実践報告

PBL カリキュラムにおけ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―

小野　和宏[1]・松下　佳代[2]・斎藤　有吾[3]

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

本研究の目的は、高次の統合的能力である問題解決能力の変容を、パフォーマンス評価を用いてプログラムレベルで把握することである。対象は新潟大学歯学部の PBL カリキュラムで学んだ学生 82 名である。独自に開発した評価ツールである改良版トリプルジャンプ（MTJ）を、第 2 学年から第 3 学年におけるカリキュラムの異なる 3 時点で実施し、縦断的に学生の問題解決能力を評価した。一方、MTJ に用いた 3 つのパフォーマンス課題の難易度を、課題自体の複雑さと難易度に関する学生の認識という 2 つの側面から検討した。その結果、学年が進み PBL の学習経験が増えるにつれて、MTJ の評価得点は中程度～大の効果量をもって有意に上昇していることが明らかになった。また、パフォーマンス課題はカリキュラムの学習内容を反映して、学年が進むにつれて複雑になっており、課題の難易度に関する学生の認識でも同様な結果が得られた。以上のことから、MTJ の評価得点の上昇は、課題の難易度が下がったことによるものではなく、学生の問題解決能力が向上したことによると解釈された。このように、課題の難易度という観点を加えることによって、複数科目でのパフォーマンス評価の結果から、プログラムレベルでの長期的な問題解決能力の変容を評価できることが明らかになった。

キーワード：PBL、カリキュラム、学習成果、パフォーマンス評価、保健医療福祉分野

1. 問題と目的

1.1. 直接評価による学習成果の評価

中央教育審議会答申「学士課程教育の構築に向けて」（学士課程答申）が 2008 年 12 月に出されて以降、教員が学生に何を教えたかではなく、その結果、学生は何を学んだか、すなわち学習成果が大学教育で重視されるようになった。この学習成果には、「学士力」やその元となった AAC&U（全米大学・カレッジ協会）の「本質的学習成果」（AAC&U, 2007）にみられるように、知識だけでなく、論理的思考や問題解決、コミュニケーションなどの能力が含まれている。

現在、各大学には 3 つのポリシー（ディプロマ・ポリシー、カリキュラム・ポリシー、アドミッション・ポリシー）の策定・公表が義務づけられている。杉谷（2019）によると、480 学部から回答を得た 2018 年の調査では、7 割以上の学部がディプロマ・ポリシー策定に際し、「能力（コンピテンシー）の涵養」を「とても重視した」と回答している。また、吉田（2008）によると、1000 学部から回答を得た 2003 年の同様の調査では、そのような能力として、課題探求能力、批判的精神といった学士課程教育全体で、すなわち複数科目にまたがってプログラムレベルで育成される能力が重視されていることが示されている。だが、そうした目標を達成できるようなカリキュラムを組むこと、また、それらが獲得されているかを評価することには大きな困難が伴う。

学習成果を評価する方法は、大きく直接評価と間接評価の 2 つに分類することができる（山田, 2012）。直接評価とは、学生の知識や行為を通じて、すなわち「何ができるか」を学生自身に提示させることで、学習成果を直接的に評価することである。一方、間接評価とは、学生の学習行動や自己認識を通じて、すなわち「どのように学習したか」や「何ができると思っているか」を学生自身に答えさせることによって、学生の学習成果を間接的に評価することである（松下, 2012）。先述した学士課程答申では、学習成果を「プログラムやコースなど、一定の学習期間終了時に、学習者が知り、理解し、行い、実演できることを期待される内容を言明したもの」と定義しており（中央教育審議会, 2008）、この定義にしたがえば、学習成果の評価では直接評価を中心とすることが妥当であろう。

これまで、プログラムレベルの学習成果を把握するのに多用されてきたのは、間接評価の一手法である質問紙

調査であった（例えば、畑野・上垣・高橋, 2015；山田, 2012）。それに対し、直接評価によってプログラムレベルの学習成果を把握することは可能なのだろうか。

1.2. パフォーマンス評価を用いた縦断研究の必要性

直接評価の中でも、本研究ではパフォーマンス評価を取り上げる。パフォーマンス評価とは、ある特定の文脈のもとで、さまざまな知識や技能などを用いながら行われる、学習者自身の作品や実演（パフォーマンス）を直接的に評価する方法のことである。パフォーマンス評価には、①評価の直接性（パフォーマンスを実際に行わせ、それを直接、評価する）、②パフォーマンスの文脈性（パフォーマンスは具体的な状況の中で可視化され、解釈される）、③パフォーマンスの複合性（それ以上分割すると本来の質を失うという、一まとまりのパフォーマンスを行わせる）、④評価の分析性と間主観性（そうした質の評価のために評価基準と複数の専門家の鑑識眼を必要とする）といった特徴がある（松下, 2010）。「文章コミュニケーション、問題解決、分析的推論などのスキルは、学問分野の中で適用されるときにこそ、もっともよく学ばれ、評価される」（Banta & Pike, 2012, p. 25）とすれば、問題解決のような高次の統合的能力を直接評価するうえで、パフォーマンス評価はもっとも適した評価方法といえるであろう。

ただし、パフォーマンス評価によって、複数科目にまたがるプログラムレベルの学習成果を把握する際、乗り越えなければならない課題がある。仮に、問題解決能力を評価するために、学年をまたいで２つの時点でパフォーマンス評価を行ったとする。また、評価基準に共通のルーブリックを用いて、それを得点化したとしよう。ここで得られた２つの得点の差異は、そのまま問題解決能力の変容の度合いと解釈することが可能なのだろうか。

より具体的に述べよう。上の例において、共通のルーブリックを用いて、ある学生が２回とも「レベル２」と評価され、それを２点と換算したとする。この場合、得点上は変化が０となる。これを、単純に、問題解決能力の変容はなかったと解釈してよいのだろうか。もし、２回目のパフォーマンス課題が、より専門性が高く複雑になるなど、難易度が高くなっていた場合、得点の変化はみられずとも、問題解決能力は高まっていると解釈するほうが妥当ではなかろうか。つまり、共通のルーブリックを用いたとしても、パフォーマンス課題の難易度が異なるのであれば、そこから得られた量的な指標の解釈には慎重になる必要がある。このように、パフォーマンス評価は、「パフォーマンス課題」とルーブリックをはじめとした「評価基準」によって構成されているため、縦断的な検討を行う場合、評価基準をそろえればそれでよいというわけではなく、課題の難易度も考慮する必要があるのである。

これに対して、課題の難易度に応じて、評価結果を調整すればよいという意見もあるだろう。たしかに、テスト理論においては、異なるテストの得点を比較可能にするために、等化という手続きがとられる。等化とは、２つの尺度について、どちらか一方の尺度に０点の位置や目盛の幅をそろえることを指す（加藤・山田・川端, 2014）。これにより、難易度や識別力といった項目の特性が異なるものによって構成されたテスト間の得点を調整することが可能になる。しかしながら、等化の手法を実施するには、一般に大きなサンプルサイズや異なるテスト間で共通受験者あるいは共通項目が必要であり、多くの時間や人的リソースを要するパフォーマンス評価では困難である。なお、２回とも同じパフォーマンス課題にするというやり方もあるだろうが、１回目の記憶がそのまま２回目に影響を与えるであろうことは容易に想像される。また、カリキュラムを考慮すれば、低学年から高学年にかけて、パフォーマンス課題の難易度が高くなるほうが自然であり、実際の教育現場で用いるには無理がある。

そこで、本研究では、パフォーマンス課題の難易度に関して、課題自体の複雑さと難易度に関する学生の認識という側面から検討し、それらも考慮したうえで、プログラムレベルでの学生の変容の把握（複数科目をまたいだ学習成果の評価）を試みる。

1.3. 本研究の目的と構成

以上から、本研究では現在の大学教育で要請されている高次の統合的能力について、学生の変容をプログラムレベルで把握することに取り組む。とりわけ、高次の統合的能力のうち、問題解決能力を取り上げ、問題基盤型学習（Problem-Based Learning、以下、PBL）を軸としたカリキュラム（PBL カリキュラム）を通じて育成される問題解決能力の変容について明らかにする。

具体的には、PBL を中心に構築した新潟大学歯学部口腔生命福祉学科のカリキュラムで学んだ学生を対象として、問題解決能力をカリキュラムの３時点で縦断的に捉え、その変容を検討する。口腔生命福祉学科のカリキュラムは、第２学年から学士課程を通じて、複数の PBL 科目を継続的に実施しており、複数科目をまたいだ学習成果を評価するうえで適した対象フィールドといえる。PBL で育成された問題解決能力を直接評価するためには、独自に開発したパフォーマンス評価である「改良版トリプルジャンプ」（Modified Triple Jump、以下、MTJ）（小野・松下・斎藤, 2014）を用いる。３つのパフォーマンス課題それぞれの難易度を、各シナリオの複雑さと学生の難易度の認識に

関する質問紙調査によって検討し、それらを考慮したうえで、PBL カリキュラムを通して獲得された学習成果について考察を加える。

2. 方法

2.1. PBL カリキュラム

（1）教育目標

新潟大学歯学部は、歯学科（6 年制、定員 40 名）、口腔生命福祉学科（4 年制、定員 20 名）の 2 学科からなり、口腔生命福祉学科は「変化の激しい現代社会のなかで、国民の多様な価値観を受け入れ、質の高い保健医療福祉サービスを提供するために、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、口腔保健、歯科医療、福祉を総合的に思考・展開できる能力を有する人材を育成する」ことを教育目標に掲げている（新潟大学歯学部口腔生命福祉学科, 2020）。「口腔保健、歯科医療、福祉を総合的に思考・展開できる能力」とは、歯科と福祉それぞれの領域における知識・スキルを有し、それらを統合して「食べること」に関する問題を解決に導く力と定義しており、教育目標は、低学年から高学年に向けて、専門性を高めながら問題解決能力を育成することを意味している。

（2）PBL カリキュラムのアウトライン

一般に、PBL の効果として、統合された深い知識・理解の習得、問題分析・問題解決能力の育成、対人関係能力の育成、継続的な学習意欲の涵養があげられている（Barrows, 1998）。そこで、新潟大学歯学部口腔生命福祉学科では、教育目標としている問題解決能力を育成するために、カリキュラム全体を通じて PBL を導入している。口腔生命福祉学科のカリキュラム（図 1）は 4 年制で、学期を学習の大きな基本単位として捉え、それぞれの学期で中心となる学習内容を設定している。第 1 学年は「大学学習法と人としての成長」、第 2 学年前期は「口腔の重要性理解と歯科医療従事者としての自覚」、同後期は「軽度な歯科疾患の診査・処置・予防」、第 3 学年前期は「進行した歯科疾患の診査・処置・予防ならびに社会福祉と社会保障の理解」、同後期は「高齢者・障害者の理解と対応」、第 4 学年は「個人および社会の視点からみたオーラルヘルスプロモーションの実践」となっている（新潟大学歯学部, 2020）。

初年次には、大学学習法により学習態度の転換を図り、本カリキュラムを履修していくうえで必須な問題解決、論理的思考、コミュニケーションの基礎を育成する。また、一般教育科目を通じて、多様なものの見方にふれさせ、さまざまな文化や価値を受容し、個性を尊重する態度を涵養する。第 2 学年からは PBL を全面的に導入し、統合された知識を習得させるとともに、問題解決能力、対人関係能力を向上させる。また、早期から継続して実際の患者に接する機会を設け、歯科医療従事者としての自覚と態度を涵養する。第 3 学年から福祉に関する学習が本格的に開始され、第 2 学年に引き続き PBL により、口腔保健学にとどまらず、口腔保健学と福祉学の知識を統合させる。また、歯科医療に関する実習を通して、歯科医療従事者としての基本的な技能を身につけさせる。第 4 学年には、臨床実習ならびに社会福祉現場実習を実施するとともに、両実習を通じて自ら見いだした課題に対する研究レポートを作成させる（新潟大学歯学部口腔生命福祉学科, 2020）。

なお、PBL 科目として、第 2 学年前期には「PBL 入

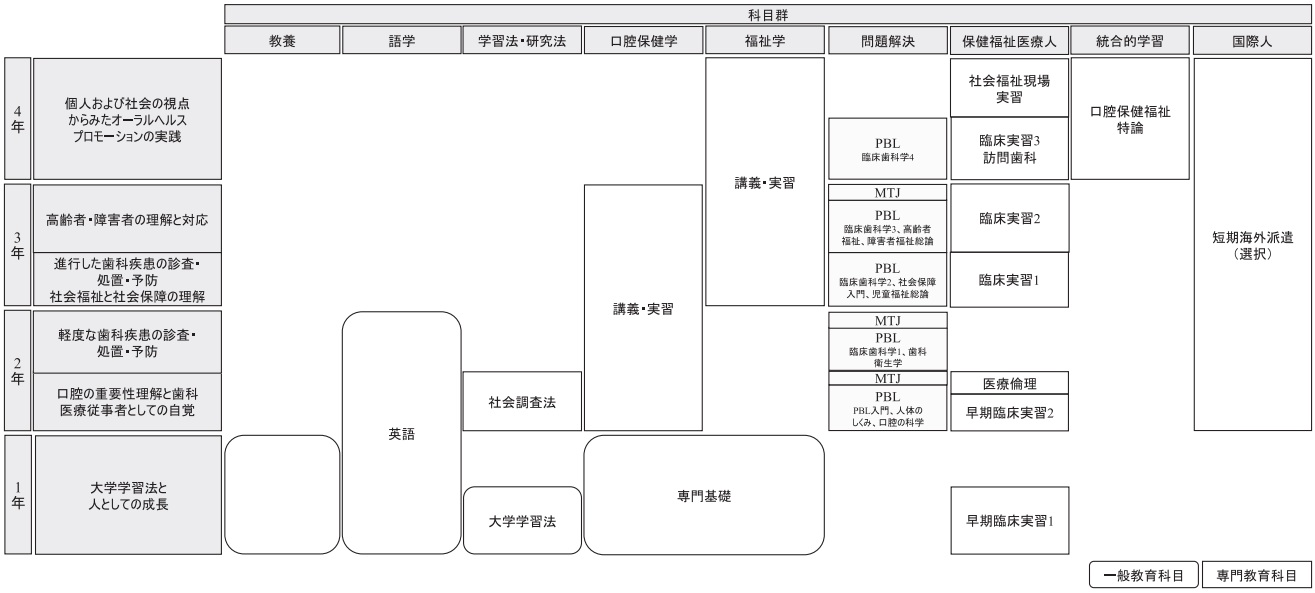


図 1　新潟大学歯学部口腔生命福祉学科のカリキュラムと PBL ならびに MTJ の位置づけ

門」「人体のしくみ」「口腔の科学」の３つの授業科目が、同後期には「臨床歯科学１」「歯科衛生学」の２つの授業科目がある。また、第３学年前期には「臨床歯科学２」「社会保障入門」「児童福祉総論」の３つの授業科目が、同後期には「臨床歯科学３」「高齢者福祉総論」「障害者福祉総論」の３つの授業科目が配置されている。さらに、第４学年前期には「臨床歯科学４」の授業科目がある（図１）。

2.2. 改良版トリプルジャンプ

（1）評価方法

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いだし、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。

一方、教員は、学生が提出したワークシートを「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観点、「レベル３」から「レベル０」の４段階からなるステップ１・２のルーブリック（表 1a）により評価する。また、ロールプレイを「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」の４つの観点で、ステップ１・２と同様に４段階からなるステップ３のルーブリック（表 1b）により評価する。

ステップ１・２ならびにステップ３のルーブリックの観点は、AAC&U の VALUE ルーブリックの「問題解決」（松

表１　MTJ のルーブリック枠組み

a：ステップ１・２

	ステップ１			ステップ２		
観点	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
レベル３	問題を見いだし、シナリオの事実から、推察しうる原因も含め、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	問題を見いだし、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	問題を見いだしているが、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。					

b：ステップ３

	ステップ３			
観点	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
レベル３	歯科医師が歯科治療恐怖症と考えた理由、虐待と関連した兆候、カンファレンスでの養護教諭との話の内容など、児童虐待の可能性を吟味する上で必要な追加情報を、すべて、的確に収集している。	児童虐待の可能性と対応について、全身的ならびに歯科的な兆候、医療従事者の役割、法律の点から深く柔軟に理解しており、歯科医師からの追加情報も十分統合できている。	自分とは異なる相手の考えを受け入れ、歯科治療恐怖症の可能性も十分考えられること、自分の考えが間違っているかもしれないことなど述べながら、自分の考えを相手に聴いてもらえるよう努めている。	話の順序や組み立てがよく整えられており、専門の違いに配慮して、相手の理解を得られるよう専門用語をわかりやすく言いかえたりしながら、説明・提案している。
レベル２	歯科医師が歯科治療恐怖症と考えた理由、虐待と関連した兆候、カンファレンスでの養護教諭との話の内容など、児童虐待の可能性を吟味する上で必要な追加情報を、ある程度収集している。	児童虐待の可能性と対応について、全身的ならびに歯科的な兆候、医療従事者の役割、法律の点から適切に理解しており、歯科医師からの追加情報も一部統合できている。	自分とは異なる相手の考えを受け入れ、歯科治療恐怖症の可能性を考慮に入れているが、自分の考えとの関連づけが不十分である。	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専門用語の使い方や表現にやや問題がみられる。
レベル１	歯科医師が歯科治療恐怖症と考えた理由、虐待と関連した兆候、カンファレンスでの養護教諭との話の内容など、児童虐待の可能性を吟味する上で必要な追加情報のごく一部を収集している。	児童虐待の可能性と対応について、全身的ならびに歯科的な兆候、医療従事者の役割、法律の点から理解しているが、文字情報としての理解にとどまっており、歯科医師からの追加情報の統合は行っていない。	おもに自分の考えを相手に伝えることに終始している。	話の順序や組み立てが混乱しており、説明・提案が理解しにくい、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を満たさない場合はゼロを割り当てること。			

（注）ステップ３のルーブリックは課題特殊的なため、記述語は３回目の MTJ（シナリオ３）のものを例として示す

シナリオ 1

あなたは新潟大学歯学部の2年生です。4月から専門科目の授業が始まり、解剖学や生理学を学んでいますが、「PBLテュートリアル」という新しい学習方法にまだ慣れず、また学習内容も急に難しくなったように感じられ、不安を抱えながら悪戦苦闘の毎日です。

そんなある日、全学のサークルで知り合った友人の佐藤彰くん（工学部3年生）が左の頬を腫らして近づいてきました。彼は3日前に近くの歯科医院で下顎の左側智歯を抜去したとのことですが、いまだに左側の下唇に麻酔がかかっているような感じが残っており、ご飯粒がついていても、わからないといいます。また、抜歯してから口を大きく開けられなくなり、飲み込むときにノドの左側に痛みもあるとのことで、食事がとりにくいと訴えています。

あなたが歯学部の学生であることから、なんでこんなことが起きているのか説明してほしいと頼まれましたが、あなたはどう答えてよいかわからず黙っていました。彼は不安そうな顔をして、「次のサークルのときでいいから」と言って別の友人のところに去っていきました。彼の後ろ姿を見ながら、将来、歯科医療従事者になるものとして、せめて「大変だね」のひと言でもかけてあげればよかったと、あなたはちょっと後悔しました。

シナリオ 2

あなたは新潟大学医歯学総合病院の歯科衛生士です。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。私はね、タバコをやめるくらいなら死んだ方がましだと思っているよ。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」と書いたくらいだ。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、やめた方がいいと思いますが…。
高橋：糖尿病は関係なかろう。ここは歯科だろう。おやおや、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。

シナリオ 3

あなたは新潟大学歯学部口腔生命福祉学科をこの春卒業し、4月から小林歯科医院で歯科衛生士として働き始めました。院長の小林信一先生（32歳・男性）はスタッフそれぞれの専門性を重んじ、またスタッフからの意見や提案を積極的に取り入れてくれるので、あなたはやりがいのある充実した毎日を過ごしています。後で聞いた話では、高齢化が進んだ現在、歯科医院にも福祉に強いスタッフが必要との判断から、社会福祉士の資格を持つあなたが採用されたといいます。

そんなある日、今年から笹山小学校の学校歯科医を務めることになった小林先生と歯科検診に行くことになりました。先生が検診している後ろのテーブルで歯科疾患の罹患状況を記録用紙に記入していたあなたですが、2年生の川崎太郎くん（7歳・男児）だけう蝕が著しく多く、口の中いたるところに歯肉膿瘍があることに違和感を覚えました。「最近の子どもにしては珍しいな」と思い、検診を受けている川崎くんを見たところ、他の生徒に比べて貧弱な体格で、不安で落ち着きのない様子が感じとれました。

あなたが検診の後始末をしている間に、小林先生は養護教諭の安藤直子先生（43歳・女性）とのカンファレンスを済ませて戻ってきました。小林先生も川崎くんを覚えており、いわゆる「歯科治療恐怖症」であろうと言いました。あなたは大学で学んだ「児童虐待」という言葉が一瞬頭に浮かびましたが、何も言わず黙っていました。

図 2　評価課題（シナリオ）

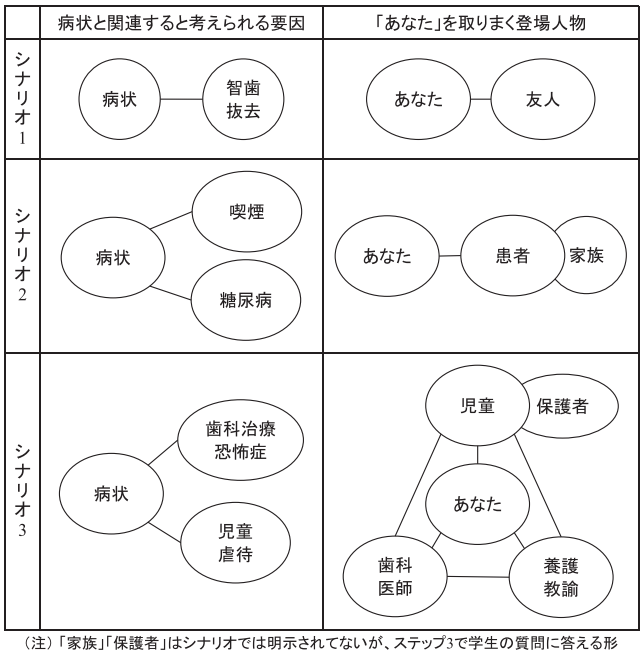
下，2012；Rhodes, 2010）をふまえて構成されている。「共感的・受容的態度」や「コミュニケーション」は、「問題解決」とは別のカテゴリーであるが、医療職の場合、相手に働きかけることで問題解決を実行するため、ステップ 3 のルーブリックにこの 2 つの観点を追加した。

なお、ステップ 1・2 のルーブリックは、さまざまな評価課題に対応できる一般的ルーブリックである。それに対して、ステップ 3 のルーブリックは、シナリオの内容に依存する課題特殊的なルーブリックで、シナリオごとに別個のルーブリックを使用している。しかし、観点は同一であり、記述語だけがシナリオに合わせて部分的に異なる。また、ステップ 1・2 ならびにステップ 3 のルーブリックのレベルを評語で表せば、「レベル 3」は「大変よい / 模範的である」、「レベル 2」は「十分である / できた」、「レベル 1」は「やや不十分である / 許容範囲内である」、「レベル 0」は「不十分である / 許容範囲外である」となり、各レベルにおける到達度は標準化されている。

新潟大学歯学部口腔生命福祉学科は、先述したように、PBL を中心にカリキュラムを構築しており、その学習成果を評価するために、第 2 学年前期（1 回目）と後期（2 回目）、第 3 学年後期（3 回目）の計 3 回にわたって（図 1）、MTJ を 3 名の教員が評価者となり実施している。なお、現在のところ、MTJ は主に知識を問う学期末試験を受けるための受験要件として位置づけられており（新潟大学歯学部, 2019）、MTJ の評価結果自体は単位修得や進級に関係しない。ただし、学生には、どの回においても、到達目標として最低レベル 1 以上、なるべくレベル 2 以上を目指して取り組むように指導している。

（2）評価課題（シナリオ）

MTJ のシナリオは、学期や学年での学習内容を反映させ、第 2 学年前期（シナリオ 1）は「智歯を抜去した友人への病状説明」、後期（シナリオ 2）は「糖尿病をもつ歯周病患者への病状説明と禁煙指導」、第 3 学年後期（シナリオ 3）は「歯科検診での児童虐待の発見と対応」としている（図 2）。



（注）「家族」「保護者」はシナリオでは明示されてないが、ステップ3で学生の質問に答える形で追加情報を提示する

図 3　評価課題（シナリオ）の複雑さ

それぞれのシナリオにおいて、問題となる病状と関連すると考えられる要因ならびにシナリオの主人公「あなた」（評価を受ける学生自身がこの役割を演じる）を取りまく登場人物の関係を概念的に図示した（図 3）。このように、学年が進行するにしたがって複雑に、また歯科と福祉の領域複合的な課題となるように作成している。

2.3. 調査と分析

本研究は目的に対応するため、2 つの調査から構成され

ている。一つは、PBL カリキュラムで学んだ学生を対象として、パフォーマンス評価の結果を 3 時点で追ったものである。もう一つは、その 3 つのパフォーマンス評価の結果から PBL カリキュラムにおける学習成果を解釈するために、それぞれのパフォーマンス課題の難易度に関して、3 つの課題をすべて経験ずみの学生を対象にして行った質問紙調査である。

（1）パフォーマンス評価の縦断的調査

対象は、MTJ をカリキュラムに導入した 2013 年度から 2017 年度の 5 年間に、前述のカリキュラムを履修し、MTJ で計 3 回の評価を受けた 2012〜2015 年度入学の学生 82 名である。1 学年の学生定員が 20 名と少ないことから、学生が共通して体験したカリキュラムの学習成果を取り出しやすくするために、4 学年分の学生を調査対象とした。評価結果の使用にあたって、調査の目的、個人情報の保護等について、学生に口頭ならびに文書で説明し、文書で同意を得た。

第 2 学年から第 3 学年における 3 回の評価結果を用いて、ルーブリックのレベルを得点に置き換えて分析する。MTJ は 3 名の教員で評価を行っているが、5 年間 4 学年分の学生を調査対象としており、その間に教員の異動等もあったことから、本研究では、すべての対象学生を 3 回にわたり評価した筆頭著者 1 名の評価結果をデータとして使用した。なお、教員 3 名による評価得点の信頼性係数は 0.70 を超えており（小野他, 2014；斎藤・小野・松下, 2015）、評価者間信頼性が十分高いことは確認ずみである。

各回におけるステップ 1・2 の 6 観点ならびにステップ 3 の 4 観点の計 10 観点、すなわち MTJ ステップ 1〜3 の合計得点（30 点満点）を「合計得点」、観点別の得点（各観点 3 点満点）を「観点別得点」として求めるとともに、3 回の得点の変化を多重比較（Holm 法による調整）により分析する。また、効果量（*d* 値）を算出し、*d* = 0.20 を効果量小、0.50 を効果量中、0.80 を効果量大とする（Cohen, 1988; Kirk, 1996）。

（2）評価課題の難易度に関する質問紙調査

図 3 で示したような概念的なつながりの複雑さは、学生の難易度の認識にも現れるのか。それを確認することで、難易度も考慮して評価得点の変化を解釈する手がかりが得られる。そこで、評価課題の難易度に関する学生の認識を検討するために、2016 年度入学の学生 21 名に対して、カリキュラムを修了する直前の 4 年次に質問紙調査を行った。これまで受験した 3 つの評価課題について、「いま現在、あらためて、それぞれのシナリオをみたときに、難易度がどうであると思うか」を、「とても難しい」「難しい」「やや難しい」「適当」「やや簡単」「簡単」「とても簡単」の 7〜1 の 7 段階で回答してもらい、得点化した。また、そのように判断した理由もあわせて記述してもらった。記述に関しては、文意が読み取れない記述を除外し、シナリオ 1 は 12 記述、シナリオ 2 は 14 記述、シナリオ 3 は 19 記述について、意味内容の類似性にしたがい分類し、〈カテゴリーネーム〉を付けて内容分析を行った。

3. 結果

3.1. 改良版トリプルジャンプの評価得点の変化

第 2 学年前期（1 回目）、同後期（2 回目）、第 3 学年後期（3 回目）の MTJ ステップ 1〜3「合計得点」の平均値±標準偏差は、それぞれ 13.4 ± 4.3、16.8 ± 4.4、18.6 ± 4.2 で、第 2 学年から第 3 学年へと学年が進むにしたがって上昇していた（図 4）。各回の間で統計学的有意差が認められ（図 4）、また 1 回目と 2 回目、2 回目と 3 回目、1 回目と 3 回目の変化の効果量（*d* 値）は、それぞれ 0.77、0.42、1.22 であり（表 2）、中程度〜大の効果量が得られた。

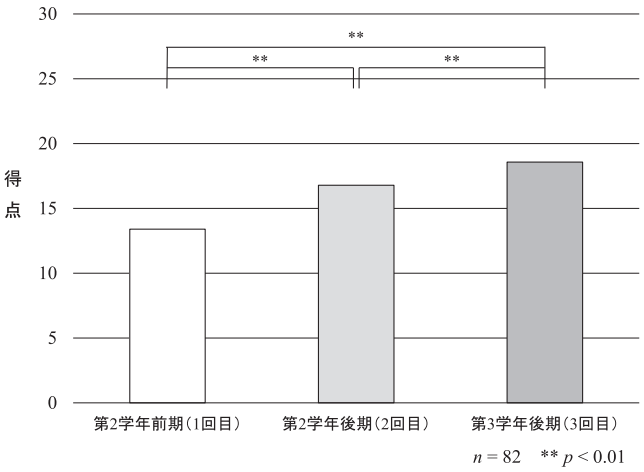


図 4　各回の合計得点

表 2　評価得点の変化の効果量（*d* 値）

	合計得点	観点別得点									
		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	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
2 回目–1 回目	0.77	0.23	0.73	1.15	0.74	0.13	0.47	−0.43	0.71	0.84	0.32
3 回目–2 回目	0.42	−0.02	−0.06	−0.22	0.06	0.58	0.21	1.25	0.64	0.12	0.54
3 回目–1 回目	1.22	0.23	0.75	1.04	0.86	0.74	0.61	0.77	1.31	1.02	0.96

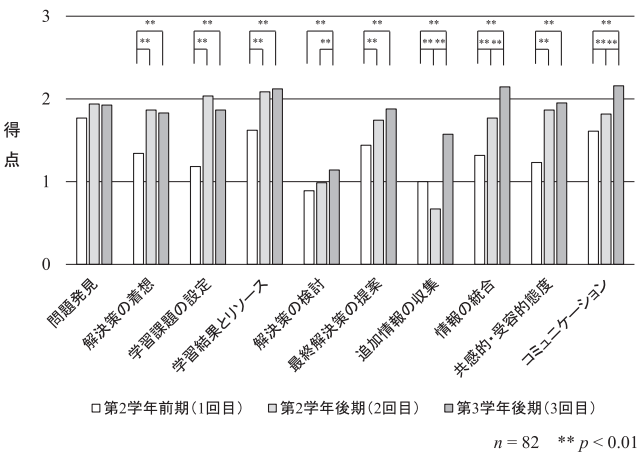


図 5　各回の観点別得点

「観点別得点」については、観点間でばらつきがあり、全般に「解決策の検討」や「追加情報の収集」では低かったものの、やはり第 2 学年から第 3 学年へと学年が進むにしたがって、おおむね上昇していた（図 5）。特に、1 回目と 3 回目では、「問題発見」以外のすべての観点において統計学的有意差（図 5）ならびに中～大の効果量がみられた（表 2）。

3.2. 評価課題の難易度に関する学生の認識

課題の難易度に関する学生 21 名の認識は、シナリオ 1、2、3 の順に、中央値、最頻値ともに 4、4、5 で、多くの学生は、シナリオ 1 およびシナリオ 2 の難易度を「適当」、シナリオ 3 を「やや難しい」と判断していた。難易度の平均値±標準偏差は、それぞれ 4.6 ± 0.9、4.3 ± 0.5、5.0 ± 0.7 で、シナリオ 2 とシナリオ 3 の間に有意差（$p < 0.01$）が認められた。

難易度の判断の理由としては、シナリオ 1 では、「臨床経験がまだ浅く、教科書の内容理解が追いつかないところがあった」とする〈学習内容の理解〉が 6 記述（50.0%）、「年齢的にも経験しそうな身近な話であり、教科書等で調べるのにちょうどよい難易度だった」との〈身近な事例〉〈学習課題の調査〉が各 3 記述（25.0%）であった。シナリオ 2 では、「歯周病、糖尿病、タバコをうまく関連づけて解決するのが難しかった」との〈歯周病・糖尿病・喫煙の関連づけ〉、「やる気がない人に対して、どうしたら動機づけができるか考え実行するのが大変だった」とする〈患者の動機づけ〉の難しさをあげる意見が各 7 記述（50.0%）みられた。シナリオ 3 では、「歯科と福祉の 2 つの領域を関連づけながら解決するのは難しかった」とする〈歯科と福祉の領域複合課題〉が 13 記述（68.4%）とほとんどを占め、他に、「解決策が何通りも考えられ判断が悩ましい」との〈さまざまな解決策の可能性〉、「誰にどこまで介入すべきか判断が難しい」との〈解決の目標設定〉が各 3 記述（15.8%）であった。

4. 考察

4.1. 学生の問題解決能力の変容の把握

PBL カリキュラムで学習した学生を対象として、第 2 学年前期と後期ならびに第 3 学年後期の 3 回、MTJ を用いて同一の評価者が問題解決能力を直接評価した。その結果、「合計得点」は第 2 学年から第 3 学年へと学年が進むにつれて、言い換えれば、PBL の学習経験が増えるにつれて有意に上昇しており（図 4）、中程度～大の効果量が得られた（表 2）。教育の分野では、実際に意味のある効果量として、最低でも $d = 0.33$ とする研究者が多いことも考慮すれば（Gall, Borg, & Gall, 1996）、1 回目と 2 回目、2 回目と 3 回目、1 回目と 3 回目のどの比較においても、少なくとも得点上は意味のある変容があったと解釈できる（表 2）。

学生が受験した 3 つの評価課題（シナリオ）は、先述したように、学年が進行するにしたがって複雑に、また歯科と福祉の領域複合的な課題となるよう作成しており、さらに評価課題の難易度に関する学生の認識でも同様な結果がみられることから、この得点の上昇は、評価課題が易しくなったことによると考えるより、学生の問題解決能力が向上したことによると考えるのが妥当であろう。

なお、問題解決能力を「観点別得点」でみると、「問題発見」の観点は、得点が 1 回目から比較的高く、3 回目にいたっても有意な上昇はみられない（図 5）。なぜ「問題発見」だけが特異な傾向を示したのか。MTJ では、出発点である問題の認識で学生間に大きな違いが生じると、解決の方向が拡散し、評価が難しくなると危惧されることから、問題を発見しやすいシナリオを評価者側が意識して作成している。このような評価方法のデザイン上の工夫が、「問題発見」の変化の小ささをもたらしたと考えられる。

本研究に関連する先行研究である小野・斎藤・松下（2018）では、計 3 回の MTJ を経験した 2012～2014 年度入学の口腔生命福祉学科の学生 64 名を対象として、MTJ 経験の意味について質問紙調査（自由記述）を行い、3 回目は「成長」といった自分自身の成長に関する語が記述を特徴づけていることを見いだしているが、これは学生の認識にもとづいた質的な間接評価から得られた結果である。それに対して、本研究ではパフォーマンス評価による直接評価の結果から学生の成長を確認できた。この 2 つの結果の整合性は、お互いにそれぞれの評価の妥当性を補強し合うものといえる。学生が MTJ の回を重ねて成長を認識していたことに関しては、本研究によって教員による直接評価の裏付けがなされたことになる。また、後述するよ

うに、MTJ は学生に深い学びや省察を促すことも狙いとしており、単なる問題解決能力の測定ツールではない。複数回の MTJ という形でパフォーマンス評価を縦断的に実施した結果、当該能力が向上しており、かつ学生自身もこの評価を通じて成長したという認識が得られたのであれば、目的に合致した影響を学生に与えていたといえる。すなわち、結果的な妥当性（平井, 2006；村山, 2012）を示す証左になりえるだろう。

4.2. PBL カリキュラムの有効性

PBL カリキュラムの有効性については、これまで、医学教育を中心に、さまざまな研究がなされてきた。効果量の点から明らかな効果はないとする研究（Colliver, 2000）もあるが、過去のレビュー等（Albanese & Mitchell, 1993; Clark, 2006; Dochy, Segers, Van den Bossche, & Gijbels, 2003; Hung, Jonassen, & Liu, 2008; Neville, 2009）をまとめると、講義中心のカリキュラムと比較して、①知識の獲得に差はみられないが、記憶定着率は優れている、②スキル（知識の活用、問題解決、臨床能力、生涯学習能力）の習得は優れている、とする報告が多い。なお、医学生のように高度に選抜された学生では、国家試験のような資格取得に関わる試験で比較しても、教育方法の違いによる影響は現れにくい、とする意見（Norman & Schmidt, 2000）がある。上記のように、知識の獲得については従来型カリキュラムと差がみられないという知見が多いが、例えば、Albanese & Mitchell（1993）は米国の資格試験である National Board of Medical Examiners（NBME） 試験のスコアを中心に検討した結果から結論を導いているにすぎず、Norman & Schmidt（2000）の意見にしたがうならば、PBL カリキュラムの優位性が覆い隠されているという可能性もある。

今回、PBL カリキュラムにより問題解決能力の向上がみられたのは、PBL の学習プロセスが問題解決プロセスと同じであることから（Barrows, 1998）、PBL 自体の効果が大きいと推察される。しかし、学生は第 2 学年から第 3 学年のカリキュラムを通じて PBL 以外の教育も受けており、他の要因の関与も否定できない。例えば、MTJ で問題解決能力に含めた「共感的・受容的態度」や「コミュニケーション」は、PBL の授業でも育成されるが、PBL と同時期に行なっている「早期臨床実習（アーリーエクスポージャー）」や「臨床実習（クリニカルクラークシップ）」（図 1）での学習の効果もあるであろう。これらの実習には共感的・受容的態度が必要な実際の患者とのコミュニケーションが含まれているからである（新潟大学歯学部, 2019）。一般に、PBL カリキュラムでは、早期から学生に現場を

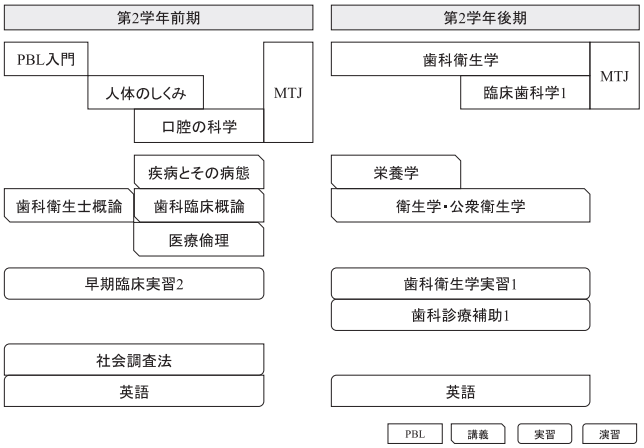


図 6　年単位の PBL カリキュラム（第 2 学年）

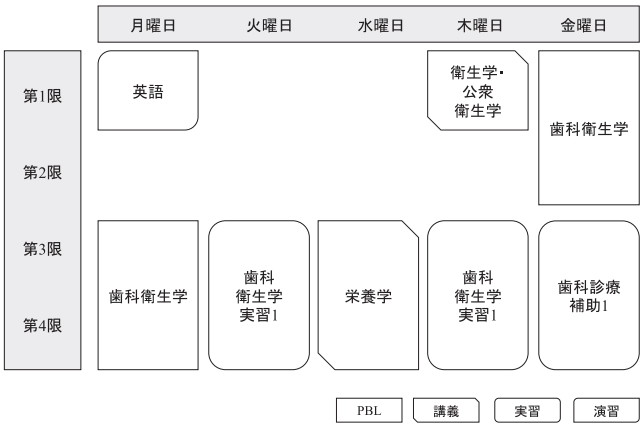


図 7　週単位の PBL カリキュラム（第 2 学年後期前半）

体験させ、その経験と教室での学びを統合させるようにカリキュラムを編成するが、このことが学生の社会的・認知的スキルを向上させる要因とも言われている（Koh, Khoo, Wong, & Koh, 2008）。

このように、PBL 科目と他の授業科目との関連から新潟大学歯学部口腔生命福祉学科の PBL カリキュラム（図 1）を年単位（図 6）あるいは週単位（図 7）でみると、内化と外化の学習サイクル（松下, 2015）が、PBL を中核として、どのスパンでも回るように組み込まれていることがわかる。PBL、講義、実習を有機的に配置し、それぞれの学習内容に関連性をもたせることにより（小野・松下, 2015）、PBL 科目において出会う問題と既有知識や経験との間で生じるコンフリクトにより動機づけられた学生は、その解決を目指して、関連する知識を講義科目で習得、すなわち内化を行い、また得られた知識を実習科目で活用し、外化を行う。この学習サイクルを繰り返すことにより、理解が深まるように編成されているカリキュラムといえる。MTJ により評価した問題解決能力が、よりポジティブに変容する傾向にあった学生は、幅広い知識を記憶し、理解しようと学習に取り組んでいた者であること（斎藤他, 2015）を考えると、カリキュ

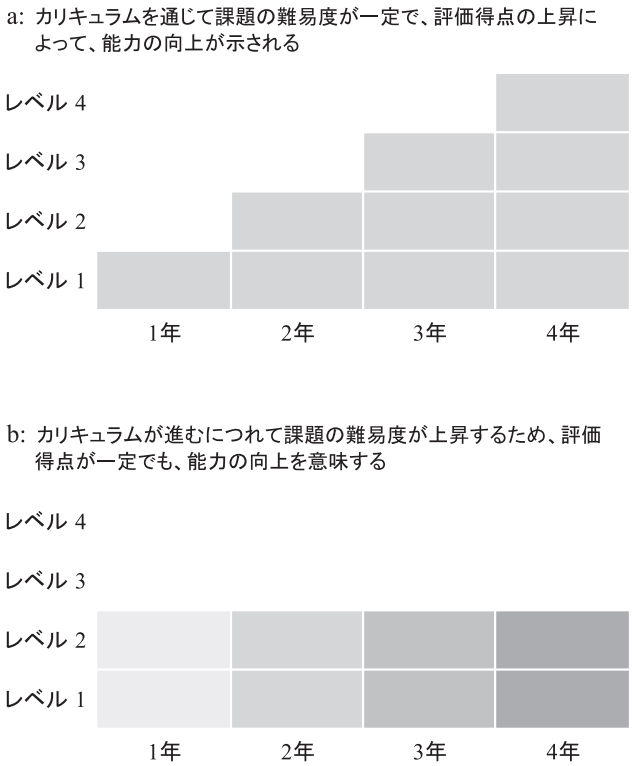
ラムの構造自体も問題解決能力を向上させた要因の一つと考えられる。

また、今回、PBL の学習成果を評価した MTJ は、学習の評価としての役割だけでなく、学生が遭遇しうる現実世界の課題、知識の活用を求める課題、明確な評価基準の提示と学生の自己評価、学習の改善に向けた個別で継続した教員からのフィードバックにより、学生に深い学びや省察を促す「学習としての評価」として機能していることが確認されており（小野他, 2018）、このような形成的評価としての機能をもつ MTJ を系統的にカリキュラムに組み込んでいることも、問題解決能力を向上させた要因といえるかもしれない。なお、ここでの「学習としての評価」は、Earl（2013）よりも広い意味で用いており、単に評価が学生自身の学習の自己調整になるというだけでなく、さらに評価それ自体が学習経験になること（Alverno College Faculty, 1994；松下, 2016）を意味している。

5. 結論

本研究では、高次の統合的能力をプログラムレベルでどう直接評価するかという課題に対し、保健医療福祉分野における問題解決能力に焦点をあて、新潟大学歯学部口腔生命福祉学科の PBL カリキュラムを事例としながら実証的な検討を進めてきた。まず、PBL を集中的に行っている第 2 学年から第 3 学年の異なる 3 時点で、MTJ というパフォーマンス評価を用いて問題解決能力を直接評価した。そのうえで、評価結果を比較したところ、学年が進むにつれて、「合計得点」は有意に、また十分な効果量をもって上昇していることが明らかになった（図 4、表 2）。

この変化を、第 2 学年から第 3 学年にかけて問題解決能力が向上したと解釈するには、厳密には「合計得点」の等化が行われている必要がある。しかし、教育実践において、テスト理論における等化の手続き（加藤他, 2014）にしたがって、複数のパフォーマンス課題の得点を等化させて、その変化を検討することは現実には困難である。そこで、本研究では、課題の難易度の違いについて、課題の複雑さを病状への関連要因と登場人物間の関係という側面から概念的に検討するとともに、当該カリキュラムで学んだ卒業直前の学生に、過去に体験した 3 つの評価課題の難易度について振り返ってもらうという形で、学生自身の認識の側面からも実証的に検討した。このような課題の難易度の検討結果を、学生の各課題における「合計得点」と合わせて考えることによって、「合計得点」の上昇を問題解決能力の向上と解釈できると判断した。仮に、課題の難易度が下がっているのであれば、「合計得点」の上昇を問題解決能力の向上とは解釈できないが、課題の難易度が同じか上がっているのであれば、「合計得点」の上昇は問題解決能力の向上と解釈できるからである。このように、縦断的なアプローチで能力の変容を捉えるには、得点の変化だけでなく、課題の難易度にも着目する必要があることが明らかになった。



（注）セルの色の濃さは課題の難易度を表わす

図 8　課題の難易度と評価得点の関係模式図

AAC&U の VALUE ルーブリックに代表されるように（Rhodes, 2010）、複数の学年・科目・課題をまたぐ長期的ルーブリックを使って、学生の長期的な変容を捉えようという試みが少しずつ広がっている。だが、そこで、課題の難易度の違いをどう処理するかという問題は不問に付されている。本研究はその問題に対して一つの方法を提案したものである。

複数科目をまたいだ学習成果の評価をめぐるこれまでの考え方は、学士課程の 4 年間（ないし 6 年間）を通じた学生の能力の変容を、長期的ルーブリックでの評価得点の向上によって捉えようとするものであった。だが、そのような考え方は、本来、評価課題の難易度が一定であると仮定した場合（図 8a）でなければ成り立たない。これに対し、通常の教育実践では、パフォーマンス評価は各科目の授業内容と結びついた「埋め込み型評価（embedded assessment）」（Suskie, 2009）として行われることが多いと推察され、この場合は、科目の違いをこえて課題の難易度を一定にすることが困難である。むしろ、低学年から高学

年にかけて、パフォーマンス課題の難易度が高くなるほうが自然であり、その場合には、課題の難易度を一定としてルーブリックのレベルが向上したかをみるよりも、課題の難易度の上昇をふまえたうえで、各科目の目標について一定のレベルに達したかをみるほうが重要となる（図 8b）。複数科目をまたいだ学習成果の評価にあたっては、本研究での検討のように、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、すなわち課題の難易度と到達レベルの 2 つの側面から、学生の能力の変容を慎重に捉えることが必要であろう。課題の難易度という観点を加えることによって、複数科目でのパフォーマンス評価の結果から、高次の統合的能力の長期的な変容を把握することが可能となる。

謝辞

本研究において、調査に協力してくださった新潟大学歯学部の教員の皆さま、学生の皆さまに感謝します。

本研究は、JSPS 科研費 JP15H03473、JP18H00975 の助成を受けたものです。

引用文献

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine*, *68*, 52–81.

Alverno College Faculty (1994). *Student assessment- as-learning at Alverno College*. Milwaukee, WI: Alverno College Institute.

Association of American Colleges and Universities (2007). *College learning for the new global century: A report from the National Leadership Council for liberal education & America's promise*. Washington, DC: Association of American Colleges and Universities.

Banta, T., & Pike, G. (2012). Making the case against – One more time. In R. Benjamin et al., *The seven red herrings about standardized assessments in higher education* (pp. 24–30). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Barrows, H. S. (1998). The essentials of problem-based learning. *Journal of Dental Education*, *62*, 630–633.

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/b_menu/shingi/chukyo/chukyo0/toushin/1217067.htm）（2020 年 3 月 23 日）

Clark, C. E. (2006). Problem-based learning: How do the outcomes compare with traditional teaching? *British Journal of General Practice, 56*(530), 722–723.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: Research and theory. *Academic Medicine*, *75*, 259–266.

Dochy, F., Seger, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*, 533–568.

Earl, L. M. (2013). *Assessment as learning* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research* (6th ed.). White Plains, NY: Longman.

畑野　快・上垣友香理・高橋哲也（2015）.「アクティブラーニングの経験は学修成果と関連するのか―3 年間の学士課程教育における両者の変化に着目して―」『大学教育学会誌』37, 86–94.

平井洋子（2006）.「測定の妥当性からみた尺度構成―得点の解釈を保証できますか―」吉田寿夫（編）『心理学研究法の新しいかたち』誠信書房, pp. 21–49.

Hung, W., Jonassen, D. H., & Liu, R. (2008). Problem- based learning. In M. Spector et al. (Eds), *Handbook of research on educational communications and technology* (3rd ed.) (pp. 485–506). Hillsdale, NJ: Erlbaum.

加藤健太郎・山田剛史・川端一光（2014）.『R による項目反応理論』オーム社.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*, 746–759.

Koh, G. C., Khoo, H. E., Wong, M. L., & Koh, D. (2008). The effects of problem-based learning during medical school on physician competency: A systematic review. *Canadian Medical Association Journal*, *178*, 34–41.

松下佳代（2010）.「学びの評価」佐伯胖（監修）・渡部信一（編）『「学び」の認知科学事典』大修館書店, pp. 442–458.

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（2015）.「ディープ・アクティブラーニングへの誘い」松下佳代・京都大学高等教育研究開発推進センター（編著）『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. 1–27.

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，pp. 3–25.

村山　航（2012）.「妥当性概念の歴史的変遷と心理測定学的視点からの考察」『教育心理学年報』51, 118–130.

Neville, A. J. (2009). Problem-based learning and medical education forty years on. *Medical Principles and Practice*, *18*, 1–9.

新潟大学歯学部（2019）.『専門科目シラバス（講義概要）』新潟大学歯学部 .

新潟大学歯学部（2020）.『Passion for Dentistry 2020』（https://www.dent.niigata-u.ac.jp/d/passion2020.pdf）（2020 年 3 月 23 日）

新潟大学歯学部口腔生命福祉学科（2020）.『歯学部口腔生命福祉学科の三つのポリシー（口腔保健福祉学）』（https://www.niigata-u.ac.jp/information/2020/68112/）（2020 年 4 月 28 日）

Norman, G. R., & Schmidt, H. G. (2000). Effectiveness of problem-based learning curricula: Theory, practice and paper darts. *Medical Education*, *34*, 721–728.

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（編著）『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（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, 123–132.

小野和宏・斎藤有吾・松下佳代（2018）.「PBL を評価する改良版トリプルジャンプにおける「学習としての評価」の要因」『京都大学高等教育研究』24, 35–44.

Rhodes, T. (2010). *Assessing outcomes and improving achievement: Tips and tools for using the rubrics*. Washington, DC: Association of American Colleges and Universities.

斎藤有吾・小野和宏・松下佳代（2015）.「PBL の授業における学生の問題解決能力の変容に影響を与える学習プロセスの検討―コースレベルの直接評価と間接評価の統合―」『大学教育学会誌』37, 124–133.

杉谷祐美子（2019）.「学部調査にみる日本の教養教育の動向」『IDE』610 号（2019 年 5 月号），35–40.

Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass. サスキー，L.（2015）.『学生の学びを測る アセスメント・ガイドブック』（齋藤聖子訳）玉川大学出版部.

山田礼子（2012）.『学士課程教育の質保証に向けて：学生調査と初年次教育から見えてきたもの』東信堂.

吉田　文（2008）.「学際的カリキュラムの陥穽―人文・社会系学部の学士課程カリキュラム―」『名古屋高等教育研究』8, 155–172.

Educational Practice Report

Direct Assessment of Long-Term Learning Outcomes in the PBL Curriculum: Based on a Longitudinal Study of Performance Assessment at the Niigata University's Faculty of Dentistry

Kazuhiro Ono[1], Kayo Matsushita[2], and Yugo Saito[3]

([1]Graduate School of Medical and Dental Sciences, Niigata University, [2]Center for the Promotion of Excellence in Higher Education, Kyoto University, [3]Headquarters for Management Strategy, Niigata University)

The purpose of this study was to examine the transformation of students' problem-solving ability, one of the higher-order skills, by performance assessment at the educational program level. The subjects were 82 students at the Niigata University's Faculty of Dentistry, who had studied under our problem-based learning (PBL) curriculum. For direct assessment of problem-solving ability, we used modified triple jump (MTJ) which had been developed as an assessment tool for our PBL program. We performed MTJ at three different time points during the second and the third year and assessed problem-solving ability of the students longitudinally. At the same time, we also examined the difficulty level of the three performance tasks used in MTJ from two aspects, conceptual contents and student perceptions. As a result, the assessment scores of MTJ improved significantly with moderate to large effect size along with their experience of PBL. The performance tasks used in MTJ became more complex along with the accumulation of learning contents in the curriculum, and student perceptions of difficulty level of the performance tasks showed similar results. In conclusion, the reason for the score improvement was not because the tasks of MTJ became easier but because problem-solving ability of the students improved. By considering the difficulty level of the performance tasks, we were able to assess the transformation of students' problem-solving ability at the educational program level from the results of embedded performance assessment in PBL courses.

Keywords: PBL, Curriculum, Learning outcomes, Performance assessment, Medical and welfare area