

12 プログラム評価^{*1}

山脇 正永^{*2}

1. はじめに

プログラム評価は学修者評価と並んで、医学医療教育機関及び教育者にとって重要であることは論を待たない。特に近年では、内部質保証、外部質保証の観点からも構造化されたプログラム評価が求められている¹⁾。本稿ではプログラム評価について、その理論的背景と主要な評価方法についてまとめた。

2. カリキュラムとプログラム

教育の目的に合わせて一連の学修内容・教育内容を表したものとして、プログラムあるいはカリキュラムという用語が用いられている。

プログラムとは、一定の教育目的の下で授業科目（コース course）や学習活動などが一定のまとまりをもって組織化されたもので、一連の授業単位、ユニットやコース、臨床実習単位から、より大きな範囲である卒後研修や卒前教育全体まで、様々な範疇で用いられる。カリキュラムも学修者に与えられる学習経験の総体を意味しているが、行政用語としては教育課程の意味として用いられている。また、制度によってプログラムあるいはカリキュラムを別々の定義づけを行っていることもある（例えば、専門医制度では別に定義づけされている）。本稿では「プログラム評価」として統一して稿を進める。

プログラム評価は教育・学修の指標として不可欠のものであり、十分に計画して実施すべきものである。

3. プログラム評価の考え方

プログラム評価の基本は「変化」を評価することである²⁾。評価の対象として最も重要なのが学修者の変化であることは明らかであるが、プログラムには、学修者、教育者、教育管理者、他職種など、様々な内外のステークホルダーが関与しており、これらのステークホルダーにとっての評価もプログラム評価に重要である³⁾。また、変化に関する評価としては、変化の有無、変化の内容、変化による効果などの視点が指標となるが、変化についてはあらかじめ意図したものだけでなく、意図しないものも含まれる。また、プログラムとそれを取り巻く環境は絶えず変化している動的なものであるため、プログラム評価の結果をリアルタイムにプログラムに反映させることも重要である。

以前は、プログラム評価は単純な因果関係（線形関係）として、原因と結果の構図で、プログラムの要素とアウトカムの関係としてとらえられていた。しかしながら近年では、教育プログラム評価は様々な内的・外的要素を包含した複雑系としてとらえられており、必ずしも線形関係、因果関係の構図としてはとらえられていない。

本稿ではアウトカム基盤型教育（OBE）が出現した後のプログラム評価について論を進める。

4. プログラム評価の定義

プログラム評価について ACGME では、“Systematic collection and analysis of information related to the design, implementation, and outcomes of a program, for the purpose of monitoring and improving the quality and effectiveness of the program.” と定義されている⁴⁾。プログラム評価には、プログラムの全体

^{*1} Program Evaluation

^{*2} Masanaga Yamawaki 東京医科歯科大学大学院医歯学総合研究科臨床医学教育開発学

像の理解, プログラムの目指す目的, 次のアクションへのつながりを明らかにすることが含まれる. さらに, プログラム評価は, 望ましい結果だけでなく望ましくない結果も含めて, プログラム内外に関連するアウトカムを評価しようとするものである⁴⁾.

プログラム評価に必要な情報として通常用いられるデータは, 種々の測定プロセスによって得られるものである. プログラム評価を実施する際には, 教育プログラムの長所や短所をデータとして明らかにできるよう, 測定ツール, 測定方法, 評価法の選択が重要であり, この選択には多くの要素が影響する. また, プログラム評価においては, 学修者評価 (個々の学修者の評価) のデータはもちろん重要であるが, これ以外の評価項目 (教育者に関する評価, 教育環境に関する評価, 教育マネジメントに関する評価等) も重要である.

5. プログラム評価の目的

教育者にとってプログラム評価の目的には, 内部的なものと外部的なものがある. 内部的なものとしては, 現行の教育カリキュラムの課題発見, 教育カリキュラム改革時の評価, 学生・教員や大学運営部門へのアカウンタビリティに関するものなどである.

外部的な目的としては認証評価等の外部質保証がある. 認証評価については本稿では詳細には扱わないが, 構造化された評価プロセスによるプログラム評価は社会的アカウンタビリティに関わるものとなっている. 卒前の医学教育プログラム評価としては, 国際的には世界医学教育連盟 (WFME), 北米医学教育連絡協議会 (LCME) の認証評価があり, わが国では日本医学教育評価機構 (JACME) による医学教育分野別評価, 大学改革支援・学位授与機構等の機関別認証評価が実施されている. 卒後教育においては, 米国卒後研修認定評議会 (ACGME) では各プログラムに Program evaluation committee (PEC) の設置が義務付けられており, 日本では卒後臨床研修評価機構 (JCEP) による臨床研修プログラム評価が導入されている.

これらの認証評価においても後述するプログラム評価方法が使用されている.

これら内部質保証及び外部質保証に必要なプログラム評価においては, 教学 IR (Institutional Research) 部門が重要な役割を果たしている.

6. プログラム評価の理論的背景

教育領域のデータ取得及び解釈は, 科学研究デザインで実施する場合と異なり種々の制限があることは日常我々が経験することである. プログラム評価についての考え方も, 科学的で還元主義的な考え方のみならず, 一般システム理論や複雑性理論に基づく包括的で統合的な評価方法も用いられてきている²⁾.

還元主義的な考え方では, アウトカムやカリキュラムから得られた結果の全体像は, プログラムの構成要素や因子の総体であるという考え方に基づく. プログラムの評価結果は, それぞれの構成要素の因果関係に則った線形の結果と考える. プログラムのある部分の変更が, 対応したアウトカムに影響するという考え方である. これらは, ロジックモデルや Before, During and After モデルとして知られている. これらの評価方法としては, 始まりから終了までのロジックの流れ, インプットとアウトプットの因果関係に注目することとなる. アウトカムにかかわる要素 (教育内容) の欠落が明らかになった場合には, その要素をプログラムに盛り込むことが必要となるという考え方である. この還元主義的な考え方は, 現在でもある部分については有効である.

その後システム理論に基づく考え方が出現したが, これは還元主義的考察に対して, アウトカムは単なるプログラム構成要素の集合ではなく, それら要素の内的及び外的な相互関連性やプログラムの置かれた環境 (コンテキスト) が重要であるとする考え方である. これは Bertalanffy の提唱した一般システム理論を基盤としたもので, それぞれの要素の相互関係及びそれらを取り巻く環境は常にダイナミックに変化しており, この相互作用と変化をシステムとしてモデル化し, 全体像を把握するという考え

方である²⁾。システム理論では教育プログラムは生物と同様に、外部からの影響を受け変化する開放系と位置付けられており、開放系システムの特徴として等結果性が知られている。初期状態として様々なバラエティのあったものでも、最終的な状態やアウトカムは同様の結果となり、そこまで到達するには様々な道筋・方法があるというものである。この理論は我々の日常の教育経験ともよく合致するものである。

複雑系理論はシステム理論を一步進めて、プログラム全体は部分の寄せ集め以上の効果があることを前提としたものである。教育プログラムは絶えず変化しており、例えば、学修者や教育者の特性、ステークホルダーや規制の影響、不断に変化する医学的知識、そのプログラムがおかれた状況（COVID-19による社会の変容など！）などによる、内部的なあるいは外部的な影響を常に受けている。医学教育プログラムはこの意味で複雑系であり、様々な影響要因から動的な影響を受けており、不確実性と不透明性の部分も十分に考慮して検討を行うべきものである。プログラムの成功は学修者の成果だけでなく、ステークホルダー相互の関係性、環境や社会との関係性についても述べるべきであり、プログラムの持つ多義性を検討すべきである。複雑系理論に基づく考え方は我々教育者にとってもプログラム評価の単純化／短絡化を防ぐ効果もある。この複雑性理論はプログラムのプロセスの評価のみならずコンテキスト評価

を重視し、プログラム因子の関係性を評価するCIPPモデルや、Before, During and After (BDA)モデルもこの範疇である⁶⁾。

7. プログラム評価の実際

代表的なプログラム評価の方法については表1のようにまとめられる。実際のプログラム評価ではこれらの手法を組み合わせで使用することが重要である²⁾。還元主義的方法に代表される実験的方法であっても、条件が整えば有効な評価ツールとなりうる。

(1) 実験的／準実験的方法

実験的／準実験的方法（experimental/quasi-experimental methods）は古くから用いられてきた方法であり、科学をバックグラウンドとする我々にもなじみが深いものである。手法としてはプログラムの構成要素とアウトカムの関係性を、還元主義的に線形の因果律を用いて検討するものであり、対照や条件を統一した注意深い研究デザインによってはじめて成り立つ。しかしながら生物学的研究と異なり、複雑な教育プログラムを評価する場合は、真に実験的方法がとりにくいこと、コントロールを置くことが困難であることなどのリミテーションに直面する。この弱点を克服するために、ランダム割付を考慮せず介入群と対照群を比較する準実験的方法がとられることも多い。しかし、実験的／半実験的手法がうまくデザインされる場合には

表1 プログラム評価方法の分類と理論的背景

Evaluation Category	Subcategory	Reductionism	System theory	Complex theory
Experimental/Quasi-experimental	Intact Group design	○		
	Time-series design	○		
	Ex Post Fact design	○		
Kirkpatrick model		○	○	
Logic model	Logic model	○	○	
	Bennett model	○	○	
Donabedian model			○	
CIPP			○	○

有効であることも事実である。

Intact-Group デザイン：ランダム化比較試験様式にデザインして2群（例えば、プログラム1群とプログラム2群）に分け、それぞれの個人成績の比較評価を行うものである。2つの集団間で意図しない差異（バイアス）が隠れている可能性があり、理想的には集団を変えて複数回の評価を行うべきであるが、現実的には困難なことが多い。本方法はインプットとアウトプットの因果関係が明白で、ランダム化が可能であれば有効な結果を導くことができる。

時系列分析（Time-series Experimental Design）：評価する対象を時系列に沿って評価するものであり、プログラム前後の比較評価などがこれにあたる。本方法は知識及び行動を対象として、短期間での変化を評価するのに適している。本方法では、学修者個人の修得度の差や同時に行われている他のプログラムの影響など外的な要素を排除することはできない。また、旧カリキュラムと新カリキュラムの比較など異なるグループ（学年）を比較する方法もこれに含まれる。本方法ではプログラム要素とアウトカムの因果関係が明白で、短期間にアウトカム測定が可能であり、比較するグループ間が比較的均一である場合に適応できる。

事後分析（Ex Post Fact Experiment design）：本方法も科学研究デザインで頻用されるものである。対象をランダム化せずにアウトカムの変数を共変量（説明変数）あるいは交絡因子としてモデル化するものである。このモデル化には予想される重要なデータを共変量としてあらかじめ評価項目として設定しておかなければならず、また、強い相関がなければその因子は無関係であるとされてしまう危険がある。

(2) カークパトリック（Kirkpatrick）モデル

Kirkpatrick モデル⁷⁾は医学教育領域で頻用するものであり、プログラムのアウトカム評価として実用的な分類法を提唱している（図1）。

その評価は、レベル1) 学修者の満足度、プログラムへの反応、レベル2) プログラムで予想される能力の獲得、レベル3) 学修者の行動変容、

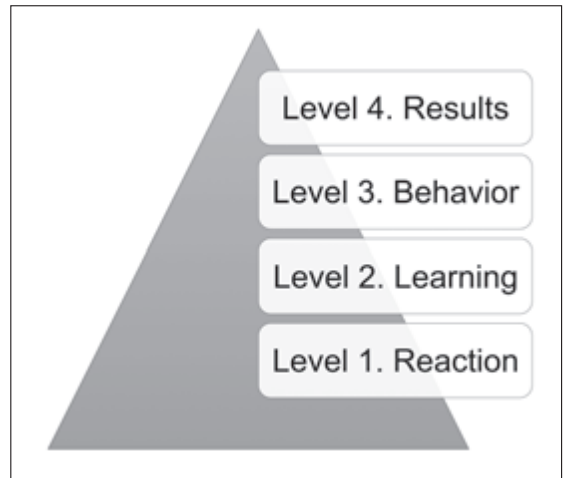


図1 Kirkpatrick モデル

レベル4) より大きな文脈でのアウトカム、の4段階の評価プロセスでまとめられている。このモデルでは、特に第3段階と第4段階が重要と考えられる。第3段階では、学修者のコンテキストの中での行動変容を評価することが挙げられており、学修者評価としての Workplace-based Assessment (WBA) もその評価において重要であることが認識される。また、第4段階のアウトカムとしては、患者アウトカム、経済的アウトカム、医療の効率性のアウトカムなどよりインパクトのある指標が位置付けられている。Kirkpatrick モデルへの批判としては、学修者のモチベーション、学修者のレディネス、プログラムに関する文脈の多様性、教育資源の多様性などが考慮されていない点、教育プログラムの全体像の評価やプログラムの効果の理由については明らかにできない点が課題とされている。しかしながら、他のモデルと併用することによりプログラム評価の有力なツールになることは事実である。

(3) ロジックモデル

ロジックモデルには一般システム理論の影響がみられ、プログラム要素および各要素の相互関係性を重視している評価方法である。一方で、これまでの理論と同様に線形の因果関係あるいは論理的な構造に沿ってプログラム評価を行うものである。本方法はプログラムの変化のプロ

セスを重視するものであり、プログラムのプランニング段階でも有用であり、認証評価等のプログラム評価でも用いられている。インプット、アクティビティ、アウトプット、アウトカムの4段階のプロセスを評価するものである。インプットについては、教育資源、教員のスキル・時間、テクノロジー及び教育文化などすべての要素の記載が求められる。アクティビティについては教育プログラムの内容の記載となる。アウトプットはプロダクトとも言い換えられるが、数的な評価である。アウトカムは短期、中期、長期の結果も含むものとなっている。

ロジックモデルはプログラム改定や新プログラム作成の計画段階において非常に有用であり、プログラムの社会的、文化的、政治的な文脈に沿って評価を行うことができる。ただし、構造上ロジックモデルは線形の因果関係の評価とならざるを得ない。ロジックモデルはあらかじめ計画されたモデルの実証となるので、アウトカムから得られた新知見を解析することや、そこからあらたな仮説を提唱することは困難であることが多い⁶⁾。

ロジックモデルの類型として、Bennett's Hierarchy モデル (inputs, activities, participations, reactions, KASA, practice change, impact)⁹⁾ も提唱されている。

(4) CIPP (Context/Input/Process/Product) モデル

CIPP モデル¹⁰⁾ は Stufflebeam により提唱されたもので、複雑で動的な教育プログラムの理解と評価に用いられている。CIPP はプログラムを4つの評価軸、1) コンテキスト評価、2) インプット評価、3) プロセス評価、4) プロダクト評価、に沿って、教育プログラムを総体として評価しようというものである。現在の認証評価は、この内容に沿って実施されていることが多い。CIPP の具体的な評価内容・要素についての一例を表2に示す。

(5) その他の評価方法

Donabedian により提唱されたプログラム評価方法として、ストラクチャー評価、プロセス評価、アウトカム評価の3段階で分析する方法がある (Donabedian モデル)¹¹⁾。プロセス評価の後にパフォーマンス評価を加えることもある。ヘルスケア領域の quality measure としても用いられる評価方法であり、プログラムや政策を包括的に評価する方法である。プログラムの構造、プログラム内容のプロセス (時間的プロセスも含む)、パフォーマンスを含むアウトカムについてそれぞれの評価内容を分析し、評価の階層を構造化したものである。

8. 新たなプログラム評価の展開

COVID-19 パンデミックに対応した評価、システム思考による評価の報告もなされている¹²⁾。本報告では、上記の一般システム論をさらに進

表2 CIPP モデルによる評価内容

Context evaluation study	Input evaluation study	Process evaluation study	Product evaluation study
- Document review - Demographic data analysis - Interviews - Surveys - Record analysis	- Literature review - Visiting exemplary programs - Consulting experts - Inviting proposals addressing the identical needs	- Observation - Document review - Participant interviews	- Stakeholders' judgment - Comparative studies of outcomes - Assessment of achievement - Group interviews - Case studies - Surveys - Participant reports

めたシステム思考 (system-thinking) により, 教育プログラムの全体像を捉え, 様々な要素とのつながりを把握したうえで, 最も効果的な解決 (最適解) へ向かうという評価アプローチを提唱している. 本報告では, COVID-19 により変容した環境が教育プログラムに与えた影響を構造的に評価し, この変化にリアルタイムに対応する教育プログラムの必要性を論じている. Programmatic assessment は学修者評価で頻用されているが, programmatic program evaluation も提唱されている¹³⁾. Programmatic assessment とは, プログラム中の学修者の progress のベンチマークとなるデータの種類及びその取得ポイントをあらかじめ設定しておき, 学修者の評価を経時的に行うものである. このコンセプトと同様に, 教育プログラムについてもプログラムのコンテキストやステークホルダーのニーズに応じて, あらかじめ時系列的に評価を行ってゆこうというものである. 本取り組みについてはいまだ報告はないものの, 速く大きな社会変容に応じた医学教育プログラムの評価方法の1つとなる可能性があることが論じられている.

医療を取り巻く社会や環境が大きく変容している時代に, 医学教育プログラムの計画・運用を行うことは容易ではない. しかしながら, 学修者及びパブリックに対しての説明責任を果たし, さらにより良いプログラムを作成するためにも, プログラム評価は今後ますます重要な位置付けとなると考えられる.

■文 献

- 1) Haji F et al. Rethinking programme evaluation in health professions education : beyond 'did it work?' *Med Educ* 2013 ; **47** : 342-51. doi : 10.1111/medu.12091.
- 2) Frye AW and Hemmer PA. Program evaluation models and related theories : AMEE guide no. 67. *Med Teach* 2012 ; **34** : e288-99. doi : 10.3109/0142159X.2012.668637.
- 3) Cook DA. Twelve tips for evaluating educational programs. *Med Teach* 2010 ; **32** : 296-301. doi : 10.3109/01421590903480121.
- 4) https://www.acgme.org/Portals/0/PDFs/ab_ACGMEglossary.pdf?ver=2018-05-14-095126-363.
- 5) Schwartz AR et al. A novel approach to the program evaluation committee. *BMC Med Educ* 2019 ; **19** : 465. doi : 10.1186/s12909-019-1899-x.
- 6) Durning SJ et al. The Structure of Program Evaluation : An Approach for Evaluating a Course, Clerkship, or Components of a Residency or Fellowship Training Program. *Teach Learn Med* 2007 ; **19** (3), 308-18. doi : 10.1080/10401330701366796.
- 7) Kirkpatrick D. 1996. Revisiting Kirkpatrick' s four-level model. *Train Dev J* : 54-9.
- 8) Onyura B et al. Five ways to get a grip on the shortcomings of logic models in program evaluation. *Can Med Educ J* 2021 ; **12** : 96-9. doi : 10.36834/cmej.71966.
- 9) Deadrick DL et al. Using hierarchical linear modeling to examine dynamic performance criteria over time. *J Management* 1997 ; **23** : 745-57. doi : 10.1016/S0149-2063 (97) 90027-1.
- 10) Stufflebeam D, Shinkfield A. 2007. Evaluation theory, models, & applications. San Francisco : Jossey Bass/John Wiley & Sons, Inc.
- 11) Donabedian A. Evaluating the Quality of Medical Care. *Milbank Q* 2005 ; **83** : 691-729. doi : 10.1111/j.1468-0009.2005.00397.x
- 12) Colbert-Getz J et al. Times of Medical Education Crisis Require New Evaluation Approaches : Proof of Concept of a System-Based Program Evaluation Model. *Med Sci Educ* 2021 17 ; **31** : 1-6. doi : 10.1007/s40670-021-01271-5.
- 13) Sinani S and Naamani K. Programmatic Evaluation : A Prospect in Program Evaluation Design. *Oman Med J* 2021 ; **36** : e260. doi : 10.5001/omj.2021.122.