

放課後プログラムが生徒のアウトカムに与える影響  
システマティック・レビューのためのプロトコル

Susan Goerlich Zief

Sherri Lauver 博士

Rebecca Maynard 博士

## I. 表紙

### タイトル

放課後プログラムが生徒のアウトカムに与える影響

### レビューワ

#### **Susan Goerlich Zief**

Graduate School of Education University of Pennsylvania 3700 Walnut Street Philadelphia,  
PA 19104 Phone: 215-898-1974 Email: zief@dolpin.upenn.edu

#### **Sherri Lauver, Ph.D<sup>1</sup>**

Institute of Education Sciences

U.S. Department of Education 555 New Jersey Ave, NW Suite 502A Washington, DC  
20208 Phone: 202-208-7162 Email: Sherri.Lauver@ed.gov

### レビューワへのアドバイザー

#### **Rebecca A. Maynard, Ph.D.**

University Trustee Professor of Education and Social Policy Graduate School of Education  
University of Pennsylvania 3700 Walnut Street Philadelphia, PA 19104 Phone:  
215-898-3558 Email: rmaynard@gse.upenn.edu

### 助成元

Smith Richardson Foundation

Hewlett Foundation

### 提出日

2004年5月9日

---

<sup>1</sup> Dr. Lauver は、キャンベル共同計画の試験的レビュープロジェクトのコンサルタントとして、ペンシルベニア大学に在籍した折に、この仕事に従事した。

## II. レビューの背景

アメリカ国内全体で、5歳から14歳までの子どものうち、約800万人が、放課後監督者がいない状態に置かれており（NIOST [国立放課後時間研究所]、2003年）、低・中所得世帯の子どもの3分の2は、親の仕事の関係上、放課後親の目が届かない状態にある（Long および Clark、1998年；米国労働統計局、2000年）。低所得世帯の者は、働かなければ公的支援制度を受けられないし、低・中所得世帯と父子・母子世帯に相関があるとすれば、この統計データは驚くにはあたらない。

調査研究の結果、こうした親の目の届かない時間が、子どもの危険を顧みない行動を増やし、被害者数の増加や学習成績低下と結びついていることは明らかである（Dwyerら、1990年；Newmanら、2000年；Osofsky、1999年；Posner および Vandell、1999年；Richardsonら、1989年；米国保健社会福祉省、1995年、米国教育省および米国司法省、2000年）。政策立案者も社会全体も、組織化されず監督を欠いた放課後は、「リスクと機会」（Hofferth、1995年）を孕んだ時間と考えるようになり、同時に、放課後プログラムが、都市部の低所得世帯の子どもの望ましくない行動を減らし、プラスのアウトカムを増進する措置として、注目されるようになっている。

ここ数年の放課後プログラムの展開は、目を見張るものがある。連邦政府、州、地方自治体、および民間財団が、プログラムに多額の資金と資源を投じるようになった。たとえば、1998年には4,000万ドルだった21世紀コミュニティ学習センター（21<sup>st</sup> Century Community Learning Centers）のプログラムに対する助成金は、現在10億ドル近くまで増えている。

この短期間に、放課後プログラムの支援団体はその数を増やし、力を伸ばした。こうした団体の影響力を示す例が、最近発表された21世紀コミュニティ学習センター（CCLC）の第1回全国評価の結果に対する反響である（米国教育省、2003年）。この報告<sup>2</sup>に対しては一部から批判もあったが、効果がないに等しいとする報告結果への強い反発というより、21世紀コミュニティ学習センター（CCLC）への助成金の40%削減を勧告するために、単独の機関が公然と実験的な調査研究を行ったことに対する反発とも考えられる。

この結果、放課後プログラムへの多額の投資やプログラムの拡大を継続的に支えているのは、大規模研究や権威を持った研究ではないことを世間に訴えるため、21世紀コミ

---

<sup>2</sup> プログラムのサンプリング、プログラムの実施状況、およびデータ収集の時間枠に対し、批判がなされた。

ユニティ学習センター (CCLC) への助成金の維持、または増額を目指すロビー活動や、草の根的な活動が盛んに行われた。放課後プログラムが、低所得世帯の子どもの健全な育成と、情緒面での発達を促すとともに、学習成績の向上や、生徒による犯罪・暴力行為への関与の低減に寄与すると思われる根拠として、よく引き合いに出される準実験や、非実験的研究はいくつか存在している (Baker および Witt, 1996 年; Foley ら, 2000 年; Huang ら, 2000 年; Jones および Offord, 1989 年; Le および Hamilton, 2001 年; McLaughlin および Irby, 1994 年; Posner および Vandell, 1994 年; Ross ら, 1992 年; Schinke ら, 2000 年; 米国教育省および米国司法省, 2000 年; Grossman ら, 2002 年; Welsh ら, 2002 年)。研究の一部は、プログラムに参加している子どもと不参加の子どもを比較している。だがこうした研究の計画では、プログラムに進んで参加した子どもとそうでない子どもの差異 (モチベーション等) を完全に比較できない。

どのような点が問題になるのか。近年の研究から、準実験の評価に強いバイアス (偏り) があることが明らかにされている。特にプログラムへの自主的な参加を対象とする研究では、この傾向が強くなる (Guyat ら, 2000 年; Agodini および Dynarski, 2001 年; Weisburd, Lurn および Petrosino, 2001 年; Wilson および Lipsey, 2001 年; Glazerman, Levy および Myers, 2003 年)。こうした理由から、プログラムに参加する子どもと参加しない子どものアウトカムの比較によって、プログラムが参加者の成長にどのように寄与したか、的確に評価することはできないのである。現在、準実験で実験結果が正確に反復できるかどうか、そしてどのような条件であれば反復できるかを評価するために、さらなる作業が進められている (Glazerman, Levy および Myers, 2003 年)。発生し得るバイアスにどう対処すべきか、そして (または) 実験計画に無理なく代わる準実験があるとすれば、どのようなタイプの研究か、といった課題について、放課後プログラムでの信頼ある回答を得られるまでは、プログラムの効果を問う質問に対し、信頼性のある影響評価と、そうではない評価を組み合わせるのは賢明な策ではないと思われる。

放課後プログラムに対し、質の高い実験のレビューから得られる知見はある。放課後プログラムに目覚しい発展が見られるようになったのは、エビデンスを提示するため、あるいは、プログラムが参加者の学習や行動上のアウトカム、社会性や情緒面でのアウトカムに与えていると思われる影響に対する理解を深めるために、プログラムの財源が拡大されてからのことである。目下の財源拡大により、レクリエーション、学習、および子どもの育成を目指す内容を組み合わせたプログラムを提供する地区や学校が増加した。従って、こうしたプログラムは、計画的な放課後プログラムとして子どもの大半の参加が見込まれる活動となり、政策立案者も大きな関心を寄せている<sup>3</sup>。

---

<sup>3</sup> 本レビューの「セクション IV : 方法論」では、現在増えているプログラムのうち、本レビュ

## 近年行われたエビデンスのレビュー

放課後プログラムが生徒のアウトカムに与える影響をめぐる近年の研究に対するレビューのうち7件が選ばれた。このうち3件は、子どもの健全な育成に目的を限定したプログラムを対象とするものである（Catalano ら、2002年；米国学術研究会議 [National Research Council]、2001年；Roth ら、1998年）。残り4件のレビューは、学習成績の向上など、上記以外の目標を加えたプログラムを検討対象としている。この4件のレビューは、放課後プログラムに対するわれわれの関心と密接に関わっているため、以下検討する。

「時間延長プログラムと放課後プログラムおよびその効果のレビュー」(*Review of Extended-Day and After-School Programs and their Effectiveness*) は、効果を裏付けるエビデンス、または効果の兆しが見られた主な項目、および／または普及と反復が可能な項目を持つプログラムを特定し、解説している（Fashola、1998年）。このレビューには達成度などのアウトカムを評価する実験および準実験（「よくマッチした実験群と対照群」P7）の両方が含まれている。Fashola は、有効なプログラムは多いものの、集められた研究には厳密な計画に基づくものはほとんどなく、ほぼすべてに選択バイアスが見られ、これらが確信的な結論を出す上で妨げとなっていると推断している。

Fashola がレビューで取り上げた情報や課題は、研究報告書の執筆時点における関連分野の状況を知るとともに、さらなる研究が求められる分野を提案するための指針として有用であった。ただし現時点で、放課後プログラムの指針として用いるとすれば、次の3点の限定が必要である。第一に、レビューに含まれる研究のすべてが、放課後プログラムに関連したものではない。著者は、放課後プログラムとしても反復可能であると述べているが、一部は通常授業の範囲で進められていたプログラムであった。第二に、評価にはメンタリング・プログラムと個別指導プログラムが含まれていたが、これはわれわれが従来型の放課後プログラムと定義するものとは基本的に大きく異なっている。第三に、影響がゼロ、または小さい研究がレビューにどの程度まで含まれているかが明確でない。

Eccles と Templeton (2002年) は、Fashola (1998年) のレビューに類似したレビューを行い、放課後プログラムの定義範囲を大幅に拡大して、メンタリング・プログラム、性教育プログラム、あるいは集中的に行なわれていないプログラム、(または) 短期プロ

---

一の観点から最も興味深いと思われるプログラムに限定している。

グラムを含めている。このレビューには、準実験と実験に基づく研究も含まれている。Eccles と Templeton も Fashola と同様、この領域の研究は誕生して間もないため、整合性に欠けているとしており、実験はごく小数で、研究の結果が同じになることは稀であり、影響に関する研究結果を解釈する上で、参考になる実施面のデータや、プロセス関連のデータが不足していると指摘している。だが、Eccles と Templeton は、実験と準実験から予備的な結論を引き出しており、「防止と促進を目的とする子どもに対するプログラムが行なわれていることで、現実にはプラスのアウトカムが増え、マイナスのアウトカムが減っていることを裏付けるエビデンスが増えつつある」(P172) と同時に、「学習指導を明確な目的としないプログラムでも、学習成績、出席率、ハイスクール卒業率の向上... (ならびに) 学校関連の問題行動の減少につながっている」(P172) ことを示唆している。

「放課後プログラムの評価：方法論のメタ評価および報告形式による研究結果の統合」(*Evaluations of After-School Programs: A Meta-Evaluation of Methodologies and Narrative Synthesis of Findings*) は、研究対象の選定基準として、上記と同じ基準を採用している (Scott-Little ら、2002 年)。レビューは、放課後プログラムは、共通テストの点数と宿題の達成率にプラスの影響をもたらし、子どもの年齢が低い場合ほど、あるいは学習面で「脱落の可能性が高い」生徒を対象とする場合ほど、有効性が高いことを示唆している。レビューには、実験と準実験の両方が含まれていることから、レビューアは、この両分野における因果関係については触れていない。だが、実験と準実験の 2 つの実験研究の結果から、放課後プログラムは、参加者の社会性や情緒面でのアウトカムにプラスの影響を与える場合があると述べている。ただし、こうした結果は、大半の子どもが体験した従来型プログラムの範囲を超えたプログラムから得られたものであることから、結果の一般化については疑問が残る。

Hollister (2003 年) は「放課後プログラムの発展：状況、問題点、および評価の影響」(*The Growth in After-School Programs: Status, Issues, and Evaluation Impacts*) の中で、放課後プログラムの効果について、レビューを行っている。前述のレビューアとは異なり、Hollister は、検索対象を実験に限定した。だが同時に「放課後プログラム」の定義範囲を拡大した (メンタリング、個別指導、補習授業、および総合サービスプログラムを含めた)。実験に基づく 10 件の研究をレビュー対象として検討した結果、メンタリング・プログラムと個別指導プログラムが、校内および放課後のアウトカムにプラスの影響を与えたこと、親の参加と親への指導がプログラムの一環として効果を挙げたこと、および生活技能面のカリキュラムが放課後プログラムのアウトカムの一部にプラスの影響を与える可能性があることが明らかになった。様々な特性を備えたプログラムを含めたため、学習プログラムに、レクリエーション・プログラムと青少年プログラムを組み合

わせた伝統的プログラムに参加する子どもの大多数に、この研究結果を当てはめるのは難しい。

近年行われた上記4レビューについては、重要な疑問が3点ある。

1. 過半数を超える子どもが参加しているプログラム、すなわち学習プログラムにレクリエーション・プログラムと青少年プログラムを組み合わせた伝統的プログラムのみをレビュー対象とすれば、レビュー結果は変わるか。
2. システマティックで一貫性のある方式により、質の高い研究を特定・選択すれば、レビュー対象となる研究の数は変わるか。
3. プログラムの総合的影響の評価に実験的評価のみを採用すれば、レビュー結果は変わるか。

#### 本レビューが貢献できる点

上記で著者が指摘しているように、急激に展開しているこの分野で入手可能な評価研究の数が限られているために、レビュー範囲が狭くなっている。このことは、レビューする研究の選定基準に関わる別の要因についての結論上、障害になった。上記レビューには実験と準実験、および／または計画・サービス提供面でまったく異なるプログラムの研究が含まれていたため、こうした研究を用いてプログラムの効果を問う質問に答えることは困難であった。

本レビュー案が前レビューと大きく異なる点が2点ある。第一に本レビューでは、放課後プログラムで近年発表された研究のうち、前レビューで取り上げられていないものを用いる。第二に、レビュー対象とする研究の選定基準については、傑出した研究に限定し、プログラムの効果など政策関連の問いに対して、より強い確証をもって答えられるようにし、また、より均質性に優れた一連のプログラム・モデルを活用する。

### **III. レビューの目的**

放課後プログラムの普及が進むとともに、プログラムの参加者のアウトカムが改善されるとの確信が強まったが、プログラムが目標の達成に近づいているかどうかを示すエビデンスはほとんどない。現在に至るまで、プログラムの拡大と改善をめぐる基本的な意思決定の指針となる科学的な知識基盤はごく限られている。

次のような問いに確実に答えられるか否かについて、また答えられる場合はその回答の

内容を知ることができれば、政策立案者とプログラム管理者は大いに役立つだろう。

- (1) 放課後プログラムへの参加は、学習上のアウトカム、社会性や情緒面でのアウトカム、および／または行動上のアウトカムに影響を与えるか。その場合、重要なアウトカムとプログラム・モデル全体の影響評価の範囲はどのようなか。
- (2) 放課後プログラムの有益度が高かった、あるいは低かったのはどのような子どもか。プログラム・モデルおよび対象の選定における変動範囲は一般化を行うのに十分か。
- (3) 評価を行ったプログラムのモデルと設定内容は、子どもへの有益性の点で他のプログラムに勝ると思われるか。こうしたプログラムの成功度を左右する特性は何か。

プログラムへの参加が子どものアウトカムに与える影響を評価する研究のうち、周到な計画のもとに実施された研究を特定・編集・分析するシステムティックな取組みを行う。このことを通じて、利用可能な研究を活用することにより、上記の問いに対して可能な範囲で最高の回答を提示するとともに、われわれの知識の限界を明確にすることができる。

選定基準を満たす研究については、システムティック・レビューとメタ分析によって次のことを行う。

- ・ プログラム、評価方法論、採用したアウトカムの評価法、および分析手法の記述。
- ・ 学習・行動上のアウトカム、および／または社会性や情緒面でのアウトカムについて有効性のエビデンスが見られたプログラムを特定し、これらのプログラムと実施項目の設定内容を別の評価対象プログラムと対比する。可能な場合には、厳密な実験計画に基づく研究から得られたプログラム別の推定により、プログラム全体がアウトカムに与えた影響を確認する。作業の中心となるこの分析を「ベンチマーク」として用い、今後の研究によりベンチマーク・レベルを引き上げることができる。
- ・ 可能であれば、目標や活動内容が異なるプログラムを対象とする研究のアウトカムとの比較によって分析を行う。生徒の様々な学習・行動上のアウトカム、および／または社会性や情緒面でのアウトカムを改善する上で、有効性が勝るのはどのような種類のプログラムかという問題に、政策立案者が寄せる関心は強まる一方である。
- ・ 可能であれば、プログラムが比較的有益であった対象を特定し、その特性およびプログラム体験を、プログラムの有益性を裏付けるエビデンスが見られなかった子どもの特性およびプログラム体験と比較する。

- ・ 前述の重要質問項目への回答の改善に役立つデータを、研究で提供できるようにするために、将来の評価案を作成する。

## IV. 方法論

### 選定基準

報告形式による本レビューおよびメタ分析の対象となる研究は、以下の基準を満たしていなければならない。

#### 介入の特性

プログラムは放課後に実施するものでなければならず、登校前の活動項目は含めても含まなくてもよい。プログラムは、学習障害、身体障害、情緒面の問題、または行動上の問題など特定かつ特殊なニーズを持つ子どもを対象とはしない。サマースクール・プログラムや重要な校内活動の一部をなすプログラムは対象外とする。プログラムは、学習、レクリエーション、および／または子どもの育成に関わる活動を提供しなければならない。ただし、プラスのアウトカムの達成を目指す基本的な手段が、個別メンタリングや個別指導の形式をとることがあってはならない。こうしたプログラムの大半は放課後に行われるが、プログラムの計画と提供に際して、プログラム担当教師／ボランティアと子どもとの関係はまったく異なると推定されることから、個別方式に基づくプログラムは本レビューの範囲外である。

言い換えると、本レビューで選ぶプログラムは、これまでのレビューで選ばれたプログラムの一部よりも、伝統的な放課後プログラムという大まかな範疇に入ろう。こうした伝統的なプログラムは、プログラムの財源拡大以来、めざましい発展を遂げ、その結果プログラムを提供する地域や学校が増えたと言ってよい。従ってこの種のプログラムは系統立った放課後プログラムとして、子どもの大半が参加していると思われる。こうした事情から政策立案者は、現在こうしたプログラムに多大の関心を寄せている。ただしわれわれは、プログラム・モデル間にはある程度の不均質性が存在すると予想している。たとえば、学習活動よりレクリエーション活動の方を数多く提供するプログラムもあれば、プログラム時間の大半を学習活動に費やし、レクリエーション活動は週1回というプログラムもあろう。

プログラムは、学校、コミュニティセンター、宗教施設など様々な環境で実施することができる。ただし報告結果の評価に際しては、プログラムの実施場所別に行うべきであ

る。最後に、本レビューの対象を北米で行われている介入の研究に限定する。<sup>4</sup>米国とカナダのプログラム以外は、介入の背景が相当異なると思われるため、国際的な研究は特にレビュー対象とはしない。ただし、この分野における今後のレビューアの参考になると判断した国際的研究はもれなく言及する。

### 母集団の特性

プログラムの参加者は、通常の公立または私立の幼稚園から12年生までの子どもであり、5歳から19歳までの子どもを対象とするプログラムを想定している。母集団の年齢範囲を広くしたのは、広範囲を対象とする研究が除外されないようにするためである。また該当すると思われるプログラムの一部は、13歳以上を対象として、危険を顧みない行動の防止を目的としている。

### 研究の種類

レビュー対象として検討する評価は、念入りに計画実施された実験研究を用いたものでなければならない。類似した介入を受けたことがあるグループを対象としてはならない。無作為割り当ての前に、親から調査研究への同意を得ておかななければならない。無作為割り当てを行った後で調査研究に同意する親には選択のバイアスが生じると考えられるからである。さらに、計画がどの程度厳密に策定されたかをレビューアが判断できるように、方法論の記述は明確かつ完全でなければならない。また研究にはプログラムの目的と活動に関する記述も求められる。

本レビューでは、2つの理由により、1982年以降に発表された研究のみを取り上げる。つまり、第一に、Lauver (2002年)のこれまでの検索では、これ以前の研究は特定されていない。第二に、1990年代までは、この分野のプログラムと評価に対する公的支援はほとんど存在しなかった。

### 研究の質

上記基準に合致する研究を、レビュー対象とする。次に、研究計画と実施プロセスの質の評価に用いる具体的な規準を適用する。プログラム・影響の評価が公平に行われている可能性が高いと見なされた研究だけが、プログラムの推定に関する統計報告に含めら

---

<sup>4</sup> Lauver (2002年)のこれまでの検索によると、米国とカナダを除き、関連分野の国際的研究は存在しないことがわかった。

れる。研究の質を評価する際の検討範囲については、Appendix A を参照されたい。

### アウトカムの評価

レビューでは、学習上のアウトカム（テスト成績、評点）、社会性や情緒面でのアウトカム（自信、自尊心、抱負、統制の所在）、および／または行動上のアウトカム（危険を顧みない態度、時間の使い方、出席率）を扱った研究を取り上げる。評価者がプログラムの明確な目的ではなかったアウトカムについての評価を行う可能性は承知している。ただし、こうしたアウトカムを分析から外すべきであるとは考えていない。これまでの研究とレビューでは、一見両立しない複数のプログラムの目的とアウトカムの評価結果との間の交差について仮説を立てるために、非実験的研究を用いた。一例を挙げると、Eccles と Templeton（2002 年）は、学習指導を明確な目的としないプログラムが、学習成績、出席率、およびハイスクール卒業率の向上につながることがあると示唆している。近年行われたこれ以外の複数の研究も、子どもの育成という目標が学習成績の向上を促す可能性があるという仮説を支持している（米国学術研究会議および医学研究所 [Institute of Medicine]、2002 年）。さらに、放課後プログラムのモデルの多くは、学習上のアウトカムと社会性や情緒面でのアウトカムの両方を含み、両者は密接な関係にある可能性を認めている（ハーバード家庭研究プロジェクト [HFRP: Harvard Family Research Project]、2003 年；Lauver、2002 年；Dynarski ら、2001 年）。

レビューでは、結果の報告に際して、学習上の目標を設定しなかったプログラムに、そうした目標の達成を義務付けないよう配慮する。同様に、行動上の目標を設定しなかったプログラムに、そうした目標を義務付けるべきではない。ただし、政策上きわめて重要なのは、どのようなアウトカムが達成されつつあり、どのようなタイプのプログラムでアウトカムが達成されているかを把握しておくことである。

### 準実験的研究の利用

近年のレビューと研究で力説されているのが、準実験に基づく評価に影響を与える強いバイアスはどうやら避けられないということである。これは、特に、放課後プログラムのような子ども向けプログラムへの自主的参加に関する研究で見られる傾向である（Guyat ら、2000 年；Agodini および Dynarski、2001 年；Weisburd、Lum および Petrosino、2001；Wilson および Lipsey、2001 年；Glazerman、Levy および Myers、2003 年）。しかしながら、放課後プログラムの介入に対する実験的評価の可能性は、資金や実現可能性のために従来から制約されてきたため、同分野の研究は準実験で非常に多く見られる。

現在、準実験で実験結果が正確に反復できるかどうか、どのような条件であれば反復できるかを評価するために、さらなる作業が進められている（Glazerman、Levy および Myers、2003 年）。発生し得るバイアスにどう対処すべきか、そして（または）実験に無理なく代わる準実験があるとすればどのようなタイプの研究か、といった課題について、放課後プログラムでの確証が得られるまでは、信頼性のある影響評価とそうでない評価を組み合わせるのは賢明な策ではないと思われる。しかしながら、Glazerman、Levy および Myers（2003 年）がまとめた下記基準を満たす準実験の参考文献を挙げておく。

- ・ 線形回帰または傾向スコアの照合を用いる方式。
- ・ 実験群との比較可能性の確率が高くなるような方法で選んだ対照群。<sup>5</sup>
- ・ 実験群と対照群の初期差を調整するために行う介入前のアウトカムの評価。

上記基準に基づき、準実験の参考文献を特定すれば、種々の方法論を用いた研究間の影響の評価の感度を求めるために、今後行うと思われる研究が容易になる。

## 適切な研究を特定するための文献検索方法

### 個人的接触

放課後プログラムは、始まって日が浅いが、21 世紀コミュニティ学習センター（CCLC）が助成を行うようになってから急速に発展している。ハーバード家庭研究プロジェクト（HFRP、2003 年）および Lauver（2002 年）は、厳密な評価を実施、あるいは実施中の現行プログラムについて広範な情報を持っている。ハーバード家庭研究プロジェクト（HFRP）は、先頃オンライン放課後評価（Out-Of-School-Time Evaluation）データベース（HFRP、2003 年）を発表した。データベースには、夏季プログラムと放課後プログラムからなる校外活動の研究 54 件のプログラム・評価記述書が含まれている。Lauver（2002 年）は、低所得世帯の子ども向け放課後プログラムを対象とする「質の高い」報告形式の影響評価を求めて、2001 年 1 月主要データベースを徹底的に検索した（ERIC [教育学関係文献情報データベース]、PsychINFO [米国心理学会のデータベース]、Sociology Abstracts [社会学論文抄録データベース]）。さらに Lauver はインターネットを検索し、研究を特定するために、直接プログラムの管理者、評価者、およびその他放課後プログラムの担当者に連絡をとった。HFRP も Lauver も同分野で利用可能な研究を

---

<sup>5</sup> 研究は、「対照群を評価対象範囲から選んだ場合のほうが国内のデータセットから選んだ場合より」バイアスが少なかったとしている。ただし、「対照群と実験群が局所的に同じであり、対照群が、異なる研究調査現場における同種または同一のプログラムの評価において選ばれた対照群であった場合」である。（Glazerman、Levy および Myers、2002 年、P47）。

特定するためにあらゆる手を尽くし、明確に定義された基準に基づいて研究を選んだ。こうして本レビューの対象として検討すべき研究が集まった。われわれはさらに放課後プログラムの専門家に連絡し、われわれの検索結果を最新の評価によって更新する際に協力を依頼する予定である。

### データベースの検索

検索するのは ERIC (教育学関係文献情報データベース)、Education Index(教育分野の研究文献索引データベース)、PsychINFO (米国心理学会のデータベース)、および Dissertation Abstracts(学位論文抄録データベース)である。キーワード検索は次の単語で行うことになると思われるが、指定の件名標目またはデータベースに合わせて調整してもよい。

- 放課後プログラム (after-school programs) <sup>6</sup>
- 放課後教育 (after-school education)
- 放課後センター (after-school centers)

および

- 評価 (evaluation)
- アウトカム (outcome)
- 影響 (impact)

Lauver (2002 年) のレビューと HFRP のデータベースに含まれる研究の多くは、ピアレビュー誌で発表されたものではなく、評価者や助成者が出版したものである。このため専門誌への発表に付随するバイアスは問題とはならないだろう。コード化により研究ソースと財源の特定を行い、潜在的バイアスという特殊な問題は分析の際に検討する。

HFRP と Lauver(2002 年) がすでに特定したほとんどの研究、および前述のレビューは、最初、プロジェクト報告として出されたが、ピアレビュー誌で発表されることはなかった。このことと、時間とリソース上の制約を考慮し、本レビューではジャーナル検索データベースは扱わない。ただし、主要データベースの検索終了日を記録し、今後のレビューで同日以降に専門誌で発表された研究すべてを収集し、データベース・カタログに保管できるようにする。この方法では、データベースでカタログが作成されていない専門誌は収集しないが、対象分野を把握しておくことにより、厳密な実験計画に基づく放

---

<sup>6</sup> 予備検索を行ったところ、「時間延長 (extended day)」をキーワードにすると、授業時間や年間授業日数の増加をテーマとする、ほとんど関連性がない研究が検索結果として得られた。

課後プログラムに検索範囲を限定した場合、そうしたプログラムを見逃さずに済むと思われる。

### これまでのレビューと文献のリスト

放課後プログラムについて、これまでに行われたレビューに含まれる研究は、すべて検討対象とするほか、検索プロセスを通じて特定した、その他の研究の文献リストを検索することにより、取り上げる研究を特定する。

### インターネット検索

google.com（グーグル）などの検索エンジンを用いた、キーワード検索も行う。この方法により、新しい重要な研究が専門誌で発表される前に、そのような研究を特定できると考えている。

### **研究の選択**

レビューアは、両名とも基準に合致した論文抄録のレビューを行ってから、入手すべき研究論文を明らかにし、上記基準を満たすと思われる研究論文のみ入手する。両レビューアの推奨する研究論文が一致しない場合は、問題となる論文をすべて入手するという、あらかじめ取り決めた基準により解決する。両レビューアはそれぞれ入手した研究論文を読んでから、どちらをレビュー対象とすべきか推奨する。レビューアの間で決定内容が一致しない場合は、レビューアのアドバイザーである Rebecca Maynard 博士が問題の解決にあたる。

### **研究のコーディング法**

コード化の際の検討点を補遺 B に示す。両レビューアは 5 件の研究からなる第 1 群のコード化を行う。次にコード化担当者間の信頼性を確立するために、コード化上の不一致を解決する。時間とリソース面での制約により、残りの研究の二重コーディングが行えない場合は、1 名が残りの研究すべてのコード化を単独で行い、もう 1 名がこうした研究の無作為標本のコード化を行う。アウトカムの評価等の重要項目については、残りの研究すべてについて、二重コーディングを行う。両レビューアともレビューに不可欠な重要データが欠落している研究論文があれば、その著者に問い合わせるようにする。

## データの分析

本レビューに含める放課後プログラムと評価計画の種類については、拡散的な分類を行ったため、メタ分析の対象となる研究の数は、それほど多くなると考えている。放課後プログラムの計画・実施方法は、まったく同じように見えるかもしれないが、「りんごとオレンジ」の比較のように、異なるアウトカムが異なる方法で評価されている可能性がある（これまでのレビュー結果を前提にした場合）。このため個々の研究の「効果」を評価した後に、手元にある研究すべての「総合的效果」を評価するのは賢明ではない。そこで放課後プログラムがある地域で他の地域より優れた効果を上げているか否かを判断しようとする際には、代わりに GPA（学業評定平均値）、数学の成績、読解力の成績、出席率、行動などの重要なアウトカム項目について、個別にメタ分析を行う。研究が重要なアウトカムを複数の方法で評価している場合には、評価結果の独立性を確保するために、重要アウトカム 1 項目の分析に割り当てるデータ・ポイントを研究 1 件あたり 1 個のみとする。<sup>7</sup>

放課後プログラムの中には時間をかけて複数の評価を行ったものもある。このような場合には、調査継続期間によってアウトカムを分析する（例えば、1 学期、2 学期、1 年、2 年が終了した時点のアウトカム）。十分な数の研究論文が入手できれば、効果量の経時的変化も調査する。

メタ分析を行う場合は、**Biostat**（バイオスタット）が開発したメタ分析統合ソフトウェアを用いる予定である。ただし、アウトプットの報告形式の違いや、メタ分析用サポートデータの欠損という問題に対処する必要が生じることも想定している。欠損データの補完の時期と方法については **Cooper** および **Hedges**（1994 年）の研究を指針とし、異なる報告形式の使用については **Lipsey** および **Wilson**（2001 年）の研究を参考にする。アウトカムの評価結果とその報告方法が明らかになるまで、メタ分析による評価で効果量を評価する際にどのような基準を用いるかを定めることはできない。われわれは、効果量を提示し、フォレストプロットによりプロットを行うとともに、平均値や比率を報告する予定である。レビューの利用者にとっては、平均値や比率の方が解りやすく適切と思われる。<sup>8</sup>

---

<sup>7</sup> たとえば、「研究 1」が主要科目の評定平均値を報告し、「研究 2」が GPA(学業評定平均値)を報告している場合、レビューアは「研究 1」の GPA(学業評定平均値)を補完し、これを分析の対象とする。

<sup>8</sup> 平均値や比率に基づきなされた評価や一部のテスト成績を比較するのは、適切でない場合があることに留意する。たとえば、地区 1 の「A」は地区 2 の「A」と同等ではない場合がある。レ

## 固定効果と変量効果

本レビューの実施に先立つ放課後プログラムの実験的評価で得られた知見から、特定する研究の数は限定される。従って、これらの研究は、本レビューが対象とする介入の特性と、同様の特性を備えた放課後プログラムについて行われる研究の母集団ではない。こうした理由により、メタ分析による評価にはすべて固定効果モデルを採用する (Cooper および Hedges, 1994 年)。

## 定性的研究の取扱い

本レビューの目的を達成するために、分析に用いるデータには定性的データを含める。システマティック・レビューには、周到な計画に基づく補完的プロセス、実施プロセス、および運用プロセスの調査研究から得られた関連影響に関する定性的な結果データが含まれる。観察とインタビューから得たデータにより、プログラムの目標、活動、授業日への関連付け、およびプログラム内容の詳細を報告する。定性的データを使って、プログラムの方式、リーダーシップ、および背景が、放課後プログラムのアウトカムにどのように影響する可能性があるかを報告する。

## V. 予定スケジュール

<u>作業項目</u>	<u>終了予定日</u>
文献の検索	2002 年 11 月
研究の特定と研究論文の入手	2002 年 12 月
研究コードの試験的検査	2003 年 1 月
研究論文からのデータ抽出および二重コーディング	2003 年 2 月
分析	2003 年 4 月－5 月
報告書の作成	2003 年 5 月－11 月
最終報告書の提出	2004 年 5 月
レビューへのコメントに対する回答	受け取り次第

## VI. レビューの更新計画

---

ビューアはレビューの読者に対して、異なるレビューのアウトカムを比較することの妥当性への注意を喚起し、こうした場合は効果量を標準的評価基準とすることに利点があると思われることを伝えておく。

著者は、マスマティカ・ポリシー・リサーチ社 (Mathematica Policy Research) が現在実施している、21 世紀コミュニティ学習センターによる全国評価の 2 年目の結果が公表され次第、レビューを更新する。以後は妥当と思われる場合に更新するが、すべて助成期間内とする。

## VII. 謝辞

本プロトコルは、Rebecca Maynard 博士の計画によるシステマティック・レビューおよびメタ分析の実施に際して、ペンシルベニア大学大学院教育学研究科セミナーのメンバーをはじめとする関係者多数の力添えにより、完成したものである。

## VIII. 利害の不一致に関する表明

Lauver は文献のシステマティック・レビューに含めるプログラム候補として、レビューを行う放課後プログラムの影響評価を行った (無作為割り当て)。Lauver 自身は本研究調査をレビューに加えるか否かの決定には加わらない。Zief および Maynard が本研究調査のコード化を行い、選定基準への適合性について決定を下す。

## 補遺 A

### 研究の質を評価するためのチェックリスト<sup>9</sup>

セクションおよび 題目	項目 番号	記述子	報告 あり	コメント
<b>序文</b>				
タイトルと抄録	1	プログラム参加者を介入の対象とした方法 （「無作為割り当て」「ランダム化」、「無作為 に割り当てた」等）		
背景情報	2	評価者と介入との関わり		
	3	研究助成者と介入との関わり		
	4	介入の論理的根拠に関する記述		
目的	5	具体的目標／目的および前提条件		
	6	論理モデルまたは変化の理論		
<b>方法</b>				
参加者	7	参加者の参加基準（目標母集団等）		
	8	募集手順の説明		
介入	9	対象となる介入に関する詳細の明示		
	10	介入の実施に関する詳細の明示		
	11	対照群の活動に関する情報		
	12	潜在的な悪影響に関する情報		
アウトカム	13	明確な定義に基づく一次アウトカムと二次ア ウトカムの評価基準		
	14	介入の目標を踏まえたアウトカムの評価		
	15	評価ツールの説明およびその妥当性と信頼性 に関する情報		
	16	データ品質を向上させるために用いた方法 （補足的な研究、複数の評価、データ収集者 の教育）		
標本サイズ	17	実験群と対照群の標本サイズ		
	18	パワー分析による標本サイズの決定		
ランダム化の手順	19	制限項目を含むランダム割付手順の作成に用 いる方法の説明（ブロック化、層別化など）		
統計的手法	20	一次アウトカム（1個または2個以上）およ びサブグループの分析などの追加分析につい て、複数のグループを比較する場合に用いる 統計手法		
	21	採用した手法の妥当性		

<sup>9</sup> このチェックリストはドラフトであり、本レビュー・プロセスの間に更新される可能性がある。

	22	ベースラインで収集し分析に統合したアウトカムとその他の重要変数を対象とする介入前の評価		
<b>結果</b>				
参加者の減少	23	各グループから離脱した参加者の数		
	24	各グループで追跡調査ができなくなった参加者の数		
	25	分析から除外した参加者の数(理由を挙げる)		
	26	減少率>20%：プログラム完了者と非完了者の統計的比較		
	27	減少率>20%：ミッションデータ処理		
ITT (包括) 解析	28	「ITT 解析」で分析を行ったか		
アウトカムとデータ報告	29	アウトカム項目ごとのグループ別結果の概括		
		平均値と標準偏差の報告		
		P 値と自由度の報告		
		効果量の報告		
		その他の値の報告 (具体的に挙げる)		
<b>結論</b>				
解釈	30	研究仮説を勘案した結果の解釈、および潜在的バイアスまたは不正確性の原因		
	31	観察データを通じての影響結果の把握		
外的妥当性	32	結果の一般化可能性		
	33	介入の反復可能性		
総合的エビデンス	34	現在のエビデンスに照らした結果の総合的解釈		

What Works Clearinghouse の研究計画・実施ツール Study Design and Implementation Device (バージョン 0.6)、およびアメリカ医学会誌 (JAMA) の論文原稿基準と関連情報 Manuscript Criteria and Information、ランダム化比較試験報告書 Reports of Randomized Controlled Trials ([http://jama.ama-assn.org/ifora\\_current.dtl#SEC13](http://jama.ama-assn.org/ifora_current.dtl#SEC13)) のデータを編集したもの。