

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—

松下 佳代¹・小野 和宏²・斎藤 有吾³

(¹ 京都大学高等教育研究開発推進センター・² 新潟大学大学院医歯学総合研究科・³ 新潟大学経営戦略本部)

本稿では、科目レベルの評価とプログラムレベルの評価をどのようにつなぐかを検討し、歯学教育プログラムの事例を用いて新しい評価方法を提示する。パフォーマンス評価は、教員の負担は大きい、知識の統合や高次の能力を評価するために有効であると考えられている。私たちはこれまでに、PBL 科目の「改良版トリプルジャンプ」など、科目レベルでのパフォーマンス評価を開発してきた。この経験にもとづき、本稿では、科目レベルとプログラムレベルの評価をつなぐ方法として、「重要科目に埋め込まれたパフォーマンス評価 (Pivotal Embedded Performance Assessment, PEPA)」を提案する。この方法では、パフォーマンス評価の範囲を、プログラムの教育目標に直結しカリキュラムの結節点に配置された重要科目に限定し、他の科目の評価を個々の教員のエキスパート・ジャッジメントに委ねる。PEPA は、カリキュラムの体系化と重要科目の選択、教員団によるパフォーマンス評価の設計と実施、形成的評価の機能を組み込んだ上での合格基準の設定、学位プログラムの修了認定という手続きで行われる。このようにして、PEPA は、評価の妥当性と信頼性を担保しながら、評価の実行可能性と単位制度との両立を維持することを可能にするのである。

キーワード：PEPA (Pivotal Embedded Performance Assessment)、プログラムレベルの評価、カリキュラム、PBL (Problem-Based Learning)、歯学教育

1. 問題設定

1.1. 学習成果の評価の多様性

学習成果をどう評価するかは、今日、国の違いを問わず、高等教育の重要な課題となっている。

学習成果の評価は近年多様化してきているが、それらは、①直接評価と間接評価、②量的評価と質的評価、③機関レベル・プログラムレベル・科目レベルの評価、といった観点で整理できる (Banta & Palomba, 2015; 松下, 2017; Suskie, 2009)。このうち、直接評価 (direct assessment) と間接評価 (indirect assessment) の違いは、その評価方法が、学習成果の直接的なエビデンスにもとづくか、間接的なエビデンスにもとづくかに拠っている。例えば、学習者の知識や能力の表出を通じて——「何を知り何ができるか」を学習者自身にやってみさせることで——行われる評価は直接評価であり、一方、学習成果についての学習者の自己報告を通じて——「何を知り何ができると思っているか」を学習者自身に答えさせることによって——行われる評価は間接評価である (Palomba & Banta, 1999; 松下, 2012; 山田, 2012)。

こうした分類軸を組み合わせることによって、評価方法の特徴を把握することが容易になる。例えば、松下 (2012, 2016) は、①と②を交差させることにより、学習成果の評価の 4 つのタイプとそれぞれの特徴を明らかにした (図 1)。

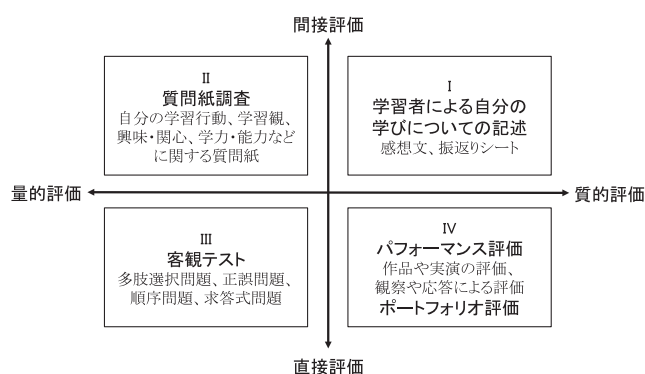


図1 学習評価における成果の評価の4タイプ

また、アメリカのアクレディテーション団体の一つ、Middle States Commission on Higher Education (MSCHE, 2007)

は、①と③を交差させることにより、機関・プログラム・科目の3つのレベルの評価方法を直接指標（direct measures）と間接指標（indirect measures）によって整理している。

これらの多様な評価方法をその特徴や機能に応じて使い分けるとともに、それらを統合することによって、学生の学習を多面的に把握し、それを教育改善につなげることが、国の違いにかかわらず、今日の大学には求められている（Banta & Palomba, 2015; Kuh et al., 2014）。

1.2. 科目レベルとプログラム・機関レベルの評価のギャップ

ただ、具体的に使われている評価方法をみると、国による違いがかなり大きい。例えば、アメリカにおいて機関レベルの学習評価として最もよく用いられているのは、「全米規模の学生調査」（85%）、「ルーブリック」（69%）、「クラス単位のパフォーマンス評価」（66%）であり、「最も価値がある（もしくは重要である）」とみなされているのもこの3つである（Kuh et al., 2014）。一方、日本の場合、「課程を通じた学生の学修成果の把握方法」は、標準テスト（32.2%）、質問紙調査（20.3%）、学修ポートフォリオ（12.7%）、ルーブリック（6.8%）となっている（文部科学省, 2017）。近年、プログラム・機関レベルの評価への要請が高まるにつれて、標準テスト（図1でいえば、タイプⅢ）や質問紙調査（タイプⅡ）の活用が増加しているのに対し、学修ポートフォリオやルーブリック（タイプⅣ）の利用は、まだ一部の大学にとどまっている。

なぜ、日本では、プログラム・機関レベルの評価において、質的で直接的な評価方法（タイプⅣ）が用いられないのか。もちろん、日本でも、科目レベルでは、パフォーマンス（レポート・制作物などの作品やプレゼンテーション・実技などの実演）による評価が数多く行われているのだが、しかし、それらの科目レベルの評価をプログラム・機関レベルの評価とつなぐための知識・技術や人的・時間的リソースが欠けているのが、その大きな要因と考えられる（松下他, 2014; 山田, 2016）。

これに対し、アメリカでは、標準テストへの批判から、全米大学カレッジ協会（Association of American Colleges and Universities, AAC&U）による Valid Assessment of Learning in Undergraduate Education（VALUE）ルーブリックが開発され（Rhodes, 2010; Rhodes & Finley, 2013）、それをきっかけに、「クラス単位のパフォーマンス評価」や「ルーブリック」がプログラム・機関レベルで使用される割合が向上した（Kuh et al., 2014, p. 33）。Banta & Palomba（2015）によれば、「VALUE ルーブリックの目的は、各科目で教員が行う評価と、それとは別に教員団や評価専門家がプログラム・機関レベルで行う評価とを結びつける手助けをすること

であった」（p. 101）とされ、また実際、そのように機能しているといえよう。

とはいえ、アメリカにおいても、プログラム・機関レベルの評価をどう行うかは依然として課題であり続けている。ユーウェル他（Ewell et al., 2011）は、プログラムレベルの評価実践の調査の結果、キャップストーン科目、ルーブリック、パフォーマンス評価など、大学・学部独自の評価が標準テストよりよく用いられていること、ただし学問分野間で大きな多様性があることを報告している。

1.3. 本稿の目的と構成

本稿の目的は、日本での学習成果の評価の状況を紹介しながら、評価研究の現在の論点の一つである、プログラムレベルの評価と科目レベルの評価をどうつなぐかという課題に対し、新潟大学歯学部での実践にもとづいて、具体的な提案を行うことである。

新潟大学は日本に86ある国立大学の一つで、歯学部は、日本の歯学教育の最先端の教育を行っていることで知られる。本研究は、新潟大学歯学部の教員と高等教育における評価の専門家との間で過去数年にわたって進められてきた評価とカリキュラムに関する実践的研究の成果の一端を示すものである。

本稿では、科目レベルの評価の例として、歯学教育プログラムの重要科目の一つであるPBL（Problem-Based Learning）科目を取り上げる。PBLの評価についてはさまざまな試みがなされてきたが（Bridges, Corbet, & Chan, 2015; Feleti & Ryan, 1994; Mtshali & Middleton, 2011; Winning, Lim, & Townsend, 2005）、ここでは、PBLのパフォーマンス評価の新しい方法として改良版トリプルジャンプ（Modified Triple Jump, MTJ）を提案する¹。また、PBL科目以外にも複数の重要科目を取り上げ、これらの科目において実践されているパフォーマンス評価がプログラムレベルの評価とどう結びつけられるのかを示す。

科目の評価でありながら、学生たちが当該のクラスで学んだことだけでなく、プログラム・機関レベルの目標をどの程度達成しているかについても情報を提供するという二重の機能をもった評価は、「埋め込み型評価（embedded assessment）」（Cummings, Maddux, & Richmond, 2008; Shavelson et al., 2008; Suskie, 2009）と呼ばれる。この概念を使えば、本稿は、「科目に埋め込まれたパフォーマンス評価（course-embedded performance assessment）」をどう配置・統合してプログラムレベルの評価を実現するかを、実践的な裏付けにもとづいて論じるものであり、そのことによって、国際的な評価研究の進展に貢献することを企図している。

Banta & Palomba (2015) によれば、プログラムレベルの評価におけるパフォーマンス指標 (performance measures) の活用において、現在、大きく分けて2つの方法がみられるという。一つは、プログラムを構成する科目のなかからサンプルとなる科目を選び、さらにそこからサンプルとなる学生のワークを選んで、教員団がプログラム評価のために二次採点 (second scoring) を行うという方法である。もう一つは、各教員による各科目での評価結果をプログラムレベルの評価のために直接、統合する (aggregate) という方法である。その代表例として、デジタル・テクノロジーを用いることにより、科目・プログラム・一般教育の評価を統合する Prince George's Community College の All-in-One アプローチ (Richman & Ariovich, 2013) が紹介されている。

1 番目の方法は、アメリカの多くの大学で用いられているが、教員の評価負担が大きくなるという問題を抱えている。これに対し、2 番目の方法は、より効率的で有望なやり方だが、まだ事例が少なく模索段階にあるといえよう。新潟大学歯学部の方法は2 番目の方法に分類されるが、All-in-One とは別のやり方を取っており、新規性がある。

以下では、まず、PBL 科目での新しいパフォーマンス評価の方法である「改良版トリプルジャンプ (Modified Triple Jump, MTJ)」について紹介する。次に、科目レベルの評価とプログラムレベルの評価の基盤となる、カリキュラムと評価のアラインメント (連動、整合性) について論じる。それをふまえて、科目レベルの評価とプログラムレベルの評価の統合の方法を提案し、最後に、プログラムレベルの評価のさまざまな方法を比較することによって、本稿で提案する方法——重要科目に埋め込まれたパフォーマンス評価 (Pivotal Embedded Performance Assessment, PEPA) ——の有効性を主張する。

2. 重要科目としての PBL とそのパフォーマンス評価

2.1. 新潟大学歯学部の PBL

新潟大学歯学部は歯学科と口腔生命福祉学科の2 学科からなる。歯学科は6 年制、口腔生命福祉学科は4 年制であり、1 学年の学生定員はそれぞれ40 名、20 名である。このうち歯学科では「変化の激しい現代社会のなかで、新たな諸課題に関係者と適切に連携しながら問題解決を図っていく能力を備え、全人的医療を実践できる高い歯科臨床能力を有する」ことを教育目標に掲げ、そのような目標を具体化するために、両学科ともに PBL を早期から積極的に導入している。また、この教育目標に直接関連するいくつかの重要科目で、ルーブリックを用いたパフォーマンス評価を実施している。

新潟大学歯学部の PBL は、スウェーデンのマルメ大学

歯学部的方式 (Rohlin et al., 1998) に準拠している。この方式では、7~8 名の学生でグループを構成して、教員のファシリテーションのもと、次のような手順で PBL を行う。

ステップ1: 授業でのグループ学習

- ①最初に、シナリオと呼ばれる事例から事実を抽出する。
- ②その事実についての疑問や考えを話し合い、シナリオに含まれる問題についての解決策 (仮説) を作る。
- ③次に学生たちは、自分たちの仮説を検証するために、どのような知識が不足しているか確認し、学習課題を設定する。

ステップ2: 授業外での個別学習

- ④その後、学生たちは授業外で個々に学習課題について調査する。

ステップ3: 授業でのグループ学習 (1 週間後)

- ⑤1 週間後、調査した結果をグループでもちより、既有知識と新しい知識を統合する。
- ⑥最初に立てた自分たちの解決策 (仮説) が妥当であったか否か検討し、最終的な解決策を作成する。

このように、PBL では、授業でのグループ学習、授業外での個別学習、授業でのグループ学習という3つのステップをたどりながら学習が進められる (小野・松下, 2015; Ono & Matsushita, 2017)。事例をもとに、グループメンバーと議論しながら、問題を解決する過程を通して知識を学ぶことから、PBL の効果として、統合された深い知識の習得、問題分析・問題解決能力の育成、対人関係能力の育成、継続した学習意欲の涵養があげられている (Barrows, 1998)。

2.2. PBL のためのパフォーマンス評価方法の開発

—改良版トリプルジャンプ—

PBL で学んだ学生の問題解決能力を評価するために、私たちは改良版トリプルジャンプを開発した。トリプルジャンプとは、PBL における問題解決能力、自己学習能力を評価するために、1975 年にカナダのマクマスター大学医学部で考案された評価方法である (Blake et al., 1995)。いわば、学生と教員が一对一で行う PBL で、通常の PBL の学習過程と同様に3つのステップからなるが、いつもは学生同士のグループ学習で行うステップ1とステップ3を、教員とのやりとりで代えて学生を評価する。具体的には、ステップ1で学生はシナリオを読み、そこに書かれた事実から問題を見出し、解決策を立案する。その際、学生は自分が必要と考える追加情報を教員に質問でき、また教員はあらかじめシナリオの追加情報を準備している。解決策を検討するために、ステップ2で学生は図書館などに出向き、信頼できる情報を収集し、自己学習を行う。その後、ステップ

3として、学生は教室に戻り、既有的知識にステップ2で得た知識を統合し、最終的な解決策を教員に説明する。

通常のPBLと同様な過程で評価が進められることから、評価の妥当性、とくに「表面的妥当性（face validity）」は高いとされ、また、さまざまな専門家が協力してシナリオを作成し、吟味することにより、「内容的妥当性（content validity）」も担保されるといわれている。しかし、他方、評価の「信頼性（reliability）」に関しては、主観的であること、学生と教員のやりとりを確認する他の評価者がいないこと、口頭でのやりとりで教員が学生の説明を聞き逃すことがあること、評価資料の質、学生のパーソナリティ、そして評価者の熟練度が影響しやすいことなどから、一般に信頼性は低いとみなされている（Mtshali & Middleton, 2011）。また、評価の「実行可能性（feasibility）」に関しても、トリプルジャンプは学生が自己学習する時間が必要な上、評価に時間がかかり、教員の評価負担が大きいことも指摘されている（Newman, 2005）。これらのことから、現在では、PBLの評価法としてトリプルジャンプを実施している大学は少ない。しかし、トリプルジャンプに代わる、妥当性があり、かつ信頼性、実行可能性を兼ね備えた評価方法はいまだ見あたらないことから、本研究ではトリプルジャンプの改良に取り組むことにした。開発にあたっては、評価を受けることが学生にとって意味のある学習経験となるよう留意した。

改良版トリプルジャンプ（Modified Triple Jump, MTJ）は、オリジナルのトリプルジャンプと同様に、ステップ1でシナリオから問題を見出し、解決策の仮説を立案し、学習課題を設定するが、その過程を60分間でワークシートに記述させる。ステップ2は学習課題を調査し学習するだけでなく、その結果をもとに解決策を検討し、最終的な解決策を提案するところまでで1週間とし、その過程もワークシートに記述させる。すなわち、改良版トリプルジャンプは、これまでのトリプルジャンプの3つのステップをステップ1・2に縮め、口頭に代えて文書で評価するものであり、さらに、その評価にあたってはルーブリックを用いることが大きな特徴である。それに加え、改良版トリプルジャンプでは、新たに設けたステップ3において、シナリオの状況を再現して教員を相手にロールプレイさせることにより、解決策の実行までを、別のルーブリックを用いて評価する。ステップ3はロールプレイのすぐ後に、評価結果のフィードバックを行い、あわせて15分間としている（図2）（小野・松下, 2015; Ono & Matsushita, 2017）。

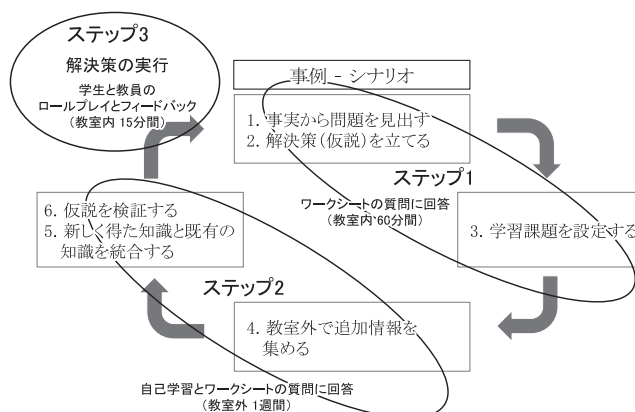


図2 PBLの学習プロセスと改良版トリプルジャンプ

より具体的に示そう。例えば「口腔と全身のかかわり」（科目名）の改良版トリプルジャンプでは、図3のシナリオを学生に与え、ワークシートによる筆記課題（ステップ1・2）とロールプレイによる実演課題（ステップ3）を学生に行わせ、教員がそれぞれを2種類のルーブリックにより評価する。ワークシート用のステップ1・2のルーブリック（表1）には、「問題発見」「解決策の着想」「学習課題の設定」「学習結果とリソース」「解決策の検討」「最終解決策の提案」の6つの観点（ディメンション）があり、問題発見から解決策の提案までをレベル3～0の4段階で評価する。ステップ3では「解決策の実行」という第7の観点を評価

「わたし、困っています」

あなたは、新潟大学病院の歯科医師です。

今日は、担当患者、高橋勇蔵（67歳・男性）の2回目の診療日です。高橋勇蔵は中等度の歯周病があり、初回は歯周検査と病状説明を行いました。

あなた：高橋さん、お口の具合はいかがですか。前回、タバコをやめるようお話しましたが、禁煙されましたか。

高 橋：してないよ。わたしはね、タバコをやめるくらいなら死んだほうがまだと思っているよ。この前、国から送られてきたアンケートにも「生きがいはタバコを吸うこと」と書いてくらいだ。相変わらず1日40本は吸っている。歯科に来て、なんでタバコをやめるよう言われなきゃならんのだね。

あなた：でも、高橋さんは糖尿病もあるし、やめたほうがいいと思いますけど…。

高 橋：糖尿病は関係なからう。ここは歯科だろう。おや、いや、内科と間違えたかな。

あなた：歯科ですけど…。とにかく、前回言ったことと同じことを言いますが、まずはタバコをやめてください、いいですね。

高 橋：ああ、わかったぞ。あんたはタバコが嫌いだな。

図3 改良版トリプルジャンプのシナリオ例

するが、これを評価するルーブリック（表2）では、「解決策の実行」の下にさらに、「追加情報の収集」「情報の統合」「共感的・受容的態度」「コミュニケーション」という下位の観点を組み込んで評価する。ステップ1・2のルーブリックは一般的ルーブリックなので、課題に取り組む前に学生に提示するが、ステップ3のルーブリックは課題特殊のルーブリックなので、学生には提示しない。

これまで口腔生命福祉学科のPBL科目で5年間（2018年現在）にわたって実施した結果、改良版トリプルジャンプにはオリジナルのトリプルジャンプに比べて以下のメリットがあることが明らかになった。まず、評価の信頼性に関しては、ステップ1・2とステップ3のそれぞれでルーブリックを用いて3名の教員で評価したところ、各観点における評価者間

の一致度は全体的に高く、十分な評価者間信頼性を得ることができた（average ICC (2, 3) = .76）（小野他, 2014）。また、評価の実行可能性については、ステップ1・2でワークシートを導入したことにより、同時に多くの学生が受験でき、教員が対面での評価の場に拘束される時間は著しく短縮された。さらに、改良版トリプルジャンプ実施後に学生アンケート調査を行った結果では、「模擬患者に理解してもらったために、しっかり調べることができ、自分の理解も深まった」「ステップ3を行うことで、PBLが将来現場に出た時に役立つものになるのだとわかった」といった肯定的な意見が数多くみられ、評価が単なる「学習の評価」であるだけでなく、それ自体が学生の学習経験にもなるような「学習としての評価」（Alverno College Faculty, 1994; Earl, 2003）とし

表1 改良版トリプルジャンプ（ステップ1・2）のルーブリック

観点	1. 問題発見～6. 最終解決策の提案					
	1. 問題発見	2. 解決策の着想	3. 学習課題の設定	4. 学習結果とリソース	5. 解決策の検討	6. 最終解決策の提案
観点的説明	シナリオの事実から、問題を見出す。	解決の目標を定め、いくつかの解決策を立案する。	問題の解決に必要な学習課題を設定する。	信頼できるリソースから、学習課題を調査する。	解決策の有効性や実行可能性を検討する。	問題に対して最終的な解決策を提案する。
レベル3	問題を見出し、シナリオの事実から、推察しうる原因を含め、問題とした理由を述べている。	いくつかの解決策を立て、これまでの学習や経験とも結びつけて、解決策の立案過程を述べている。	学習課題を的確に設定し、解決策と学習課題の関連から必要性を述べている。	利用可能なさまざまなリソースを駆使し、信頼性に注意して、正しい内容を学習している。	いくつかの解決策を比較検討し、それぞれの有効性や実行可能性を考察している。同時に、解決策の限界にも思いをめぐらしている。	シナリオの状況に適した、妥当な最終解決策を提案している。解決策をより効果的に実行するために、追加情報の必要性に気づいている。
レベル2	問題を見出し、シナリオの事実から、問題とした理由を述べている。	いくつかの解決策を立て、解決策の立案過程を述べている。	学習課題を設定し、解決策と学習課題の関連から必要性を述べているが、重要な学習課題が一部欠如している。	リソースの信頼性に注意して、おおむね正しい内容を学習している。	いくつかの解決策を比較検討し、それぞれの有効性や実行可能性を考察している。	シナリオの状況に適した、妥当な最終解決策を提案している。
レベル1	問題を見出しているが、問題とした理由の説明は不十分である。	解決策を立てているが、立案過程の説明は不十分である、あるいは解決策が1つのみである。	学習課題が漠然としており、何を学ぶべきか焦点が絞られていない、あるいは必要性の説明が不十分である。	リソースの信頼性についての注意が不十分で、学習内容にいくつかの誤りが含まれている。	解決策の検討は不十分である、あるいは複数の解決策について比較検討していない。	最終解決策の提案にいたっていない、あるいは解決策、学習結果、結論の間に矛盾や飛躍がある。
レベル0	レベル1を満たさない場合はゼロを割り当てること。					

表2 改良版トリプルジャンプ（ステップ3）のルーブリック

観点	7. 解決策の実行			
	7-1. 追加情報の収集 (追加情報の収集と問題の再把握)	7-2. 情報の統合 (追加情報の統合と解決策の内容修正)	7-3. 共感的・受容的態度 (相手への共感と受容)	7-4. コミュニケーション (相手にあわせて解決策の表現)
観点的説明	禁煙を働きかける上で必要となる追加情報を患者とのやりとりを通じて収集し、必要に応じて問題の再把握を行う。	禁煙を働きかける上で有用な情報を追加情報も入れて統合し、必要に応じて解決策の内容修正を行う。	患者の考えや価値観に配慮して禁煙を働きかける。	自分の考えを患者にわかりやすく説明する。
レベル3	歯周治療に対する患者のニーズや、なぜ「生きがいはタバコを吸うこと」なのか、糖尿病の病状など、禁煙を働きかける上で必要な追加情報をすべて、的確に収集している。	患者からの追加情報も統合し、歯周治療における禁煙の重要性を、歯周病と喫煙の関係のみならず、歯周病と糖尿病の関係とも結びつけて深く柔軟に理解している。	「生きがいはタバコを吸うこと」という患者の考えを受入れ、相手の気持ちに配慮して、患者の反応をみながら禁煙を働きかけている。	内容とその関連から、話の順序や組み立てを考え、平易な言葉で、相手の理解を意識しながら説明している。
レベル2	歯周治療に対する患者のニーズや、なぜ「生きがいはタバコを吸うこと」なのか、糖尿病の病状など、禁煙を働きかける上で必要な追加情報を、ある程度収集している。	患者からの追加情報を一部統合し、歯周治療における禁煙の重要性を、歯周病と喫煙の関係のみならず、歯周病と糖尿病の関係とも結びつけて適切に理解している。	「生きがいはタバコを吸うこと」という患者の考えを受入れ、相手の気持ちに配慮して、禁煙を働きかけている。	話の順序や組み立てはおおむね整っているが、相手の理解を得る上で話の構成にやや問題がみられる。
レベル1	歯周治療に対する患者のニーズや、なぜ「生きがいはタバコを吸うこと」なのか、糖尿病の病状など、禁煙を働きかける上で必要な追加情報のごく一部を収集している。	歯周治療における禁煙の重要性を、歯周病と喫煙の関係のみから理解している、あるいは歯周病と糖尿病に関する知識はあるが、禁煙指導と結びついていない。	患者の考えを認めているが、それに対する配慮が不十分である。	話の順序や組み立てが混乱しており説明が理解しにくい、あるいは事前に準備した内容を読み上げているだけである。
レベル0	レベル1を満たさない場合にはゼロを割り当てること。			

(注) 改良版トリプルジャンプ（ステップ3）のルーブリックは課題特殊的で、ルーブリックの記述語はシナリオの内容に依存する。ここでは、図3のシナリオのルーブリックを示す。

て機能していた（小野・松下，2015；Ono & Matsushita，2017）。以上から、改良版トリプルジャンプはPBLのためのすぐれた評価方法であるといえることができる。

3. カリキュラムと評価のアラインメント

3.1. カリキュラムと評価のアウトライン

新潟大学歯学部歯学科では、先にあげた教育目標の下に、より具体的な目標が、24項目（「知識・理解」に関する7項目、「専門的能力（分野固有の能力）」に関する6項目、「汎用的能力」に関する8項目、「態度・姿勢」に関する3項目）にわたって設定されている²。これらは、歯学教育プログラムにおける「目標としての学習成果」である。

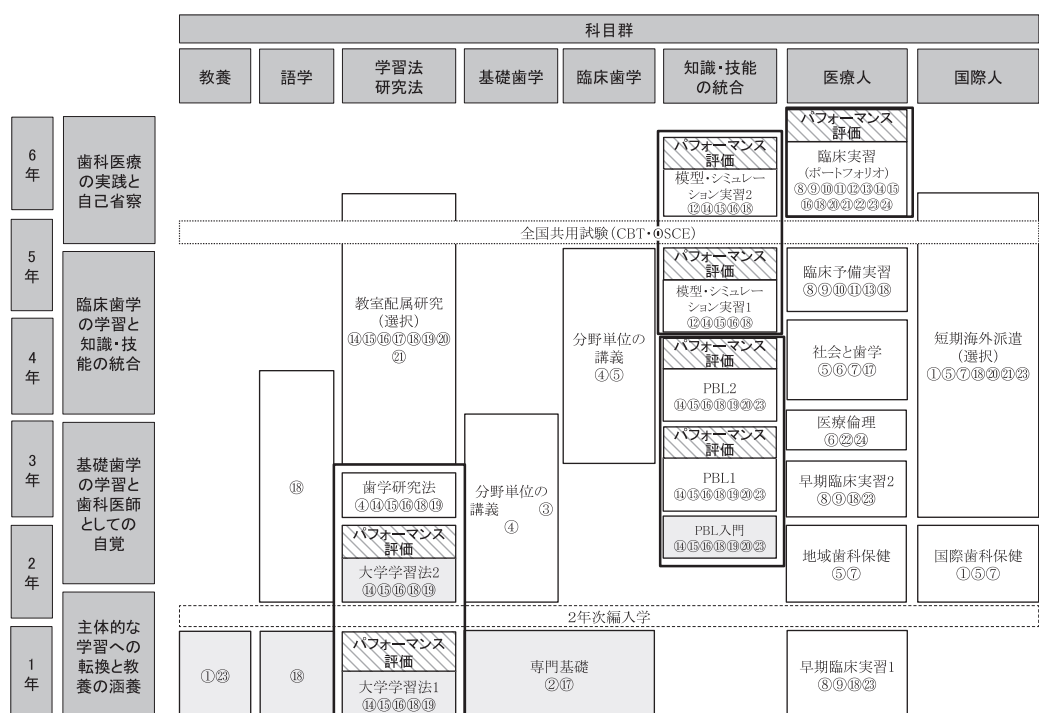
ちなみに、新潟大学では、「課題を発見し、それを解決する能力」「課題解決に必要な知識・技能を主体的に学修する能力」「課題に協働的に取り組むためのコミュニケーション能力」を、新潟大学のすべての卒業生が身につけるべき能力として掲げており、各学位プログラムでは、それらを含め、目標としての学習成果を、「知識・理解」「専門的能力（分野固有の能力）」「汎用的能力」「態度・姿勢」に区分して表すこととしている。歯学教育プログラムの目標としての学習成果は、全学の方針にしたがって定められており、また一方で、日本のすべての歯科大学・歯学部が共通して取り組むべき「コア」として策定された「歯学教育モデル・コア・カリキュラム」（モデル・コア・カリキュラム改訂に関する連絡調整委員会およびモデル・コ

ア・カリキュラム改訂に関する専門研究委員会，2017）における「歯科医師として求められる基本的な資質・能力」とも整合性を有している。つまり、新潟大学歯学部歯学科の目標は、一方で機関の目標と、もう一方で機関を越えた当該分野の目標と整合性をもつように設定されたものである。

このような能力を身につけさせるために、目標としての学習成果をカリキュラムマップにより各科目と結びつけ、一般教育を含む歯学教育プログラムを図4のように編成している。このカリキュラムツリーで示すように、プログラムは6年制で、おおむね第1学年から第2学年前期、第2学年後期から第3学年、第4学年から第5学年前期、第5学年後期から第6学年と、学習内容の面から大きく4期に分けられる。また、科目は、「教養」「語学」「学習法・研究法」「基礎歯学」「臨床歯学」「知識・技能の統合」「医療人」「国際人」の8つの科目群から構成される。

第1期は「主体的な学習への転換と教養の涵養」を重視する。「大学学習法」により学習態度の転換を図り、本プログラムを履修していく上で必須な問題解決能力、論理的思考力、表現力を育成し、パフォーマンス評価により学習成果を評価する。

第2期は「基礎歯学の学習と歯科医師としての自覚」が中心で、基礎歯学の科目を講義形式で開講し、知識を習得させるとともに、「早期臨床実習（アーリー・エクスポージャー）」で患者とのふれあいを通して医療人としての自覚と態度を涵養する。また、講義で得た知識を統合し、



（注）□ は一般教育科目を、□ は専門教育科目を表す。また、枠内の番号は表3の番号で、それぞれの科目における主な目標としての学習成果を示している。

図4 カリキュラムと評価のアウトライン（新潟大学歯学部歯学科）

表3 歯学教育プログラムにおける領域特殊な問題解決能力の育成

問題解決能力	歯科臨床能力	大学学習法 (相手なし)	PBL (ペーパー・ペイシエント ／模擬患者)	模型・シミュレーション 実習 (模型)	臨床実習 (本物の患者)
問題の定義	情報収集・分析	✓	✓	✓	✓
	診断	✓	✓	✓	✓
方略の同定	治療方針の決定	✓	✓	✓	✓
解決／仮説の提案	治療計画の立案	✓	✓	✓	✓
採りうる解決の評価		✓	✓	✓	✓
解決の実行	治療の実践		改良版トリプルジャンプ ステップ3	✓	✓
結果の評価	治療結果の評価と 治療計画の見直し		改良版トリプルジャンプ ステップ3	教員からのコメント・ フィードバック	✓

→ 専門性・総合性・真正性

問題解決能力を育成するために、講義と並行して「PBL」を実施し、パフォーマンス評価、すなわち、先述した改良版トリプルジャンプにより学習成果を評価する。

第3期は「臨床歯学の学習と知識・技能の統合」が中心となり、臨床歯学の科目を講義形式で開講する。それに並行して、第2期に引き続きPBLを、また新たに「模型・シミュレーション実習」を実施し、基礎歯学を含め、知識と技能を統合させ、より専門性を高めた形で問題解決能力を育成する。そして、その学習成果をパフォーマンス評価により評価する。

第4期は「歯科医療の実践と自己省察」の期間である。まず、医療系大学間共用試験実施評価機構（Common Achievement Tests Organization, CATO）が行う全国共用試験により、知識はCBT（Computer-Based Testing）を用いて、また技能・態度はOSCE（Objective Structured Clinical Examination, 客観的臨床能力試験）を用いて、臨床実習を行えるだけの資質・能力を身につけているかを評価する。この試験を経て、学生は「臨床実習（Clinical Practicum）」への参加を認められ、患者診療の経験により、歯科臨床能力を向上させる。歯科臨床能力の評価は、継続的にポートフォリオを用いて形成的に行い、臨床実習終了時に患者診療を直接評価するパフォーマンス評価を実施する。

3.2. 重要科目におけるパフォーマンス評価

このカリキュラムのなかで、各期における「重要科目」として位置づけられているのが、一般教育として提供されている「大学学習法」と、歯学部専門科目である「PBL」、「模型・シミュレーション実習」、「臨床実習」である（図4）。これらの重要科目は、既存の知識・技能を

統合することにより、学科の目標である汎用的な問題解決能力、および領域特殊な問題解決能力である歯科臨床能力を形成することが求められる科目である。表3において、「問題解決能力」にあげた項目は、全米大学カレッジ協会（Association of American Colleges & Universities, AAC&U）の「問題解決 VALUE ルーブリック」（Rhodes, 2010）の各観点であり、「歯科臨床能力」が歯科という領域における問題解決能力であることを示している。「大学学習法」（第1段階）から、「PBL」（第2段階）、「模型・シミュレーション実習」（第3段階）、「臨床実習」（第4段階）へと進むにつれて、問題解決能力がより専門的になり、問題解決能力の観点をより多く包含するよう総合性を高め、紙の上での問題解決から本物の患者に向き合う臨床場面へとより真正性を増していくことが見てとれよう。

これらの重要科目においては、以下のような種類の異なるパフォーマンス評価を実施する（各評価のルーブリックについては、後掲の表4も参照）。

(1) 大学学習法におけるパフォーマンス評価（第1期）

「大学学習法」では、レポート課題を学生に与え、そこから問題解決能力、論理的思考力、文章表現力を評価する（松下他, 2013; 小野・松下, 2016）。教員からレポート課題として大きなテーマが提示され、学生は多くの問題のなかから、レポートで扱う具体的な問題を設定する。そして、設定した問題に対して、調べてまとめるだけでなく、学生自身の主張と結論を述べる。教員は、「背景と問題」「主張と結論」「論拠と事実・データ」「対立意見の検討」「全体構成」「表現ルール」の6つの観点と、3～0の4段階のレベルからなるルーブリックを用いてレポートを評価し、「背景と問題」「主張と結論」から問題解決能力を、「主張と結論」「論拠と事実・データ」「対立意見

の検討」「全体構成」から論理的思考力を、「全体構成」「表現ルール」から文章表現力を把握する。

（2）PBLにおけるパフォーマンス評価（第2期）

先述したように、改良版トリプルジャンプにより、シナリオにもとづくペーパー・ペイシエントについての筆記課題、模擬患者に対するロールプレイという2種類のパフォーマンスを2つのルーブリックを用いて評価する。

（3）模型・シミュレーション実習におけるパフォーマンス評価（第3期）

「模型・シミュレーション実習」では歯科臨床能力を評価する。この科目は、口腔内を再現した模型、患者シナリオ、レントゲン写真や検査所見から、患者の問題を見出し、妥当な治療方針・治療計画を立案して、それを模型上で実践し、その結果を評価して治療計画を見直すというものであり、学生がそれぞれの内容を記載したワークシートを、教員がルーブリックによりレベル3～0の4段階で評価する。ルーブリックには、「歯科情報の収集・分析と歯科疾患の診断（病的所見と診断）」「患者情報を統合した治療方針の決定（治療方針の決定）」「治療方針にそった治療計画の立案（治療計画の立案）」「治療の実践と治療計画の再検討（治療実践後の振り返り）」の4つの観点に加え、患者へのインフォームドコンセントを意識した「歯科専門家としての表現（専門用語と表現）」という観点を含めている（秋葉他, 2017）。

（4）臨床実習におけるパフォーマンス評価（第4期）

患者診療を行う「臨床実習」では、歯科臨床能力を評価するために、形成的評価としてポートフォリオ評価を、総括的評価として臨床パフォーマンス評価を行う。

ポートフォリオ評価では、学生は、診療のたびに、「今回の診療での自分の目標」を立て、実習後に「処置・治療

内容の詳細」「診療で学んだ知識や専門的スキル」「診療で学んだ医療者としての態度・姿勢」をeポートフォリオに記載する。また、診療の到達度を、どの程度自分自身で行えたかをもとに5段階で自己評価する。そして、「今回の診療での問題と今後に向けた解決策・自己学習課題」を記入する。それに対して学生を指導した教員は、教員からみた診療の到達度を同じく5段階で評価し、「教員からのコメント・追加学習の指示」を与える（小田他, 2017）。

一方、臨床パフォーマンス評価は、臨床実習が終了に近づき、学生から受験申込みのあった患者診療を、「チェアサイド」において（つまり、診療室内で患者さんの座っている椅子のそばで）、実習担当教員である専門診療科の歯科医師が評価する。このパフォーマンス評価では、「処置の根拠となる所見の聴取・収集ができていたか」「診断にもとづいて適切な処置を選択したか」「必要な医療器具・歯科材料を準備し、使用上の注意を守って扱っていたか」「行った処置を適正に自己評価できたか」「患者への配慮がみられたか」「医療安全に対する配慮がみられたか」の6項目について、「できた」「一部できなかったが許容範囲内（許容内）」「一部できず許容範囲外（許容外）」の3段階で評価する（藤井他, 2017）。

以上述べてきたように、「大学学習法」では対象となる相手はいないが、「PBL」、「模型・シミュレーション実習」、「臨床実習」では、それぞれ、シナリオにもとづくペーパー・ペイシエントとその模擬患者、模型、本物の患者を対象とした実演という形で行われるパフォーマンスを、教員団の開発したルーブリックによって評価する。「ミラーのピラミッド（Miller's Pyramid）」（Miller, 1990）を使って表現すれば、「PBL」、「模型・シミュレーション実習」、「臨床実習」の評価は、図5に示すように高次化していくとい

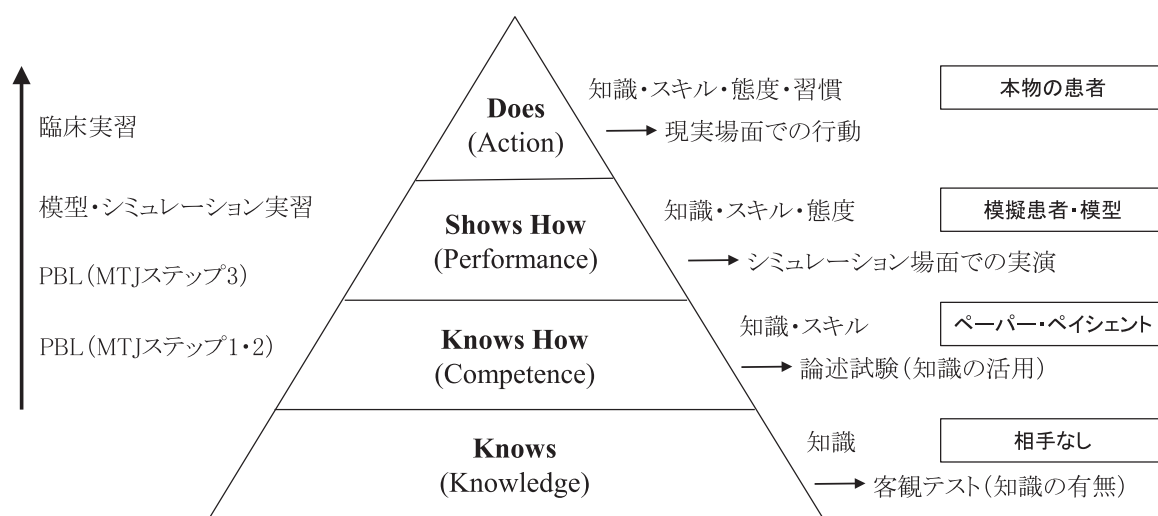


図5 重要科目のパフォーマンス評価の位置づけ

ともできる。PBL のなかでもペーパー・ペイシェントに対する筆記課題として行われる MTJ ステップ 1・2 は Knows How (knows how to apply acquired knowledge: 獲得した知識の適用の仕方を知っている) に、ロールプレイとして行われる MTJ ステップ 3 や模型・シミュレーション実習は Shows How (shows how to apply that knowledge: その知識の適用の仕方を見せる) に、臨床実習は Does (actually applies that knowledge in practice: 実践においてその知識を実際に適用する) に対応した課題を設定している。このように、種類の異なるパフォーマンス評価を重ねることによって、より高次の歯科臨床能力が評価できるようになると考えられる。

では、こうした科目レベルのパフォーマンス評価は、どのようにしてプログラムレベルの評価につながることができているのだろうか。それらは、実際に、埋め込み型評価として機能し、科目レベルだけでなくプログラムレベルの目標の達成度合いについても情報を提供できているのだろうか。

4. 科目レベルの評価とプログラムレベルの評価の関係

4.1. 埋め込み型のパフォーマンス評価による質保証

表 4 は、埋め込み型のパフォーマンス評価 (embedded performance assessment) による質保証のイメージを表したものである。重要科目のうち、「大学学習法」「PBL」「模型・シミュレーション実習」はルーブリックのパフォーマンス・レベルのうち「レベル 2」以上を合格とし、「臨床実習」は、「できた」「一部できなかったが許容範囲内 (許容内)」「一部できず許容範囲外 (許容外)」のうち「一部できなかったが許容範囲内 (許容内)」以上を合格とする。ただし、前三者についてはいずれも、「PBL1」「PBL2」のように、学期・学年をまたいで 2 つの科目があるので、各科目の「1」において合格基準に達していなくても、続く「2」においてレベル 2 以上を達成できればよい (例えば、1 年前期の「大学学習法 1」では「全観点レベル 2 以上」を達成できなくても、2 年前期の「大学学習法 2」では達成できたなど)。また、「臨床実習」でも、繰り返しさまざまな課題で臨床経験を積み、ポートフォリオによる日々の評価・指導を受けることにより (そのプロセスは e ポートフォリオに記録される)、最終的な臨床パフォーマンス評価の合格水準を超えられればよい。このように、いずれの重要科目のパフォーマンス評価においても形成的評価の機能が内包されており、その科目を修了する段階では合格基準を達成できるようにデザインされている。

もちろん、図 4 に示すように、歯学教育プログラムには、これらの重要科目以外にも数多くの科目が含まれている。これらの科目の多くは、「知識」あるいは「知識・スキル」

表 4 埋め込み型パフォーマンス評価による質保証のイメージ

初年次	大学学習法 1・2				
	観点	レベル 3	レベル 2	レベル 1	レベル 0
問題解決能力 ／ 歯科臨床能力	背景と問題		2	1	
	主張と結論		2	1	
	論拠と事実・データ	2	1		
	対立意見の検討		2	1	
	全体構成		2	1	
	表現ルール	2		1	
	PBL 1・2				
	問題発見		2	1	
	解決策の着想		2	1	
	学習課題の設定		2	1	
最終年次	学習結果とリソース	2	1		
	解決策の検討		2	1	
	最終解決策の提案		2	1	
	解決策の実行		2	1	
	模型・シミュレーション実習 1・2				
	病的所見と診断	2	1		
	治療方針の決定		2	1	
	治療計画の立案		2	1	
	治療実践後の振り返り		2	1	
	専門用語と表現	2	1		
	臨床実習				
	観点	できた	許容内	許容外	
	所見の聴取・収集				
	診断と処置の選択				
	器材の準備と使用				
	処置後の振り返り				
	患者への配慮				
	医療安全				

(注) ルーブリック内の「1」「2」は科目の番号を表し、のセルは、質が担保されたとみなされる卒業生の例を示す。

(図 5 参照) の獲得・形成をめざすものであり、そこでの評価は、個々の教員の「エキスパート・ジャッジメント」³に委ねられ、そのもとで単位の認定が行われる。

こうして、重要科目で合格水準に達し、カリキュラム全体を通じて必要な単位を修得した学生が学士課程の修了を認められることになる。また、学士課程修了により、歯科医師国家試験の受験資格が与えられる。

なお、日本では、6 年間の学士課程修了後に国家試験に合格し、診療に従事する歯科医師は、大学病院や厚生労働大臣が指定する病院などにおいて、1 年間以上の臨床研修を受けることが歯科医師法により義務づけられている。臨床パフォーマンス評価は、もともと臨床研修における評価方法として開発された経緯もあり、学士課程修了時ではなく臨床研修修了時において、すべての項目でパフォーマンス・レベルの「できた」に達することを想定して設計されている。

4.2. プログラムレベルの評価の比較

冒頭に述べた学習成果の評価の4タイプ（図1）のうち、従来、プログラムレベルの評価に用いられてきたのは、タイプⅡの学生調査、タイプⅢの客観テスト、タイプⅣのポートフォリオ評価やキャップストーン科目でのパフォーマンス評価であった（Ewell et al., 2011; Suskie, 2009）。表5では、それぞれの評価方法の特徴を、評価の妥当性・信頼性・実行可能性、および単位制度との親和性という観点から整理したものである。

まず、質問紙調査については、直接評価を代替することは困難であることから（Pike, 1996; 斎藤他, 2016）、ここでの検討からは除外される。追加型アセスメント（add-on assessment）としてプログラムの終了時などに実施される客観テストは、評価の実行可能性は高いが、必ずしも知識の統合や高次の能力の評価には適しておらず、また、単位制度との親和性も高くない。ポートフォリオ評価では、教員団による二次採点が必要となり、評価の実行可能性が低くなってしまうおそれがある（Banta & Palomba, 2015）。すべての科目でパフォーマンス評価を行う場合は、二次採点の必要はなく、単位制度との親和性も高いが、その評価負担は膨大なものになるだろう。

以上の4つの方法に対して、重要科目のみでパフォーマンス評価を行うという方法は、当該科目での評価負担は大きい、科目は限定されているため、評価の実行可能性は比較的高くなり、また、プログラムの目標としている知識の統合や高次の能力を直接評価できるという強みがある。キャップストーン科目でのパフォーマンス評価はその典型例といえよう。しかしながら、キャップストーン科目だけではプログラム全体をカバーできない。

そこで本研究では、重要科目のみをパフォーマンス評価

の対象にするとともに、その他の科目については個々の教員のエキスパート・ジャッジメントに委ねることによって、プログラム全体をカバーしながら、科目レベルの評価とプログラムレベルの評価を結びつける方法を提案した。重要科目だけを科目に埋め込まれたパフォーマンス評価によって行うというこの方法を、「重要科目に埋め込まれたパフォーマンス評価（Pivotal Embedded Performance Assessment, PEPA）」と呼ぼう。

この方法は、評価負担の大きいパフォーマンス評価の対象を、複数科目で修得した知識の統合や高次の能力を要求される重要科目のみに限定し、それをカリキュラムの各段階（新潟大学歯学部の場合は4段階）の結節点に配置することによって、評価の妥当性を確保しつつ、評価の実行可能性や単位制度との親和性も維持しようとするものである。なお、評価の信頼性（とくに評価者間信頼性）は、パフォーマンス評価において教員団が協働でルーブリックを開発し、複数の担当教員でキャリブレーションとモデレーションを含む評価を行うことによって担保されている（Lane, 2014; Rhodes & Finley, 2013）。このように、プログラムレベルの目標に直結する重要科目をカリキュラム上に系列的に配置した上で、そこでの教員団によるパフォーマンス評価とその他の科目での個々の教員による知識・スキルの評価を組み合わせるというのが、本研究において提案するプログラムレベルの評価の方法である。

先述したように、日本の歯学教育プログラムでは、学生は、臨床実習開始前には標準化された共用試験（CBTとOSCE）、学士課程修了後には歯科医師国家試験を受けることになっており（図4）、このような外部基準によってベンチマーキングを行うことも可能である。ただし、共用試験のCBTや国家試験は個別の知識と思考スキルを問う客

表5 プログラムレベルの評価の比較

タイプ	プログラムレベルの評価	評価の妥当性	評価の信頼性	評価の実行可能性	単位制度との親和性
Ⅱ	質問紙調査 (間接評価/量的評価)	直接評価の代替は困難である	高い	高い	なし
Ⅲ	客観テスト (直接評価/量的評価)	事実的な知識の評価に適している	高い	高い	評価結果とそれまで取得した単位との関係について整理が必要になる場合がある
Ⅲ・Ⅳ	標準化された共用試験 (直接評価/量的・質的評価)	CBTは事実的な知識、OSCEは基本的な臨床能力の評価に適している (評価課題や評価基準に関して大学間で調整が必要)	CBTは高いが、OSCEでは評価者トレーニングが必要である	大学間共用試験実施評価機構などの組織が必要で、実施にあたって大学の負担は大きい	評価結果とそれまで取得した単位との関係について整理が必要になる場合がある
Ⅳ	ポートフォリオとスタンダード (直接評価/質的評価)	一定期間の学習や成長の評価に適している	評価者トレーニングが必要である	評価項目数により評価の負担は増減する	評価結果とそれまで取得した単位との関係について整理が必要になる場合がある
Ⅳ	すべての科目でのパフォーマンス評価 (直接評価/質的評価)	知識の統合や高次の能力の評価に適しており、プログラム全体をカバーできる	評価者トレーニングが必要である	低い	高い
Ⅳ	重要な科目でのパフォーマンス評価 (直接評価/質的評価)	知識の統合や高次の能力の評価に適しており、プログラム全体をかなりの程度カバーできる	評価者トレーニングが必要である	比較的高い	高い

観テストであり、知識の統合や高次の能力を評価することは難しい。OSCEもシミュレーション場面でのスキル・態度の評価であり、臨床実習で育成されるような能力全体を評価できるわけではない。

一方、本研究でのパフォーマンス評価は、「大学学習法」から、「PBL」、「模型・シミュレーション実習」、「臨床実習」へと進むなかで、一般的な問題解決能力から、その領域特殊な形としての歯科臨床能力へと専門性を高めていくように配置され、また、育成・評価する問題解決能力の範囲はより総合的になり、その評価場面もより高い真正性をもつようデザインされている（表3・図5）。「臨床実習」の最後のパフォーマンス評価後の質問紙調査では、「知識だけでは足りない臨床能力を測ってもらえることができる」「評価を受けることによって改善点を認識することができる」「OSCE、CBTだけでは不十分と思う」といった肯定的な回答が多くみられた（藤井他, 2017）。

表4に示すように、最後の「臨床実習」では、「できた」「許容内」「許容外」の3段階のうち、「許容内」に達せば合格としており、「できた」は臨床研修でめざすレベルを示している。これは、学士課程修了後の臨床研修にまで続く学生の成長をこのルーブリックが捉えようとしていることを示している。それは、学生にも、自分たちの学びが生涯にわたって持続されるべきものであることを伝えるメッセージになっているといえよう。

5. 結論

すぐれた科目レベルの評価を備えていても、プログラムレベルの評価を実施することは困難である。本稿では、複数の科目で開発したパフォーマンス評価を土台としながら、科目レベルとプログラムレベルの評価をつなぐ方法として、「重要科目での埋め込み型パフォーマンス評価（Pivotal Embedded Performance Assessment, PEPA）」を提案した。PEPAは、カリキュラムに埋め込まれた評価の考え方を採用し、それを重要科目におけるパフォーマンス評価の形で具体化したものである。PEPAは以下のような手続き群からなる。

- ①カリキュラムを体系化し、プログラムの教育目標と各科目の関係を明らかにするとともに、カリキュラムをいくつかの部分に分節化する。
- ②分節化した各部分に含まれるさまざまな知識を統合し、高次の能力を育成することがめざされる——プログラムの教育目標に直結するような——重要科目を抽出し、そこではパフォーマンス評価を実施する。パフォーマンス評価の課題のデザインやルーブリックの作成、評価の実施は教員団で行う。こうして開発された種類の異なるパフォーマンス評価は、学生の学習が進むにつれ

て、専門性・総合性・真正性が高まるように系列化される。一方、それ以外の個々の科目の評価は個々の担当教員のエキスパート・ジャッジメントに委ねる。

- ③重要科目でのパフォーマンス評価については、ルーブリックの各観点で一定のレベル以上のパフォーマンスを示すことで合格とする。ただし、いずれの重要科目でも、パフォーマンス評価に形成的評価機能が内包されており、その科目を修了する段階では合格基準を達成できるようにデザインされている。また、どのルーブリックも当該科目だけでなく、科目修了後も長期的に使えるようになっており、学習プログレッションの方向性を指し示すものとなっている。
- ④重要科目でのパフォーマンス評価において合格基準に達すること、各科目で所定の単位を修得することによって、そのプログラムの修了が認められる。

PEPAと名づけられたこの方法は、評価負担の大きいパフォーマンス評価の対象を、知識の統合や高次の能力を要求される——したがってカリキュラムの結節点に位置づけられる——重要科目のみに限定し、その評価を教員団でデザイン・実施することによって、評価の妥当性と信頼性を確保しつつ、評価の実行可能性や単位制度との親和性も維持しようとするものである。「カリキュラムに埋め込まれたパフォーマンス評価（curriculum-embedded performance assessment）」という考え方は新しいものではないが（Cummings et al., 2008）、私たちのPEPAは、カリキュラムとのアラインメントにより、どんな科目を重要科目とみなして埋め込み型アセスメント（embedded assessment）にするのかを明確化した点、また、それらのパフォーマンス評価が形成的機能をもちつつ系列的に配置されるようにした点にオリジナリティがある。

本稿では、日本の大学の歯学教育プログラムを事例に取って説明してきた。PEPAは、科目レベルとプログラムレベルの学習成果の評価をつなぐ強力な方法であると私たちは考えている。もっとも、PEPAの概念や手続きが他の学問分野で有効に機能するかどうかは、現時点ではまだ明らかではない。歯学教育とカリキュラム構造が似ている医学教育や薬学教育にはおそらく適用できるだろう。さらにいえば、「ミラーのピラミッド」で示すようにKnowsからKnows How、Shows How、Doesという形で、求められる認知と行動が高次化していく分野——例えば、教員養成や法曹養成など——にも適用できる可能性がある。今後の課題は、PEPAの適用範囲とその潜在的な制約を探ることである。

注

¹ Bridges et al. (2015) では、“modified ‘Triple Jump’

assessment” という方法が提案されているが、これは、学生の既有知識を確かめるために最初の「ジャンプ」でコンセプトマップを使うというもので、本研究の MTJ とはまったく異なる。

² 新潟大学歯学部歯学科のディプロマ・ポリシー (<https://www.niigata-u.ac.jp/information/2020/68104/>) 参照。

³ 「エキスパート・ジャッジメント」は、深堀他 (2020) によれば、「大学教育一般や特定の学問分野において共有されている参照基準を参考にして設定されたプログラム・レベルの学修成果を、授業科目の中で扱う学問分野の知識・能力に具体化するとともに、その達成度を適切に評価することのできる、大学教員の判断力を指す」(p. 63) と定義されている。本稿では、とくに後半部の「その達成度を適切に評価することのできる、大学教員の判断力」に重きを置いている。

謝辞

本研究は、JSPS 科研費 JP15H03473、JP18H00975 の助成を受けたものである。調査に協力してくださった新潟大学歯学部の教員・学生みなさんに感謝します。

付記

本稿は、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 (doi.org/10.18543/tjhe-6(1)-2018pp111-142) を翻訳したものである。ただし、一部、加筆修正を行った。

引用文献

- 秋葉奈美・長澤麻沙子・小野和宏・前田健康・魚島勝美 (2017). 「新潟大学歯学部における統合型模型実習の取り組み」『日本歯科医学教育学会雑誌』33, 106–114.
- Alverno College Faculty (1994). *Student assessment-as-learning at Alverno College*. Milwaukee, WI: Alverno College Institute.
- Banta, T. W., & Palomba, C. A. (2015). *Assessment essentials: Planning, implementing, and improving assessment in higher education* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Barrows, H. S. (1998). The essentials of problem-based learning. *Journal of Dental Education*, 62, 630–633.
- Blake, J. M., Norman, G. R., & Smith, E. K. (1995). Report

card from McMaster: Student evaluation at a problem-based medical school. *The Lancet*, 345, 899–902.

- Bridges, S. M., Corbet, E. F., & Chan, L. K. (2015). Designing problem-based curricula: The role of concept mapping in scaffolding learning for the health sciences. *Knowledge Management & E-Learning*, 7(1), 119–133.
- Cummings, R., Maddux, C. D., & Richmond, A. (2008). Curriculum-embedded performance assessment in higher education: Maximum efficiency and minimum disruption. *Assessment & Evaluation in Higher Education*, 33(6), 599–605.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Ewell, P., Paulson, K., & Kinzie, J. (2011). *Down and in: Assessment practices at the program level*. Champaign, IL: National Institute for Learning Outcomes Assessment. (<http://www.learningoutcomeassessment.org/documents/NILOAsurveyreport2011%20-%20Down%20and%20In%2010-20.pdf>) (2018 年 3 月 6 日)
- Feleti, G., & Ryan, G. (1994). The triple jump exercise in inquiry-based learning: A case study. *Assessment & Evaluation in Higher Education*, 19(3), 225–234.
- 藤井規孝・竹中彰治・多部田康一・佐藤直子・秋葉奈美・小田陽平・勝美祐二・小野和宏・前田健康 (2017). 「新潟大学歯学部臨床実習における臨床能力評価」『日本歯科医学教育学会雑誌』33, 4–11.
- 深堀聰子・松下佳代・中島英博・佐藤万知・田中一孝・畑野快・斎藤有吾 (2020). 「学修成果アセスメント・ツール活用支援を通じたエキスパート・ジャッジメントの涵養と大学組織の変容—先駆的事例の分析—」『大学教育学会誌』41(2), 62–66.
- Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. L. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment. (<http://www.learningoutcomeassessment.org/documents/2013%20Abridged%20Survey%20Report%20Final.pdf>) (2015 年 11 月 23 日)
- Lane, S. (2014). Performance Assessment: The State of the Art. In L. Darling-Hammond, & F. Adamson (Eds.), *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning* (chap. 5). San Francisco, CA: Jossey-Bass, Kindle.

- 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.
- 松下佳代 (2016). 「アクティブラーニングをどう評価するか」松下佳代・石井英真 (編)『アクティブラーニングの評価』東信堂, 3-25.
- 松下佳代 (2017). 「学習成果とその可視化」『高等教育研究』20, 93-112.
- 松下佳代・小野和宏・斎藤有吾・白川優治 (2014). 「学士課程教育における共通教育の質保証—直接評価と間接評価の開発と統合について—」『大学教育学会誌』36(2), 17-21.
- 松下佳代・小野和宏・高橋雄介 (2013). 「レポート評価におけるルーブリックの開発とその信頼性の検討」『大学教育学会誌』35(1), 107-115.
- Middle States Commission on Higher Education (2007). *Student learning assessment: Options and resources* (2nd ed.). (https://www.msche.org/publications/SLA_Book_0808080728085320.pdf) (2017年1月21日)
- Miller, G. E. (1990). The assessment of clinical skills/competence/ performance. *Academic Medicine*, 65(9, Suppl.), S63-S67.
- 文部科学省 (2017). 『大学における教育内容等の改革状況について』(http://www.mext.go.jp/a_menu/koutou/daigaku/04052801/_icsFiles/afieldfile/2017/12/13/1398426_1.pdf) (2018年3月6日)
- Mtshali, N. G., & Middleton, L. (2011). The triple jump assessment: Aligning learning and assessment. In T. Barrett, & S. Moore (Eds.), *New approaches to problem-based learning: Revitalising your practice in higher education* (pp. 187-200). New York, NY: Routledge.
- モデル・コア・カリキュラム改訂に関する連絡調整委員会およびモデル・コア・カリキュラム改訂に関する専門研究委員会 (2017). 『歯学教育モデル・コア・カリキュラム (平成28年度改訂版)』. (https://www.mext.go.jp/component/a_menu/education/detail/_icsFiles/afieldfile/2018/06/15/1325989_29_02.pdf) (2018年3月6日)
- Newman, M. J. (2005). Problem based learning: An introduction and overview of the key features of the approach. *Journal of Veterinary Medical Education*, 32, 12-20.
- 小田陽平・小野和宏・藤井規孝・小林正治・前田健康 (2017). 「診療参加型歯科臨床実習におけるweb公開型電子ポートフォリオの開発と運用」『日本歯科医学教育学会雑誌』33, 65-73.
- 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, 215-240.
- 小野和宏・松下佳代 (2016). 「初年次教育におけるレポート評価」松下佳代・石井英真 (編)『アクティブラーニングの評価』東信堂, 26-43.
- Ono, K., & Matsushita, K. (2017). PBL tutorial linking classroom to practice: Focusing on assessment as learning. In K. Matsushita (Ed.), *Deep active learning: Toward greater depth in university education* (pp. 183-206). Singapore: Springer.
- 小野和宏・松下佳代・斎藤有吾 (2014). 「PBLにおける問題解決能力の直接評価—改良版トリプルジャンプの試み—」『大学教育学会誌』36(1), 123-132.
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco, CA: Jossey-Bass.
- Pike, G. R. (1996). Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education*, 37(1), 89-114.
- Rhodes, T. (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges & Universities.
- Rhodes, T., & Finley, A. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Washington, DC: Association of American Colleges & Universities.
- Richman, W. A., & Ariovich, L. (2013). *All-in-One: Combining grading, course, program, and general education outcomes assessment*. Champaign, IL: National Institute for Learning Outcomes Assessment. (<http://learningoutcomesassessment.org/documents/Occasional%20Paper%2019%20FINAL.pdf>) (2017年12月8日)
- Rohlin, M., Peterson, K., & Svensäter, G. (1998). The Malmö model: A problem-based learning curriculum in undergraduate dental education. *European Journal of Dental Education*, 2, 103-114.
- 斎藤有吾・小野和宏・松下佳代 (2016). 「パフォーマンス

- 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』40(Suppl.), 157–160.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314. (doi: 10.1080/08957340802347647)
- Suskie, R. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Winning, T., Lim, E., & Townsend, G. (2005). Student experiences of assessment in two problem-based dental curricula: Adelaide and Dublin. *Assessment & Evaluation in Higher Education*, 30(5), 489–505.
- 山田礼子 (2012). 『学士課程教育の質保証へむけて—学生調査と初年次教育からみえてきたもの—』東信堂.
- 山田礼子 (2016). 「共通教育における直接評価と間接評価における相関関係—成果と課題—」『大学教育学会誌』38(1), 42–48.

Articles by the Center Staff and Research Fellows

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

Kayo Matsushita¹, Kazuhiro Ono², and Yugo Saito³

(¹Center for the Promotion of Excellence in Higher Education, Kyoto University,

²Graduate School of Medical and Dental Sciences, Niigata University, ³Office for Educational Planning, Niigata University)

This paper addresses how to combine course- and program-level assessments and presents a new assessment method illustrated by a case of dental education program in Japan. Performance assessments are considered effective for evaluating knowledge integration and higher-order skills, while placing a burden on faculty, hence their feasibility as the program-level assessment is regarded lower than standardized tests or questionnaire surveys. We have developed several performance assessments at the course level, such as Modified Triple Jump for the PBL course. Based on this experience, we propose Pivotal Embedded Performance Assessment (PEPA) as a method for combining assessment at the course and program levels. The method limits the range of performance assessment to key courses directly linked to program goals and placed at the critical juncture points of curriculum, while entrusting the assessment of other courses to expert judgment of individual teachers. PEPA consists of the following procedures: systematization of curriculum and selection of key courses; design and implementation of performance assessments by a faculty team; setting passing criteria with incorporating the function of formative assessment; certifying the completion of the degree program. PEPA thus enables maintaining assessment feasibility and compatibility with a credit system, while ensuring assessment validity and reliability.

Keywords: Pivotal Embedded Performance Assessment (PEPA), Program-level assessment, Curriculum, PBL (Problem-Based Learning), Dental education