、とらえきれていない要素を浮き彫りにすることが必要であると考えられる。

第 3 に、全般的に学生の評価は教員評価よりも高かった。学生の評価は一般的に甘くなりがちである可能性が示されている（諸根他，2016；安原他，2012）。しかし、正当な他者評価は正当な自己評価やメタ認知の正確さとリンクするため、教育的観点からも、他者の発表に対して適切に評価する力を養成する必要があると考えられる。他者評価はアカデミックスキルなどのように訓練が必要であると考えられる。したがって、今後の日常的な発表機会において、教員だけでなく学生自身が相互評価のためにループ

研究ノート

リックを用いた訓練が必要であると考えられる。ただし、本大会で発表を評価した43名のうち、トラップ項目に適切な回答をした学生は半数に満たなかった。この点については、評価のスキルなどではなく、評価や発表を聴くことについてのモチベーションが大きく影響する可能性がある。したがって、モチベーションを含む発表を聴くことへの態度も同時に養う必要があり、そのための工夫を取り入れれば授業設計が必要になるかもしれない。

リミテイション

本稿ではあくまで実践報告にとどまり、ループリックの使用による心的変化を含む効果検証には至らなかった。なぜなら、昨年度と同様、今大会でも、ループリックの周知やその評価基準の公開が大会の1か月前ほどであり、教員、学生とも適切に使用することが困難な状態であったためである。そのため、次の研究発表大会に向けて長期的に使用していくことが必要だろう。その上で、ループリック評価項目の本質的な問題点を洗い出し、修正する必要があると考えられる。

また、学生評価者が少なかったことも、先に進めるための分析を困難にした原因の一つであると考えられる。先の項でも述べたが、発表しない学生であってもオーディエンスとして積極的に参加することができる環境を整える必要があるとともに、日常的にループリックを使用してもらい、評価項目を批判的に捉えなおすことができる環境を形成する必要があるだろう。

まとめると、ループリックを日常的に使用することがまず重要である。その上で、現行のループリックの問題点・改善点を洗い出し、修正を加えるとともに、ループリックを使用することの学修効果の検証およびモチベーションの維持といった心理的効果の検証を進める必要があるだろう。

引用文献

- 遠海友紀・岸磨貴子・久保田賢一（2012）. 初年次教育における自律的な学習を促すループリックの活用. 日本教育工学会論文誌, 36, 209-212.
- 碓山恵子・木村尚仁（2017）. 学生の協働意識を引きだす学習者主体のループリッ

- ク作成と自己評価の試み. 北海道科学大学研究紀要, 43, 35-41.
- 石上敬子・喜入暁・大島一悟 (2021). ルーブリックの法学部教育への展開——法学部学生研究発表大会における実践例と統計的検証——. 大阪経済法科大学法学論集, 84, (1)-(26).
- 加藤みずき・藤田哲也 (2019). 初年次教育科目における発表のルーブリックに対する学生の受け止め方. 経営・情報研究 多摩大学研究紀要, 23, 91-100.
- 三浦麻子・小林哲郎 (2015). オンライン調査モニタの Satisfice はいかに実証的知見を毀損するか. 社会心理学研究, 31, 120-127.
- 水本篤・竹内理 (2008). 研究論文における効果量の報告のために——基礎的概念と注意点——. 英語教育研究, 31, 57-66.
- 諸根美恵子・佐藤厚子・大野勲・大河原雄一・鈴木常義・中村仁・東裕 (2016). ピア評価と自己評価によるプレゼンテーション能力の測定の試み. YAKUGAKU ZASSHI, 136 (7), 1041-1049.
- 村澤昌崇 (2006). 高等教育研究における計量分析手法の応用(その2)——組織内合意形成の程度分析: ハーフインダール係数, rwg, 級内相関係数. 広島大学高等教育研究開発センター大学論集, 38, 353-367.
- 日本高等教育開発協会 (2022). ルーブリックバンク 日本高等教育開発協会 Retrieved from <https://www.jaedweb.org/rubricbank> (2022年1月17日).
- 沖裕貴 (2014). 大学におけるルーブリック評価導入の実際——公平で客観的かつ厳格な成績評価を目指して——. 立命館高等教育研究, 14, 71-90.
- 鈴木雅之 (2011a). ルーブリックの提示が学習者に及ぼす影響のメカニズムと具体的事例の効果の検討. 日本教育工学会論文誌, 35, 279-287.
- 鈴木雅之 (2011b). ルーブリックの提示による評価基準・評価目的の指示が学習者に及ぼす影響. 教育心理学研究, 59 (2), 131-143.
- 高木悟・林康弘・二上武生 (2019). フルオンデマンド型全学数学基盤教育におけるルーブリックの活用. 教育システム情報学会研究報告, 34, 109-112.
- 田宮憲 (2014). ルーブリックの意義とその導入・活用. 高等教育開発センターフォーラム, 1, 125-135.
- 田中耕治 (2008). 学力調査と教育評価研究. 教育学研究, 75, 146-156.
- 山口陽弘 (2013). 教育評価におけるルーブリック作成のためのいくつかのヒントの提案. 群馬大学教育学部紀要 人文・社会科学編, 62, 157-168.
- 山下純司 (2018). ルーブリックの法学教育への応用. 学習院大学法学会雑誌, 53, 103-115.
- 安原智久・山口貴史・曾根知道・柳田一夫・中村三孝・栗尾和佐子・西川智絵・山本祐実・前田定秋 (2012). Small group discussion (SGD) へのピア評価の導入と総括的評価としての妥当性. YAKUGAKU ZASSHI, 132, 1179-1188.

研 究 ノ ー ト

吉永一行 (2019). 学生の発表を中心にした大講義科目の運営手法——「橋本メソッド」の一応用例として——. 高等教育フォーラム, 9, 113-118.