

実践報告

プレFD科目のマイクロティーチング用ルーブリックの開発

長沼 祥太郎¹・鄭 漢模¹¹九州大学教育改革推進本部

本研究の目的は、プレFD科目におけるマイクロティーチングを評価するためのルーブリックを開発し、その信頼性と妥当性を検討することである。先行研究を参考にして、ルーブリック (Ver. 1) を開発した。これをプレFD科目にて使用し、2名の採点者で大学院生20名のマイクロティーチングを評価した。信頼性の指標として用いた評価者間信頼性に関しては、4つの観点のうち、3つの観点で中程度およびまずまずの一致が確認された。妥当性の検証のため、2つの外的指標との相関を検討したところ、予想通りの結論が得られた。一方で、ルーブリックに内容上の問題が見られたため、修正したルーブリック (Ver. 2) を開発した。このルーブリックを同じ2名で別の大学院生33名の評価に使用し、同様の検討を行なった。その結果、評価者間信頼性では、4つのうち2つの観点でしか好ましい結果を得られなかった。妥当性に関しては、一部で予想通りの結論が得られた。最後に、評価者間信頼性を向上させる観点から、改善案としてルーブリック (Ver. 3) を提案した。

キーワード：マイクロティーチング、プレFD、ルーブリック、信頼性、妥当性

1. 問題と目的

1.1. プレFDの普及と努力義務化

2000年代半ば以降、大学教育においてプレFD (Faculty Development) という言葉が徐々に広まってきた。プレFDについて説明を行うため、まずはFDから説明する。

FD (ファカルティ・ディベロップメント) とは、教育者、研究者、一人の人間という3つの側面から大学教員に必要とされる、教育評価スキル、競争的資金を獲得する方法、キャリアプランニング、学生指導スキル、健康管理、対人関係スキルなどの様々なスキルの獲得、向上に関する取り組みを指す (POD, 2016)。プレFDとは、大学院生やオーバードクター、ポスト・ドクターを対象に行われる (田口ほか, 2010; 今野, 2016)、FDの前段階に位置づく大学教員養成活動のことである (田口ほか, 2014; 田口ほか, 2010)。

プレFDに関して、2008年の答申「学士課程教育の構築に向けて」(中央教育審議会答申, 2008)では「教育研究上の目的に応じて、大学院における大学教員養成機能(プレFD)の強化を図る」ことが求められ、また、2019年には文部科学省(2019)が、大学教員の教育能力向上のため、大学院を持つ大学で、プレFDが努力義務化したことで、その注目度が上がった。1999年にFDが努力義務化されてのち、10年後の2008年にはFDが義務化されたという高等教育の歴史的な経緯を踏まえると、今後プレFDが義務化される可能性も見込まれ、プレFDに対する関心はさらに高まっていくと予想される。

1.2. プレFDに関する先行研究

プレFDに関して国内で先駆的に研究として報告されたのが、田口ほか(2010)による京都大学文学研究科において実施されているプロジェクトである。田口らは、京都大学文学研究科において実施されているプロジェクトの特徴と具体的な実践を紹介した上で、プロジェクト参加者運営に関わった教務補佐員に対して聞き取りを行い、プロジェクトの良い点などの評価を行なっている。林ほか(2013)では、立命館大学におけるプレFDについて紹介し、「セミナーは全体として満足できるものでしたか?」等のアンケート項目への回答と参加者の意見を紹介している。今野(2016)では、東北大学におけるプレFDプログラムの内容を紹介した上で、成果として修了生の数およびアンケート調査の結果等を報告し、肯定的な評価が得られたことを指摘している。大山ほか(2017)では、大阪大学におけるプレFDプログラムの一部である「大学授業開発論Ⅲ」をとりあげ、SoTLの考えに則り受講生が教育実践研究を計画する実践を行い、プログラムの評価と課題を主に質的な観点から検討している。根岸(2018)では、同じく大阪大学プレFDにおいて、受講生である大学院生が行う授業実践をお互いに観察し合うプログラムを開発・試行し、大学院生の授業イメージの変容の観点からプログラムの成果と課題を検討している。比喩生成課題とインタビュー調査を併用して分析した結果、参加者の授業イメージの変容を確認している。

一方で、課題も指摘されている。田口ほか(2010)は、

プロジェクト参加者が具体的にどの程度、「教育者として成長したのか」については十分な評価は行えていない点をあげている。この課題をもとにして、田中ほか (2014) では、京都大学文学研究科のプレFDを修了した大学院生、オーバードクター、ポストドクが、プログラムを通じてどのような意識の変容を被り、どのような能力を獲得したかを定性的に明らかにしている。その結果、「教育技術の向上」「知識の獲得」「教育者としての意識の向上」「研究者としての成長」が記述から得られたという。他にも、栗田 (2015) では、プレFDの課題を整理している。栗田 (2015) によると、「対象の拡大」「認知度の向上」「教員の協力」「カリキュラム」「人的資源」「TA制度との連携」「推進のための研究」を課題に挙げている。このうち、「推進のための研究」では、「効果検証」と「カリキュラムの開発」がプレFDの課題として挙げられている。これを一つの契機とする形で、以降各大学のプレFDにおいて、そのカリキュラムの記述および効果検証が行われるようになっていく。

このように、プレFDに関する研究としては、「プログラムの実践報告」に加え、「効果検証」が様々に行われるようになってきている。しかしながら、これらの効果検証の多くは調査対象者の自己認識を調査しており、「教育者として成長したのか」という田口ほか (2010) が持っていた課題意識に対して一面的にしか答えられていないのが現状である。また、これらの効果検証で用いられた指標は成績評価で用いられたものではなく、あくまで追加的に調査された指標である。成績評価に用いられる指標は、当該授業で最も伸ばしたい能力に強くかかわっていると想定され、「教育者として成長したのか」に応えるためにはそのような指標を用いることがより望ましい。そこで次節では、九州大学におけるプレFDにおいて成績評価の対象とした「マイクロティーチング」に注目して、議論を進めていく。

1.3. プレFDの核としてのマイクロティーチング

今野 (2016) では、日本国内の主なプレFDの取り組みを概観している。プレFDの具体的な内容としては、多くは高等教育政策の現状や学習目標、授業内容、教授法、評価方法などコースデザインに関する講義に加えて、模擬授業などの演習も実施されており、実践的に学ぶプログラムとなっていることが見出されている。そして、模擬授業においては、しばしば「マイクロティーチング」が利用されている (今野, 2016, p. 63)。

マイクロティーチングとは、指導教官や他の参加者からフィードバックを受け取る、一般的な授業をスケールダウンし5分から20分で行うミニサイズの授業のことを指す (Allen, D. W., & Eve, A. W., 1968; Perlberg, 1988)。なおここで

のスケールダウンの意味としては、東京大学や九州大学のように90分の授業を短い時間に凝縮すること (鄭ほか, 2021)、および東北大学のように90分の授業の一部をそのまま切り出すこと (今野, 2016)、の両方の実態を含むものとする。日本では、他にも大阪大学や名古屋大学でプレFDにおいてマイクロティーチングが実施されている (今野, 2016)。現職の教員とは異なり、実際に自分で授業を行ったことのない大学院生にとっては、マイクロティーチングは限られた時間ではあるものの、授業スキルを学ぶ非常に大きな機会となる。

1.4. マイクロティーチングの評価

マイクロティーチングにおいて問題になるのが、授業スキルをどのように評価するかということである。マイクロティーチングにおける評価は、その時点で各参加者が持つ授業スキルを把握し、それに従い改善案を提示するほか、プレFDが科目として提供される場合には最終的に公正かつ正確な成績評価を行う上で、非常に重要となる。そこで、マイクロティーチングにおいて、実演中に様々なタイミングで表出する授業スキルを、身振り手振りなど形式的なものに限らず適切に捉えるための明確な評価基準として、ルーブリックが必要になる。ルーブリックとは、「パフォーマンス (作品や実演) の質を評価するために用いられる評価基準」のことであり、「一つ以上の基準 (次元) とそれについての数値的な尺度、および、尺度の中身を説明する記述語」からなる (松下, 2012, p. 82)。田口ほか (2010) が指摘するように、参加者の教育者としての成長を把握する上でも、また、カリキュラムの問題点を把握し、それをのちのカリキュラム改善につなげるといふ栗田 (2015) の指摘した課題を克服する上でも、マイクロティーチングを評価することのできるルーブリックを開発することは非常に重要である。

マイクロティーチングにおいてルーブリックを活用した研究として、国内では、牧野 (2017) が挙げられる。牧野 (2017) では、参加者自身にマイクロティーチングを評価するルーブリックを作成、それを用いて評価を行うことを試みられた。観点としては、「伝え方」「板書」「内容」が設定されている。海外の研究としては、Gan & Saphthaswaran (2020) が挙げられる。同研究では、プレFDに参加した学生たちを対象に、マイクロティーチングの中に協同学習 (Collaborative Learning) を促す状況を設定し、そこでも「学生」たちとの相互作用を評価するためにルーブリックが採用された。観点としては、「既有知識 (prior knowledge)」「学習目標 (learning objectives)」「活動の種類 (Types of Activities)」「流れ (Sequencing)」「役割の分担 (Role Distribution)」「表現の種類 (Types

of Representation)」「評価 (Evaluation)」が設定されている。他にも大学院生を対象に、比較的多くの研究が見られ、Açıkgül & Aslaner (2020) では、小学校教員を目指す大学3年生の学生たちを対象に、ルーブリックを使用している。観点としては、「内容的知識 (Content Knowledge)」「授業内容を想定した教授法に関する知識 (Pedagogical Content Knowledge)」「技術とかかわる内容知識 (Technological Content Knowledge)」「技術とかかわる教育的 content 知識 (Technological Pedagogical Content Knowledge)」が設定されている。

このように、これまでの研究では、すでにマイクロティーチングにおいて用いられるルーブリックが複数開発・利用されているが、今回の授業で使用するにはそれぞれ以下の問題を残している。牧野 (2017) のルーブリックは、「伝え方」「板書」「内容」という3つの観点のうち、「伝え方」「板書」の2つの観点をさらに2つずつに分割し、実質的に合計5つの観点で評価を行なっているが、そのうちの4つは教員側の「伝える」という視点によっている。加えて、「内容」の観点の中に「わかりやすさ」という認知的な要素と「興味関心を引く」という情意的な要素が組み込まれている。そのため、近年重視されている「学習者」にとっての学習成果という観点が全体として薄く、バランスに偏りがある。Gan & Saphthaswaran (2020) のルーブリックは授業の流れに沿ったもので、非常に参考になるが、今回のマイクロティーチングでは必ずしも協同学習に限定しておらず、かつ論文中では観点のみの提示にとどまっているので、このままでは使用できない。また、情意面に対する視点が欠けている。Açıkgül & Aslaner (2020) のルーブリックは、TPACK の理論に依拠したルーブリックとして価値があるが、本授業の参加者の分野の多様性を考慮すると、板書で授業を行う学生もいると想定され、必ずしもテクノロジー活用は前提にはならないため、そのままの使用はできない。最後に、Park (2018) のルーブリックは記述語を含んでいないため、フィードバックとして返却した際に、学習者にとって次に目指すべきパフォーマンスがわからず、効果的なフィードバックにならないという限界がある。

以上の状況を鑑み、本研究では90分の授業を短い時間に凝縮して実施されるマイクロティーチングを評価するための新たなルーブリックを開発することを試みた。先行研究と比較すると、①学習者目線を重視したこと、②授業の特定の形式に限定されない (すなわち協同学習やテクノロジー活用に必ずしも縛られない) 汎用的なものであること、③情意面への配慮を観点に含めること、を特徴としたルーブリックの開発が必要と考えた。先行研究で見られたような、より領域や状況に限定されたルーブリックと比較して、さまざ

まな分野の学生が参加するプレFDにおけるマイクロティーチングを想定したルーブリックを開発し、その信頼性と妥当性を検討しようとする点に、本研究の独創性を見出すことができる。

1.5. 本研究の目的

そこで、本研究ではプレFDのマイクロティーチング用ルーブリックを開発し、信頼性・妥当性を検討することを目的とする。この目的のため、本研究では、まず先行研究を参考にしてルーブリック Ver. 1を開発した。この Ver. 1についてまず評価者間信頼性および基準関連妥当性を検討した (研究1)。その後、Ver. 1の問題点を克服するために Ver. 2のルーブリックを開発し、同様の検討を行った (研究2)。

2. 研究1の方法

2.1. ルーブリック Ver. 1の開発

ルーブリックの内容的妥当性を担保するために、まず Feldman (1997) で提示されている学習の理解度に影響を与える要素を参照した。このうちマイクロティーチングという短い時間のパフォーマンスを評価可能で、かつ、できる限り授業スキルの様々な側面に焦点を当てた項目として「説明の明確さと理解しやすさ」「教員による知的な刺激」「クラスの水準や進捗への理解」を選択した。また、近年アクティブ・ラーニングが普及している流れを受けて教員と学生間、あるいは学生間の「双方向のコミュニケーション」も評価の対象に加えた。これらの観点は、一般的なマイクロティーチングにおいて求められる (話し方や資料の提示の仕方、問いかけなどといった授業スキルの習得という目的よりも高度なものとなっている。その理由としては、初等中等教育と異なり教育実習のような制度が確立されていない大学教員養成においては、より高度な目的をプレFDで掲げ、そこに至ることができるように指導することが必要と考えたためである。後述のように、「わかりやすい授業を行うことができる」という学習目標に到達できるようにすることをこの授業では意図しているため、学習の理解度に影響を与える要素を大学教育の文脈で列挙している Feldman (1997) は評価観点として参照するに妥当と判断された。なお、話し方や資料の提示の仕方等については、後述のピア評価の項目に含め、こうした観点については他の受講生からフィードバックをもらえるように意図した。評価にかかる教員の負担も考慮し、以上の4つの観点で、4段階 (レベル0-レベル3) を評価に用いることとした。これらを学生にもわかりやすい言葉で表現することを意図して、最終的に「説明の明確さと理解しやすさ」「教員による知的な刺激」「クラスの

表1 ルーブリック (Ver. 1/Ver. 2)

観点	レベル3 模範的	レベル2 発展段階	レベル1 標準的	レベル0 要改善
説明の わかりやすさ	説明は非常にわかりやすかった	説明はある程度わかりやすかった	何を伝えたいのかいまいちよく分からなかった	レベル1を 満たさないもの
動機付け	内容の面白さや重要性など、学生が授業を集中して聴きたくなるような刺激を十分に与えた	内容の面白さや重要性など、学生が授業を集中して聴きたくなるような刺激をある程度与えた	学生が授業を集中して聴きたくなるような刺激をほとんど与えず、淡々と授業を進めた	
学生の理解度の 確認	クラスの水準や進捗を、理解度を 確認する工夫によりしっかりと理解 した上で授業を進めた	クラスの水準や進捗を、理解度を 確認する工夫によりしっかりと理解 していたが、それを授業に十分 に活かすことができていなかった	クラスの水準や進捗をほぼ気にせ ず授業を進めていた	
学生との コミュニケー ション*	教室内では十分な双方向性が保 たれていた	学生が教員あるいは他の学生とコ ミュニケーションを取りやすい雰 囲気を作ろうとしていたが、十分に 機能していなかった	ほぼ一方的に授業を進め、教 室内でのコミュニケーションへの 配慮はなかった	
目標や構造の 提示**	構造や目標を明確に提示しており、 学生が非常に取り組みやすい 授業だった	構造や目標を明確に提示している が、十分に機能していなかった	構造や目標はほとんど提示されて いなかった	

* Ver. 1のみ ** Ver. 2のみ

水準や進捗への理解」「双方向のコミュニケーション」を順に「説明のわかりやすさ」「動機付け」「学生の理解度の確認」「学生とのコミュニケーション」と名付け、観点とした。ルーブリックの記述語に関しては採点者2名で話し合い、洗練させた。ルーブリックの Ver. 1 を表1に示す。

2.2. 授業と調査の概要

(1) 授業の概要

科目「大学の授業をデザインする」は、2020年度より新規に開講された2単位の科目である。2020年度は夏学期、夏季集中講義、秋学期、2021年度はこれまでに春学期と夏学期に実施された。

この科目は、大学で将来教員として授業を持つ上で重要な教育技術に関して実践的に学ぶものであり、学習目標として以下の4点が掲げられている（マイクロティーチングは2点目に対応している）。

- ・質の高い「シラバス」を作成できる
- ・わかりやすい授業を行うことができる
- ・質の高い「評価方法」を作成できる
- ・多様性に配慮して授業デザインを見直すことができる

これら4つの学習目標を達成するために、15コマのスケジュールは表2のように構成されている。本科目の特徴として、授業の実施に必要な知識を学習する機会を設けつつ、それらを実践することで学びを深めていく点、そしてウィギンズ・マクタイ (2012) の「逆向き設計論」を内容の中心に据えて授業がデザインされている点が挙げられる。この授業の中で、受講生は計2回、マイクロティーチングを実施する。初回は本授業ではまだ具体的な事柄については学

表2 全15コマの内容

回	タイトル
1	オリエンテーション
2	授業デザインとは何か
3	<u>マイクロティーチング 1</u>
4	<u>マイクロティーチング 1</u>
5	目標・評価 (理論編)
6	シラバス (理論編)
7	目標・評価の作成 (実践編)
8	シラバスの作成 (実践編)
9	学習の科学
10	3つの教授法と授業デザインシート
11	多様な授業方法の体験
12	授業デザインの検討
13	<u>マイクロティーチング 2</u>
14	<u>マイクロティーチング 2</u>
15	合理的配慮体験とまとめ

* 開講期ごとに、一部順番の入れ替わりあり

んでいない状態で実施し、現在の自身のパフォーマンスに関して評価を受ける。2回目のマイクロティーチングまでに実施される、第5回から第12回までの授業では、逆向き設計論の流れに従って授業設計について学び、その学びは2回目のマイクロティーチングに活かせる要素が多く含まれている。例えば、学習者の理解度の把握方法 (第5回) や、コースの設計方法 (第6回)、動機付けの方法 (第9回)、授業に学生を引き込む方法 (第11回)、セッションの設計

方法（第12回）などである。そのため、初回のマイクロティーチングにおける形成的評価が、第5回から第12回までの授業の重要性を意識させる機会となっている。そして2回目のマイクロティーチングが2名の教員によって評価され、その平均点が最終評価に用いられる。なお、1回目のマイクロティーチングは、成績評価の対象とはならず、教員にとっては、参加者の現状（授業スキル、専門分野など）を把握する形成的評価の意味を持ち、参加者にとっては、2回目のマイクロティーチングに向けてのフィードバックを与えられる機会となるように意図している。

(2) 調査参加者

本調査は、2020年度夏学期（7～8月）及び同年度夏季集中講義（8月）において行われた。対象学年は修士課程1年生から博士後期課程3年生までで、受講生計30名のうち、データの研究利用を承諾し、かつ研究に必要なデータの欠損がなかった20名分のデータを対象とした。専門分野は自然科学系から社会科学系、人文学系、複合領域系まで多様であった。

(3) 採点者

採点者は、本授業の担当教員である、高等教育学を専門とする大学教員2名である。授業前の段階で、この2名でルーブリックを作成した。教育経験年数は、授業開始時点で、1名が2年、もう1名は未経験であった。

2.3. 調査項目

(1) マイクロティーチングの教員評価

マイクロティーチングの教員による評価には、前述のルーブリック Ver. 1 を用いた。このルーブリックは、4つの観点（「説明のわかりやすさ」「動機付け」「学生の理解度の確認」「学生とのコミュニケーション」）で、4段階（レベル0-レベル3）で構成されている。

なお、マイクロティーチングの実施・評価の際には教室を2つにわけ、それぞれの部屋を教員が1名ずつ担当した。1名の教員はその場で直にマイクロティーチングを評価できるか、もう1名はできないため、もう1名は録画されたマイクロティーチングの動画を後日視聴しながら採点を行った。

(2) マイクロティーチングのピア評価

マイクロティーチングのピア評価を基準関連妥当性検証のための変数として用いた。本変数を用いた理由は、ルーブリック Ver. 1（表1）が妥当なものであれば、それは授業スキルを適切に評価するものとなっているはずであり、ルーブリックで得点が高いほどそのマイクロティーチングに参加した他の受講者からの評価も高くなると考えたためである。

3-5名の参加者が、互いのマイクロティーチングに「学習者」として参加し、マイクロティーチングを行う「授業者」

のパフォーマンスを以下の5つの項目（菅原, 2015）でピア評価した。

- ・授業者の声が聞き取れた
- ・授業者の説明・質問がわかりやすかった
- ・学習者を見る視線・表情は適切だった
- ・授業の内容は理解しやすかった
- ・教材の提示の仕方が工夫されていた

ピア評価では、「よく当てはまる (5)」「当てはまる (4)」「どちらともいえない (3)」「当てはまらない (2)」「全く当てはまらない (1)」の5件法から回答してもらった。5項目に対する回答の平均値を合計し、各マイクロティーチングに対するピア評価の得点として用いた。これらの項目は、声の聞きやすさや視線や表情の適切さなど、形式的な面も含んでいるが、良い授業の要件としては必要なものである。そのため、マイクロティーグ用のルーブリックで良い点数を得る人ほど、これらの項目でも高い得点を得ており、正の相関があると考えられる。

なお、この5項目のうち、「授業者の説明・質問がわかりやすかった」「授業の内容は理解しやすかった」の2項目については、ルーブリックの「説明のわかりやすさ」と類似している。それは、ルーブリックの観点を本研究でFeldman (1997)などを参考に作成した際に、ピア評価の項目と独立に採点したために生じた重なりである。そのため、この2項目に関してはルーブリックとの内容的重なりが見られるが、他の項目に関しては重なっていない。そのため、ルーブリックとピア評価の得点には正の相関が予想されるものの、値としては大きくないと考えられる。

(3) 教育方法に関する不安

教育方法に関する不安も基準関連妥当性検証のための変数として用いた。本変数を用いた理由は、ルーブリック Ver. 1 が妥当なものであれば、それは授業スキルを適切に評価するものとなっているはずであり、ルーブリックで得点が高いほどマイクロティーチングの実施者自身の教育不安が小さくなると考えたためである。

教育方法の不安に関する指標は、田口ほか (2006) で使用されたものを利用した。項目例は、「授業の流れや指導計画に対する不安」であり、「強く感じている (4)」「やや感じている (3)」「あまり感じていない (2)」「全く感じていない (1)」の4件法から回答してもらった。田口ほか (2006) で使用された項目数は15であるが、本調査に適さないと判断された5項目（「他の授業に劣らないような授業ができていないかに関する不安」「学生の興味に沿った授業ができていないかに関する不安」「授業中に話したことや板書に誤りがないかという不安」「自分の授業で学力がついているかに関する不安」「担当科目が学生にとって意義

表3 基礎統計量（Ver. 1）

指標	平均値	中央値	標準偏差	最小値	最大値
マイクロティーチングの教員評価					
説明のわかりやすさ	2.63	2.75	0.43	2	3
動機付け	2.30	2.50	0.66	1	3
学生とのコミュニケーション	2.23	2.25	0.60	1	3
学生の理解度の確認	2.18	2.25	0.61	1	3
合計	9.33	9.00	1.70	5.50	11.50
マイクロティーチングのピア評価					
合計	21.67	22.20	1.87	17.60	24.67
教育方法に対する不安					
合計	23.90	23.00	5.09	16	35

あるものであるのかという不安」を除き、全10項目を用いて調査を実施した。

2回目のマイクロティーチング実施日は、当該科目の最終日にあたる。最終日の授業終了時に教育方法に関する不安を質問紙で尋ねた。

2.4. 分析方法

本研究では、ルーブリックの信頼性と妥当性を検討した。信頼性は、評価者間信頼性を検討した。評価者間信頼性の検討には、2名の採点者による一致度（%）とκ係数を用いた。

次に妥当性の検討方法ではルーブリックの得点と2つの外的指標（マイクロティーチングのピア評価、教育方法に関する不安）との関連を検討した。この際、ルーブリックの集約可能性、すなわち4つの観点で得られた得点を合成変数として集約可能であることを確認するために、採点者2名の採点の平均値を求めて主成分分析を行った。マイクロティーチングのルーブリックが妥当なものであれば、得点が高い参加者ほど、ピア評価は高くなり（=正の相関になる）、不安は少ない（=負の相関が見られる）と想定される。

以上の分析においては、フリーの統計ソフト HAD（清水, 2016）を用いた。

3. 研究1の結果・考察

3.1. 基礎集計

はじめに、基礎集計として、各指標の平均値、中央値、標準偏差、最小値、最大値を算出した。結果を表3に示す。マイクロティーチングに関しては、「説明のわかりやすさ」の平均値がもっとも高くなっている。レベルが0-3までの4段階で、観点が4つであるため、合計の理論的最大値は12点であるが、それに達したものはおらず、平均点は9.33

表4 評価者間信頼性（Ver. 1）

観点	一致度	κ係数
説明のわかりやすさ	75%	0.47
動機付け	60%	0.35
学生の理解度の確認	30%	-0.08
学生とのコミュニケーション	65%	0.40
全体	57.5%	0.30

点であった。ピア評価の結果に関しては、5つの設問があり5件法の選択肢が設定されていたため合計の理論的最大値は25であるが、それに達したものはおらず、平均は21.67点であった。教育方法に関する不安は10項目で選択肢は4件法であるため、合計の理論的最大値は40であるがそれに達したものはおらず、平均値は23.90であった。また、ピア評価と教育方法に関する不安の内的整合性は、順にα=.73、.80であった。

3.2. 信頼性の検討

信頼性としては、評価者間信頼性を検討した。評価者間信頼性の検討として一致度とκ係数を観点ごとおよび全体でみると、表4のとおりであった。一致度に関しては、全体としては Reddy & Andrade (2010) の70%という基準には達していないものの、「学生の理解度の確認」という観点を除けば60%以上の一致は見られた。また、κ係数に関しては、Landis & Koch (1977) の基準に基づくと、「説明のわかりやすさ」「学生とのコミュニケーション」に関しては順に中程度の一致、「動機付け」に関してはまずまずの一致であった。一方で、「学生の理解度の確認」に関してはわずかに一致という結果であった。

3.3. 妥当性の検討

ルーブリックの得点と2つの外的指標（マイクロティーチングのピア評価、教育方法に関する不安）との関連を検討した。ルーブリックの得点について、次元性の確認のため、採点者2名の評点の平均値を用いて固有値の減衰状況を確認した。固有値は、順に2.167、0.900、0.572、0.360であり、1因子を想定するのが妥当と考えられた。寄与率が54.17%であった。また、クロンバックの α 係数は $\alpha = .71$ であった。以下の検討では、2名の評定者の平均値を、4つの観点分合計した値をルーブリックの得点として用いた。マイクロティーチングのルーブリックの得点と、ピア評価得点、教育方法に対する不安得点の相関は、順に $r = .57$ 、 $-.51$ でともに有意であった ($p < .05$)。これらの結果は、基準関連妥当性の証拠となりうる。

4. 研究2の方法

4.1. ルーブリック Ver. 2 の作成

以上、ルーブリック Ver. 1 の信頼性と妥当性の結果を報告した。一方で、Ver. 1 を使用した結果、「学生の理解度の確認」と「学生とのコミュニケーション」の観点の内容的類似性が高いことが採点者間で問題視された。その理由としては、例えば理解度を確認するために授業者が参加者に質問をした場合、「学生の理解度の確認」の得点のみならず「学生とのコミュニケーション」の得点も高くなることがわかったためである。そこで、「学生とのコミュニケーション」という観点を外し、「目標や構造の提示」を新たに観点に加えた。これは、科目全体として、目標志向性の高い「逆向き設計論」(ウィギンズ・マクタイ, 2012) を重視して指導しており、また、Feldman (1997) でも、学習の理解度に影響を与える要素として目標の提示が提示されているためである。これを経て、Ver. 2 を作成した(表1)。

なお、「学生の理解度の確認」については、評価者間信頼性は低かったものの、内容としての重要性および記述を検討した上で、ルーブリックを使用していく過程で一致度が高まる可能性も考慮し、記述語をそのまま用いた。そのため、Ver. 1 との違いとしては、「学生とのコミュニケーション」という観点をなくし、「目標や構造の提示」を加えた点である。

4.2. 授業と調査の概要

(1) 授業の概要

同じく、科目「大学の授業をデザインする」において調査を行った。今回は、2020年10-12月(秋学期受講者)、2021年4-6月(春学期受講者)、2021年8月(夏学期受講者)の科目を対象とした。

(2) 調査参加者

対象学年は、修士課程1年生から博士後期課程3年生までで、2020年度秋学期、2021年度春学期、夏学期の受講生計36名のうち、データの研究利用を承諾し、かつ研究に必要なデータの欠損がなかった33名分のデータを対象とした。専門分野は、研究1と同様、自然科学系から社会科学系、人文学系、複合領域系まで多様であった。

(3) 採点者

採点者は、ルーブリック Ver. 1 を用いて採点した際と同じ2名の教員で行った。

(4) 調査項目・分析方法

調査項目としては、研究1と同様、「マイクロティーチングの教員評価」「マイクロティーチングのピア評価」「教育方法に関する不安」の3つである。信頼性と妥当性の分析は、研究1と同様の分析を行った。

5. 研究2の結果・考察

5.1. 基礎集計

基礎集計として、各指標の平均値、中央値、標準偏差、最小値、最大値を算出した。結果を表5に示す。マイクロティーチングに関しては、合計の理論的最大値は12点であり、それに達した調査対象者が観察された。平均点は8.88点であった。ピア評価の結果に関しては、合計の理論的最大値である25点に達したものが見られた。平均は22.52点であった。教育方法に関する不安の合計の理論的最大値の40に達したものはなかった。平均値は24.76であった。また、ピア評価と教育方法に関する不安の内的整合性は、順に $\alpha = .70$ 、 $.87$ であった。

5.2. 信頼性の検討

信頼性としては、評価者間信頼性を検討した。評価者間信頼性の検討として一致度と κ 係数を観点ごとおよび全体でみると、表6のとおりであった。一致度に関しては、全体としてはルーブリックの Ver. 1 よりやや下がっている。「説明のわかりやすさ」と「動機付け」の一致度が Ver. 1 の時よりも若干下がり、「学生の理解度の確認」はほぼ同じ一致度であった。一方で、Ver. 2 で加わった「目標や構造の提示」は40%程度であった。 κ 係数に関しては、Landis & Koch (1977) の基準に基づくと、「説明のわかりやすさ」「動機付け」に関しては、Ver. 1 の時と同様、順に中程度の一致、まずまずの一致であった。一方で、「学生の理解度の確認」および「目標や構造の提示」に関してはわずかに一致という結果であった。「説明のわかりやすさ」「動機付け」「学生の理解度の確認」は、2つのバー

表 5 基礎統計量 (Ver. 2)

	平均値	中央値	標準偏差	最小値	最大値
マイクロティーチングの教員評価					
説明のわかりやすさ	2.42	2.50	0.44	2	3
動機付け	2.24	2.00	0.55	1	3
学生の理解度の確認	1.88	2.00	0.63	0.50	3
目標と構造の提示	2.11	2.00	0.68	0.50	3
合計	8.88	9.00	1.66	5	12
マイクロティーチングのピア評価					
合計	22.52	22.67	1.21	18.50	25
教育方法に対する不安					
合計	24.76	25.00	5.90	13	38

表 6 評価者間信頼性 (Ver. 2)

観点	一致度	κ 係数
説明のわかりやすさ	69.70%	0.43
動機付け	57.58%	0.26
学生の理解度の確認	30.30%	-0.05
目標や構造の提示	42.42%	0.12
全体	50.0%	0.20

ジョンで共通に使用された観点であるが、Ver. 1 および Ver. 2 を通しての訓練効果は見られなかった。

5.3. 妥当性の検討

ルーブリックの得点と2つの外的指標（マイクロティーチングのピア評価、教育方法に関する不安）との関連を検討した。ルーブリックの得点について、次元性の確認のため、採点者2名の評点の平均値を用いて固有値の減衰状況を確認した。固有値は、1.944、1.192、0.539、0.325 であり、ガットマン基準によると2因子を想定するのが妥当と考えられた。一方で、MAPは、.2323、.3675、1.0000と変化しており、1因子を支持した。Ver. 1 で1因子が支持されたこと、および、本科目では最終的には得点を合成することから、1因子を構成すると判断した。寄与率は48.59%であった。また、クロンバックの α 係数は $\alpha = .62$ であった。以下の検討では、2名の評定者の平均値を、4つの観点分合計した値をルーブリックの得点として用いた。

マイクロティーチングのルーブリックの得点と、ピア評価得点の相関は、 $r = .42$ で、5%水準で有意であった。この結果は、基準関連妥当性の証拠となりえよう。一方で、教育方法に対する不安得点の相関は、 $r = -.14$ で、有意ではなく、この結果からは妥当性の証拠を得ることはできなかった。

6. 総合考察

6.1. 本研究のまとめと意義

本研究では、プレFDにおけるマイクロティーチングの評価のためのルーブリックを開発し、その信頼性と妥当性を検討した。その結果、信頼性に関しては一部で解決すべき問題が見られるものの、妥当性に関しては概ね期待される証拠を得ることができた。

ルーブリックに関しては、これまでに信頼性と妥当性に関する検討が不足していることが指摘されていた (Reddy & Andrade, 2010)。そのため、本研究で開発したルーブリックがある程度の信頼性と妥当性を有していることを示したことは、ルーブリックの研究自体への貢献という学術的な意義をもつ。また、プレFDが普及する中で、すでにいくつかの大学で取り入れられているマイクロティーチングを評価するためのルーブリックを開発したことは、今後のプレFDにおける評価実践に対して実践的な意義をもたらすと考えられる。

6.2. 本研究の限界と今後の課題

最後に本研究の限界と今後の課題に言及する。信頼性の課題、妥当性の課題、そして調査対象者に関する課題の順に述べる。

(1) 信頼性の課題

今回の結果を見るに、信頼性の指標とした評価者間信頼性に関しては課題が残った。まず、Ver. 2においては、新たに追加した「目標や構造の提示」に関しては κ 係数が.20未満と低かった。また、Ver. 1とVer. 2で共通して「学習者の理解度の確認」に関しては κ 係数が低かった。これら2つの観点に関して、Ver. 2での採点後に採点者2名（以下、簡単のため採点者A、採点者Bとする）でルーブリックに書き込んだメモをもとに話し合い検討したところ、そのズレの理由として以下のことがわかった。

表7 ルーブリック (Ver. 3)

観点	レベル3 模範的	レベル2 発展段階	レベル1 標準的	レベル0 要改善
説明の わかりやすさ	説明は非常にわかりやすかった	説明はある程度分かりやすかった	何を伝えたいのかいまいちよく分からなかった	レベル1を 満たさないもの
動機付け	内容の面白さや重要性など、学生が授業を集中して聴きたくするような刺激を十分に与えた	内容の面白さや重要性など、学生が授業を集中して聴きたくするような刺激をある程度与えた	学生が授業を集中してききたくするような刺激をほとんど与えず、淡々と授業を進めた	
学生の理解度 の確認	タイミングと質の観点で、適切な理解度確認が行われた	タイミングと質の観点で、適切な理解度確認がある程度行われたが、改善の余地がある	理解度を確認しようとしているが、授業内容にとって無意味な確認も多い	
目標・構造と 授業との 整合性	以下のうち、3つとも達成されている。 ・目標と構造が提示されている ・目標と構造の質が高い ・授業と整合的である	以下のうち、2つが達成されている。 ・目標と構造が提示されている ・目標と構造の質が高い ・授業と整合的である	以下のうち、1つしか達成されていない。 ・目標と構造が提示されている ・目標と構造の質が高い ・授業と整合的である	

まず、「目標や構造の提示」に関して、採点者Aは記述語には明確に書かれていない「目標の達成を確認しているかどうか」を採点基準に含めていたこと、学習目標の提示の有無のみならずその質も評価しようとしていたことがわかった。そして採点者Bがマイクロティーチングの冒頭での構造の提示に注目している一方で、採点者Aはマイクロティーチング全体での構造の質を評価しようとしていたことも判明した。また、採点者Aは「レベル0」を割り当てることを控えていたこともわかった。次に、「学習者の理解度確認」に関して、採点者Bは理解度確認のために授業者が使用する質問の内容自体はそれほど問わず評価していた一方で、採点者Aは内容に関連した理解度確認かどうかに着目して採点していた。

以上を踏まえ、これらの問題点を克服したルーブリック (Ver. 3) を作成した (表7)。Ver. 3の特徴としては、まず、「学生の理解度の把握」の記述語において、量 (タイミング) と質の双方を評価対象にすることを明確化したことがある。次に、Ver. 2では「目標と構造の提示」とされていた観点を、Ver. 3では目標と構造の質自体を問う視点と、目標とマイクロティーチングの構造との整合性を要素として含んだ観点として「目標・構造と授業との整合性」という観点到修正したことである。なおここでは、チェックリストの形式を採用することで合意した。宮田・奥村 (2018) では、チェックリストとルーブリックが異なる側面を評価している可能性が高いことを指摘した上で、チェックリストの良さとして採点の容易さがあることに言及している。Ver. 2で「目標と構造の提示」の評価者間信頼性が低かったことを考えても、Ver. 3でチェックリストを一部で使用するには一定の意義があるだろう。

さらに、2名での話し合いの中で、採点結果がずれないようにアンカーとして過去の受講生の作品を添付することで

合意した。今後、こちらのルーブリックを使用し、その評価者間信頼性のみならず、本研究で行ったように信頼性と妥当性を検討していくことで、より良いルーブリックの開発につながる事が期待できる。

(2) 妥当性の課題

次に、妥当性の課題である。Ver. 1とVer. 2のルーブリックにおいてともに、ピア評価の得点と中程度の正の相関 (順に $r = .57, .42$) が見られたことは、基準関連妥当性に関しての証拠を得ることができたとともに、ルーブリックとピア評価では異なる部分が含まれること (つまり独自に評価対象としているものがあること) を示している。一方で、Ver. 1のルーブリックと教育方法に対しての不安との間には期待された負の相関が確認されたが、Ver. 2のルーブリックとの間には負の相関は確認されなかった。この理由としては、例えば、授業スキルが高い学生ほど、自分の能力をより正當に評価し、自分が克服できていない不安を正しく認識している可能性が挙げられる。妥当性に関しての証拠を探っていくことは、「永続的な作業」 (村山, 2012) とされている。本研究では、妥当性に関して一定の証拠を提供したが、今後もこのような証拠を積み上げていくことが重要である。

なお、本研究では分析結果をもとに次元性を仮定したが、寄与率から判断すると、その代表性は高いとはいえない。そのため、これらを単純加算して得点化し、本科目で設定している「わかりやすい授業を行うことができる」能力が合計点のみで代表されるように学習者に示すのではなく、過度に情報を落とさないためにも、できる限り評価情報をそのまま、すなわちルーブリックの観点ごとに返すことが重要であることが明らかとなった。

(3) 調査対象者に関する課題

最後に、調査対象数が少ない点が課題である。本研究では、「大学の授業をデザインする」の参加者を対象に調

査を行ったが、ルーブリックのバージョンを2つに分けたため、一つの調査あたりの人数は20-30名と多くはない。今後の研究では、より継続的に調査を実施し、調査対象者の数を増やして検討をしていく必要がある。

引用文献

- Açıkgül, K., & Aslaner, R. (2020). Effects of Geogebra supported micro teaching applications and technological pedagogical content knowledge (TPACK) game practices on the TPACK levels of prospective teachers. *Education and Information Technologies* 25(3), 2023-2047.
- Allen, D. W., & Eve, A. W. (1968). Microteaching. *Theory Into Practice* 7(5), 181-185.
- 中央教育審議会 (2008). 『学士課程教育の構築に向けて (答申)』 (https://www.mext.go.jp/component/b_menu/shingi/toushin/___icsFiles/afiedfile/2008/12/26/1217067_001.pdf) (2021年8月31日)
- Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Oxford, England, U. Illinois Press.
- Feldman, K. A. (1997). Identifying exemplary teachers and teaching. Evidence from student ratings. In R. Perry, & J. Smart (Eds.), *Effective teaching in higher education. Research and Practice* (pp. 368-395). New York: Agathon.
- Gan, M. J. S., & Saptaswaran, V. (2020). The Development of a Scoring Rubric for Studying Graduate Teaching Assistants' Competence in Collaborative Learning Lesson Planning and Implementation. *Asian Journal of the Scholarship of Teaching and Learning* 10(1), 40-52.
- 林 素子・沖 裕貴・松村 初 (2013). 立命館大学における PFF の取り組み, 『立命館高等教育研究』 13, 169-186.
- 鄭 漢模・長沼祥太郎・野瀬 健・丸野俊一 (2021). 「九州大学におけるティーチング・アシスタント制度改革:1年間の記録」『基幹教育紀要』 7, 189-212.
- 今野文子 (2016). 「大学院生等を対象とした大学教員養成プログラム (プレFD) の動向と東北大学における取り組み」『東北大学高度教養教育・学生支援機構紀要』 2, 61-74.
- 栗田佳代子・夏目達也・羽田貴史・佐藤浩章・小笠原正明・吉良 直 (2015). 「プレFDの現状からみえる課題と目指すべき方向性 企画」『大学教育学会誌』 37(2), 75-78.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data Published by: International Biometric Society Stable. *Biometrics* 33(1), 159-174.
- 牧野治敏 (2017). 「ルーブリック評価を導入した授業改善の試み:模擬授業を評価するルーブリックを学生自身が作成する実践」『大分大学高等教育開発センター紀要』 9, 77-84.
- 松下佳代 (2012). 「パフォーマンス評価による学習の質の評価—学習評価の構図の分析にもとづいて—」『京都大学高等教育研究』 18, 75-114.
- 宮田佳緒里・奥村好美 (2018). 「ルーブリックとチェックリストによる評価結果の関連性の検討—効果的な使いわけのために—」『兵庫教育大学 研究紀要』 52, 117-125.
- 文部科学省 (2019). 『学校教育法施行規則及び大学院設置基準の一部を改正する省令の施行等について (通知)』 (https://www.mext.go.jp/b_menu/hakusho/nc/1420657.htm) (2021年8月31日)
- 村山 航 (2012). 妥当性-概念の歴史的変遷と心理測定学的観点からの考察, 『教育心理学年報』 51, 118-130.
- 根岸千悠 (2018). 「大学教員を目指す大学院生を対象とした相互授業観察プログラムの試行と評価—授業イメージの変容に着目して—」『日本教師教育学会』 27, 146-156.
- 大山牧子・根岸千悠・佐藤浩章 (2017). 「SoTLに基づいた教育実践研究計画を作成するプレFDプログラムの試行と評価」, 『日本教育工学会論文誌』 41 (Suppl.), 225-228.
- Park, E. (2018). An Observational Case Study of Microteaching at a University in the United States. *Teacher Education Research* 57(2), 219-231.
- Perlberg, A. (1988). From Microteaching to Teaching Learning Laboratories--TLL An International Perspective. *The Annual Meeting of the American Educational Research Association*.
- POD (2016). What is Educational Development? (<https://podnetwork.org/about-us/what-is-educational-development/>) (2021年10月17日).
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education* 35(4), 435-448.
- 清水裕士 (2016). 「フリーの統計分析ソフトHAD:機能の紹介と統計学習・教育, 研究実践における利用方

- 法の提案」『メディア・情報・コミュニケーション研究』1, 59-73.
- 菅原弘一 (2015). 「研究授業・模擬授業から学び合うために」稲垣忠・鈴木克明 (編著) 『授業設計マニュアル Ver. 2—教師のためのインストラクショナルデザイン—』北大路書房, 159-168.
- 田中一孝・畑野 快・田口真奈 (2014). 「プレFDを通じた大学教員になるための意識の変化と能力の獲得—京都大学文学研究科プレFDプロジェクトを対象に一」. 『京都大学高等教育研究』20, 81-88.
- 田口真奈・出口康夫・赤嶺宏介・半澤礼之・松下佳代 (2010). 「未来のファカルティをどう育てるか—京都大学文学研究科プレFDプロジェクトの試みを通じて—」『京都大学高等教育研究』16, 91-111.
- 田口真奈・年寿西森・神藤貴昭・中村 晃・中原 淳 (2006). 「高等教育機関における初任者を対象としたFDの現状と課題」『日本教育工学会論文誌』30(1), 19-28.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design (Expanded 2nd ed.)*. Alexandria, VA: Association for Supervision and Curriculum Development.
- G・ウイギンズ・マクタイ (2012). 『理解をもたらすカリキュラム設計—「逆向き設計」の理論と方法—』(西岡加名 恵訳) 日本標準.

Educational Practice Reports

Development of a Micro Teaching Rubric for a Pre-FD Course

Shotaro Naganuma¹ and Hanmo Jeong¹

(¹Education Innovation Initiative, Kyushu University)

The purpose of this study was to develop a rubric for evaluating micro teaching in a pre-FD course and to examine its reliability and validity. The rubric (Ver.1) was developed by referring to previous studies and was used in a pre-FD course, where two scorers evaluated 20 graduate students' micro teaching. As for inter-rater reliability, an indicator of reliability in this study, fair to moderate agreement was found in three of the four perspectives. In order to verify validity, correlations with the two external indicators were examined, and the results were as expected. On the other hand, some content problems were found in the rubric, so a revised rubric (Ver. 2) was developed. This rubric was used by the same two scorers to evaluate other 33 graduate students' micro teaching, and the same analysis was conducted. The results showed that only two of the four perspectives yielded favorable results in terms of inter-rater reliability. In terms of validity, some results were as expected, while others were not. Finally, to improve inter-rater reliability, we proposed a revised rubric (Ver. 3) to be used in the future.

Keywords: Micro-teaching, pre-FD, rubric, reliability, validity