

PEPAによって学生の成長を縦断的に評価する

齋藤 有吾・松下 佳代
(新潟大学) (京都大学)

Assessing Student Growth: Based on a Longitudinal Study Using PEPA

Yugo Saito, Kayo Matsushita
(Niigata University) (Kyoto University)

Pivotal embedded performance assessment (PEPA) is a method for combining the assessment of learning outcomes at course and program levels. Compared to standardized tests, performance assessments have several disadvantages. We addressed the difficulties of inter-institutional and longitudinal comparisons and discussed how to overcome them. In particular, we examined longitudinal comparisons using an empirical study from the Faculty of Dentistry, Niigata University, which inspired the theoretical concept of PEPA. The results showed that even if a common rubric is used, the difference in scores within inter-institutional comparisons did not necessarily represent a difference in ability between groups unless a common performance task was used. Additionally, the difference in scores over time did not necessarily indicate the degree of change. This suggests that when longitudinally assessing student growth through multiple performance assessments over time, it is necessary to carefully consider two aspects in student learning assessment: task difficulty and achievement level.

〔キーワード：プログラムレベルの学修成果、パフォーマンス評価、機関間比較、経年比較、PEPA〕

1. はじめに

PEPA (Pivotal Embedded Performance Assessment: 重要科目での埋め込み型パフォーマンス評価) は、その科目の目標がプログラム全体の目標に直結し、複数の教員が関わるような重要科目の評価の中にプログラム・機関レベルの評価を埋め込み、それをパフォーマンス評価によって行うものである (Matsushita, Ono, & Saito, 2018; 松下・小野・齋藤, 2020)。その点で、それはプログラムレベルと科目レベルをつなぐ評価システムであるということが出来る。大学組織が学習システム・パラダイムへの転換を果たすためには、大学教員が担当する授業科目だけでなく学位プログラム全体を通して、学生

がどう学び成長しているかを把握することが重要であるため、PEPAはそれを実現していくために有望なアプローチの一つである。

昨年度の報告では、PEPAの実行可能性と他分野への適用可能性という課題が見いだされた (松下, 2020)。そのため、本課題研究では、東京都市大学をフィールドとして、全学的に、すなわち様々な分野におけるPEPAに取り組み始めている。その詳細は本号での伊藤他による報告を参照されたい。本報告では、PEPAの理論を創出する元になった新潟大学歯学部における実践とその研究から、PEPAによって学位プログラム全体を通じた学生の成長を把握することができるのかどうかを検討する。

2. プログラムレベルの学修成果の評価としてのパフォーマンス評価が抱える課題

(1) プログラムレベルの学修成果の評価方法

プログラムレベルの学修成果を評価する方法として、成績評価の活用（GPAやカリキュラムマップに基づいたレーダーチャート）、学生調査（質問紙調査）、標準テスト、卒業論文・卒業研究、ポートフォリオなどをあげることができる。それぞれ利点と留意点があり、例えば学生調査は機関間比較や経年比較が行いやすく、調査できる範囲が広い。ただし、自己報告のため、直接評価の代替はできない。また標準テストも機関間比較や経年比較が行いやすい。ただし、主にジェネリックスキルが測定対象であり、それが当該大学や学位プログラムのディプロマ・ポリシー（以下、DP）と合致しているとは限らないし、分野固有の知識や能力は測定できない。逆に、卒業論文・卒業研究は、単なる評価課題ではなく学習課題としての意義をもち、当該分野における卒業時点での統合的な能力を把握することができる。すなわち、DPとの整合性をもたせやすい。ただし、卒業年次まで把握できないため経年比較はできないし、評価基準を統一することは難しいため、機関間比較も困難である。

さて、PEPAもプログラムレベルの学修成果を評価するものであり、その名が示す通り重要科目におけるパフォーマンス評価によってそれを行う。卒業論文・卒業研究もパフォーマンス評価であるが、PEPAは卒業年次に限定するものではなく、カリキュラム上DPに直結する重要科目において行うものである。そのため、中間地点での統合的な能力を把握することが可能となる（松下他、2020）。また、PEPAは重要科目だけでなく、他科目で学んだ知識・スキル等がそこでどう統合されるかを通して、間接的に他科目の学修成果もみることになる。

(2) 本報告の目的

このようにパフォーマンス評価をプログラムレベルの学修成果の評価として使うことで、DPと合致した統合的な能力を直接評価することができる。ただし、一般に、パフォーマンス評価には、標準テストや学生調査と比較して、実行可能性が低く（評価負担が大きく）、機関間比較・経年比較が難しい、という問題がある（PEPAの実行可能性については、松下（2020）参照）。本報告では、パフォーマンス評価につきまとう機関間比較・経年比較の難しさを取り上げ、これらをどのように乗り越えることができるのかを検討する。特に経年比較に関しては、縦断的にパフォーマンス評価を行っている新潟大学歯学部における実証的な研究を参照しながら考

察を行う。

3. パフォーマンス評価の機関間比較

パフォーマンス評価によって機関間比較をするためには、共通の枠組みで設計されたものを複数機関で実施する必要がある。例えばAAC&Uでは、VALUE Instituteを立ち上げて、学生の成果物をもとに評価し、結果のベンチマーキングができるようにしている（AAC&U, n.d.a）。具体的な手続きとしては、まず高等教育機関が学生の成果物のサンプルを収集する。その後、成果物をデジタル・リポジトリに転送し、VALUE Instituteの認定を受けた教員がVALUEルーブリックを用いて成果物を採点する。そして、参加機関は結果のベンチマークが可能なデータや報告書を受け取る、という流れである。なお、VALUE Instituteが評価しているのは、Civic Engagement（市民参加）、Critical Thinking（批判的思考）、Ethical Reasoning（倫理的推論）、Global Learning（グローバル学習）、Integrative Learning（統合的学習）、Intercultural Knowledge and Competence（異文化知識・能力）、Quantitative Reasoning（量的推論）、Written Communication（文章コミュニケーション）といった一般的能力であり、分野固有の能力ではない（AAC&U, n.d.b）。

分野固有の能力を評価し、機関間比較が可能なパフォーマンス評価の例として、医歯学系において臨床実習前や臨床実習後に実施されるOSCE（Objective Structured Clinical Examination：客観的臨床能力試験）がある（CATO, 2020）。OSCEは、順番にステーションを移動しながら、予め定められた臨床シナリオに従い、特定の臨床能力のスキルを示すように指示される。各受験者が一連の同じタスクに対応し、標準化された評価項目を使って評価者が採点するというものである（松山・染谷, 2020）。つまり医歯学系は、このような追加型評価によって分野固有の能力の機関間比較をクリアできる。

4. パフォーマンス評価の経年比較

一方、パフォーマンス評価によって経年比較をするためには、当該プログラムの学生のみを対象とするのでもかまわないが、学年をまたいで複数回の評価を実施する必要がある。その際、複数回のパフォーマンス評価において共通のルーブリックを用いて、その評価結果の変化によって学生の成長を検討する方法がある。ただし、その評価得点によって経年比較が妥当に行えるかどうかには留意すべき点がある。パフォーマンス評価は図1に示

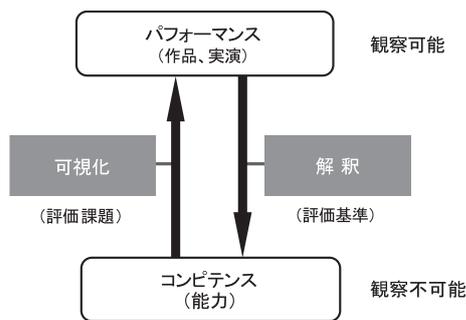


図1 パフォーマンス評価のモデル
出典：松下（2007）p.11の図2を一部改変。

すように、評価課題（パフォーマンス課題）と評価基準（ルーブリックが多用される）によって構成される。カリキュラムの段階性を考慮すれば、学年が上がるごとにパフォーマンス課題の難易度も上がることが自然である。パフォーマンス課題の難易度が上がっていく場合を考えると、共通のルーブリックの得点の変化だけでは学生の成長を適切に捉えられないという懸念がある。例えば、学年をまたいだ2つのパフォーマンス評価で、共通のルーブリックによって、ある学生がどちらも「レベル2」と評価され、それを2点と換算したとする。その場合、得点上は変化が0ということになるが、これを能力の変容もなかったと解釈するのは適切ではない。そこで、ルーブリック得点だけではなく課題の難易度を加味した解釈を行った研究を紹介する。

5. パフォーマンス課題の難易度を考慮した経年変化の解釈

紹介する研究は、新潟大学歯学部口腔生命福祉学科の、PBLにおける問題解決能力の評価である（小野・松下・斎藤，2020）。ここでは、PBLカリキュラムに埋め込まれる形で、2年前期、2年後期、3年後期の3回パフォーマンス評価が行われている。学年が上がるごとに、より複雑で真正性の高い課題に取り組むデザインである。

3回のパフォーマンス評価は、PBLにおける問題解決能力を評価するために、改良版トリプルジャンプ（Modified Triple Jump, 以下、MTJ）（小野・松下，2015）という評価方法を採用している。MTJは、いわば一人で行うPBLで、3つのステップから構成される。ステップ1で、学生はシナリオから問題を見だし、解決策を立案し、学習課題を設定して、それを60分間でワークシートに記述する。ステップ2では、1週間をかけて、設定した学習課題を調査し、解決策を検討して、最終的な解決策を提案する。そして、その内容もワーク

シートに記述し、教員に提出する。その後、ステップ3として、シナリオの状況を再現した場面において、学生は教員を相手にロールプレイを行い、解決策を実行する。ステップ3の時間は、教員からのフィードバックを含め15分間である。教員は、学生が提出したワークシートを6観点、ロールプレイを4観点、どちらも4段階からなるルーブリックにより評価する。

3回のMTJにおいて、評価課題としてのそれぞれの特徴を表すのはシナリオであり、MTJ1（2年前期）は「智歯 [= 親知らず] を抜去した友人への病状説明」、MTJ2（2年後期）は「糖尿病をもつ歯周病患者への病状説明と禁煙指導」、MTJ3（3年後期）は「歯科検診での児童虐待の発見と対応」である。カリキュラム進行に従って複雑かつ歯科と福祉の領域複合的な課題となるように作成されている。また、評価基準であるルーブリックは、3回のパフォーマンス評価で全て共通して使用している。ただし、ロールプレイのルーブリックは課題特殊なものであるため、記述語のみ多少異なる。

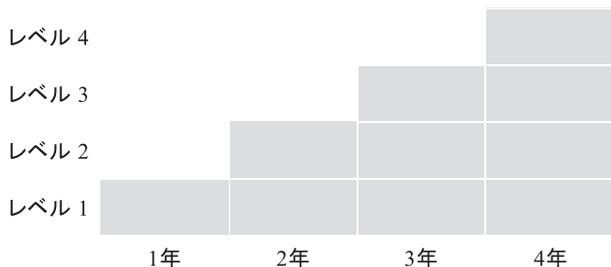
小野他（2020）では、この3回のパフォーマンス評価に関して、ルーブリック得点だけではなく課題の難易度を加味した能力の変容を解釈するため、シナリオの難易度分析、ルーブリック得点の平均的变化の分析、課題に対する学生の難易度の認識に関するアンケート調査とその分析を行った。

その結果、シナリオの難易度分析と課題に対する学生の難易度の認識に関するアンケート調査から、3回のシナリオの難易度は異なっていること、特にMTJ3は概念的なつながりが最も複雑であること、学生の難易度の認識からもこれらの知見が支持されることが明らかになった。その上で、ルーブリック得点の変化をみたところ、3回のMTJの間で全て有意差が認められ、カリキュラム進行に従って上昇していた。また、MTJ1から2にかけて大きな平均値差の効果量、MTJ2から3にかけて中程度の平均値差の効果量が得られた。

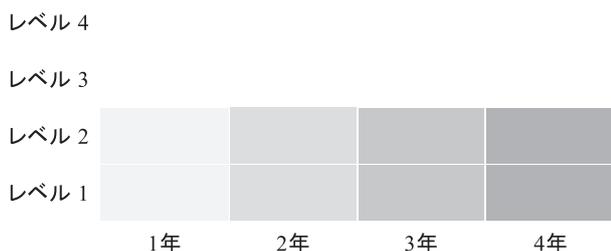
得点差と効果量だけをみれば、MTJ1から2に比べてMTJ2から3のほうが成長の度合いが小さいことになるが、しかしMTJ3が最も難易度が高いことも考慮すればそうとは限らない。どちらのほうがPBLにおける問題解決能力が大きく変容しているかを判断することは、得られているデータからは困難であるが、少なくとも、難易度を考慮せず得点の変化だけで成長を解釈しようとすると、それを過大／過小に見積もってしまう危険性が示唆されたといえる。

これらの結果から、低学年から高学年にかけて、複数回のパフォーマンス評価によって学生の成長を検討しよ

a: カリキュラムを通じて課題の難易度が一定で、評価得点の上昇によって、能力の向上が示される



b: カリキュラムが進むにつれて課題の難易度が上昇するため、評価得点が一定でも、能力の向上を意味する



(注)セルの色の濃さは課題の難易度を表わす

図2 課題の難易度と評価得点の関係模式図

出典：小野他（2020）p.9 図8

うとする場合、図2に示すように、課題の難易度を一定としてルーブリックの得点が向上したかをみるよりもむしろ、カリキュラム進行に従って課題の難易度が上昇することが自然であるため、それをふまえたうえで各科目の目標について一定のレベルに達したかという視点が重要であると結論づけている。

6. まとめと今後の課題

ここまで述べてきた課題は、プログラムレベルの学修成果の評価としてのパフォーマンス評価が共通して抱える課題である。特に、VALUEルーブリックのような共通のルーブリックを使用したとしても、共通のパフォーマンス課題を使用していない限り、機関間比較において得点の差分が集団間の能力差を表しているとは限らないし、また経年比較において得点の差分が能力の変容の度合いを表しているとはやはり限らない。特に、複数年にまたがる複数回のパフォーマンス評価によって学生の成長を縦断的に評価する際には、どのような評価課題に取り組んで、どのような評価結果を得たかという、課題の難易度と到達レベルの2つの側面から、学生の能力を慎重に捉える必要がある。

本報告で取り上げた研究は、PBLにおける問題解決能力に焦点を当てたものであるが、その検討手法は他分

野においても適用可能なものである。本課題研究の他のフィールドにおいて、PEPAの実行可能性とともに、同様の研究手法によって縦断的に学生の成長を追うことができるのかどうかを検討したい。

また、このような縦断的デザインによって、教員・学生ともにパフォーマンス評価に関わる機会が複数回生じることになる。1回のみの場合と比較して、教員のエキスパート・ジャッジメントや学生の自己評価能力に対する影響など、期待される副次的効果をあわせて検討していきたい。

謝辞

本報告の作成にあたり、新潟大学 歯学総合研究科教授 小野和宏先生より多大なご助力をいただきました。ここに記して謝意を表します。

文献

- AAC&U (n.d.a). VALUE Institute Overview. (<https://www.aacu.org/VALUEInstitute>) (2021年1月30日アクセス)
- AAC&U (n. d. b). Interest in VALUE Institute Participation. (<https://www.aacu.org/interest-value-institute-participation>) (2021年1月30日アクセス)
- CATO (医療系大学間共用試験実施評価機構) (2020) 『共用試験ガイドブック (第18版)』. (<http://www.cato.umin.jp/e-book/18/>) (2021年1月30日アクセス)
- 松下佳代 (2007) 『パフォーマンス評価』 日本標準.
- 松下佳代 (2020) 「プログラムレベルと科目レベルの評価をつなぐ—PEPAの理論と課題—」『大学教育学会誌』 42(1), 77-81.
- Matsushita, K., Ono, K., & Saito, Y. (2018). Combining course- and program-level outcomes assessments through embedded performance assessments at key courses: A proposal based on the experience from a Japanese dental education program. *Tuning Journal for Higher Education*, 6(1), 111-142.
- 松下佳代・小野和宏・斎藤有吾 (2020) 「重要科目での埋め込み型パフォーマンス評価を通して科目レベルとプログラムレベルの評価をつなぐ—歯学教育プログラムの経験にもとづく提案—」『京都大学高等教育研究』 26, 51-64.
- 松山泰・染谷真紀 (2020) 「ちゃんとした試験をすれば

授業は要らない？」錦織宏・三好沙耶佳編『指導医のための医学教育学—実践と科学の往復—』京都大学学術出版会, pp.126-139.

小野和宏・松下佳代 (2015)「教室と現場をつなぐPBL—学習としての評価を中心に—」松下佳代・京都大学高等教育研究開発推進センター編『ディープ・アク

ティブラーニング—大学授業を深化させるために—』勁草書房, pp.215-240.

小野和宏・松下佳代・斎藤有吾 (2020)「PBLカリキュラムによる長期的な学習成果の直接評価—新潟大学歯学部でのパフォーマンス評価の縦断研究にもとづいて—」『京都大学高等教育研究』26, 1-12.