

20060300/A

20060300/B

厚生労働科学研究費補助金

社会保障国際協力推進研究事業

多国間協力事業の進捗管理および評価 手法のあり方に関する研究

平成 18 年度 研究総括・分担研究報告書

平成 16—18 年度 研究総合研究報告書

主任研究者 佐藤 敏彦

平成 19 (2007) 年 3 月

平成 18 年度厚生労働科学研究費補助金

社会保障国際協力推進研究事業

多国間協力事業の進捗管理および評価手法の
あり方に関する研究

総括・分担研究報告書

目 次

I. 平成 18 年度総括研究報告

「多国間協力事業の進捗管理および評価手法のあり方に関する研究」

佐藤 敏彦

II. 分担研究報告

1. 多国間協力事業の進捗管理および評価手法の在り方に関する研究の概要
2. 国際開発機関の政策評価における重み付けについて
3. 総合評価における指標の重みづけの決定手法に関する研究
4. 政策評価システムと資源配分の意味決定
5. 日本における多国間協力事業の政策評価システム

III. 附録

主要援助機関の政策評価の概観

平成18年度厚生労働科学研究費補助金（社会保障国際協力推進研究事業）
総括研究報告書

多国間協力事業の進捗管理および評価手法のあり方に関する研究

主任研究者 佐藤 敏彦 北里大学医学部 衛生学公衆衛生学

研究要旨

本研究は、複数の拠出者（利害関係者）が関わる、多国間協力事業の進捗管理と評価において、わが国の視点と、その他の拠出者からの視点の相違を明らかにした上で、我が国の多国間協力事業を通じた国際協力の評価方法を確立することを主な目的としている。これにより、多国間協力事業の効果的な実施に関してわが国の意思を明確に提示し、国際社会におけるわが国とわが国民の利益の増進を図ることが期待される。昨年度までに、現在の進捗管理・評価手法の潮流を分析するために、既存の援助機関の進捗管理・評価指標のレビューを実施した上で、実際に進められている多国間協力事業の進捗管理・評価手法に関して分析を行った。次に、上記のレビューを踏まえて多国間協力事業に相応しい評価システム・評価指標に関して分析し、それを参考にしたうえで評価・進捗管理の枠組みを形成した。

今年度は主として、ここまでで提示された評価手法を元に、諸機関の評価システムの利点を集約した評価システムを構築し、その評価手法と日本の国益という観点も加えながら分析することで、日本独自の評価システムのより具体的に政策提言の形にまとめた。しかし、本研究においては、運営に向けての具体的な手法構築にまではいたらなかった。残された課題としては、指標にどのように重み付けを行い、総合化するかという問題がある。この問題に取り組むためには、具体的なプロジェクトや実際に実施されたプロジェクト評価などの情報を収集し、専門家からの意見収集などを通して、根拠をもとに決定していく必要がある。

<分担研究者>

井手 義雄 医療法人雪ノ聖母会

聖マリア病院

長谷川敏彦 日本医科大学

池田 俊也 慶応義塾大学医学部

長谷川友紀 東邦大学医学部

高橋 謙 産業医科大学

<研究協力者>

小塩 篤史 東京大学大学院

赤沼 美里

A 研究目的

本研究は、複数の拠出者（利害関係者）が関わる、多国間協力事業の進捗管理と評価において、わが国の視点と、その他の拠出者からの視点の相違を明らかにした上で、我が国の多国間協力事業を通じた国際協力の評価方法を確立することを主な目的としている。これにより、多国間協力事業の効果的な実施に関してわが国の意思を明確に提示し、国際社会におけるわが国とわが国民の利益の増進を図ることが期待される。本年度は、既存の援助機関の援助評価手法について関係機関の報告書等の資料分析を中心に調査を行った。また評価の上で常用な概念である重み付けについて研究し、策定手法を検討した。最後に、各機関の政策評価の概観をふまえて、日本における政策評価のあり方を検討した。

B 研究方法

調査対象機関として、多国間協力事業に限定せず、世界中の主要な援助機関を扱った。インターネット等を通じて、各機関における評価に関する情報を収集し、各機関の評価の枠組みや評価における枠組みを検証した。また、文献調査を通して、総合評価を実施するうえで重要な手法に関して研究を行った。附録として、主要援助機関の政策評価の概要をまとめ、今後の研究の基礎付けを行った。

C 結果

1. 各機関の重み付け

政策評価の重み付けに関しては、議論は確定しておらず、各機関は比較的自由に重み付けを決定している。各機関の重み付けを分類した結果、7つの領域に分類が可能であった。現状は多くの援助機関は、重み付けに関して具体的に検証していないが、重み付けは評価の結果を大きく変える可能性があることが分かった。何らかの方法論を用いて重み付けを決定する必要性を強調した。

2. 重み付けの決定手法

重み付けの決定手法としては、意見集約型の手法と統計解析型の手法があることが分かった。意見集約型としては、AHP、デルファイ法、予算配分方式などがあり、統計解析型としては主成分分析、DEA などがあった。統計解析型も指標間の構造を把握す

る上で重要な情報を提示するが、現状ではデータが少ないために、実施することが困難であることが分かった。そこで、意見集約型で重み付けを決定し、その後に統計解析型を組み合わせる方法が妥当だと考えられた。

D. 考察と今後の進め方

各機関のモニタリング・評価ツール、評価方法についてレビューした結果、「効率性、有効性、インパクト、妥当性、自立発展性」の指標に加えて、「公平性、制度発展・オー

ナーシップ、付加価値、一致性、連携、人間の安全保障（緊急性・リスクへの対応）、環境」などを政策基準とすることが可能ではないかと思われる。評価のフレームワークとしては、国際的に主流になっている「投入・産出・結果・インパクト」を用いることが妥当ではないかと思われた。重み付けに関しては、AHP という意思決定手法を用いて、決定し、ODA 大綱に定める目標に沿った政策評価システムを構築するための基礎的な提言を行った。

国際協力機関の政策評価における重み付けについて

1. はじめに

国際協力における政策評価の指標に関しては、DAC 5 基準を中心にして、合意が進みつつある。それを受けて各機関は、政策評価のプロセスの中に、妥当性、有効性、効率性、インパクト、持続可能性といった基準を加えているが、実際の運用に当たっては大きな違いがある。特に大きな違いとなっているのが、総合評価における重み付けの問題である。総合評価は、複数の基準で判断された政策・プロジェクトを総合的に判断し、評価することであるが、その際には各基準をどの程度重視するかを決定する重み付けを行う必要がある。重み付けの方法によって、評価の順位が大きく入れ替わることもあり、各機関は非常に慎重である。しかし、政策評価を資源配分につなげるためには、総合評価を行い、各政策・プロジェクトを全体の中で位置づけることは有益である。本章では、国際機関・二国間援助機関が実施している政策評価の重み付けを概観し、分類を試みる。

2. 各機関の重み付け

JICA

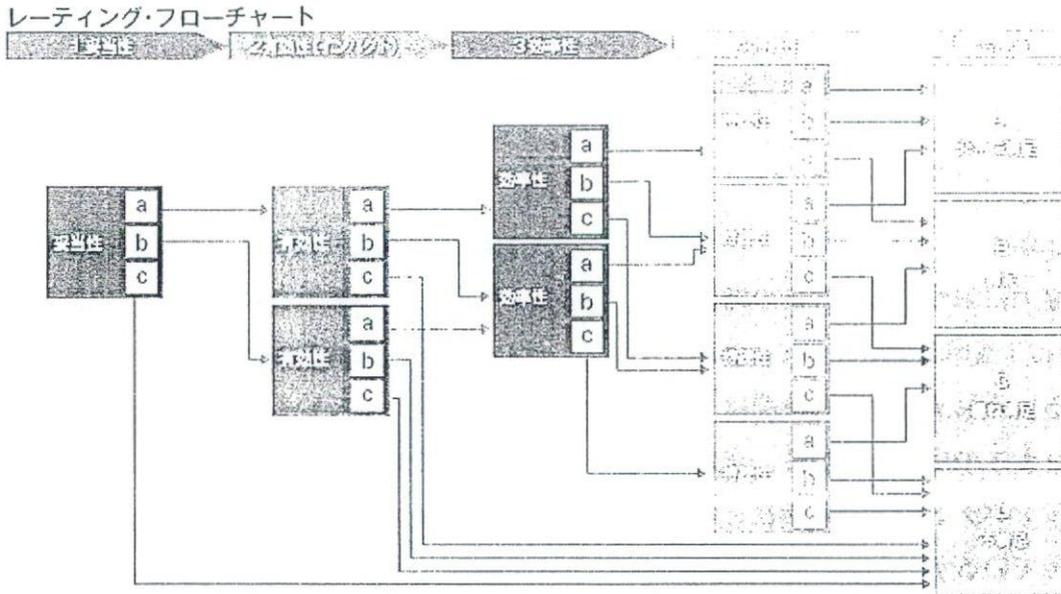
JICA（国際協力機構）は、DAC 5 基準を政策評価基準として採用している。JICA におい

ては明確な重み付けの決定はなされていない。基準として採用されている 5 項目それぞれに関して評価を実施し、個々の基準の向上を計るとしている（JICA 2004）。総合的に評価する際は、すべての項目が一定のレベルに達しているかどうかを判断基準としており、特定の重み付けを用いて評価を統合し、総合的に判断するということは行っていない。これは言い換えると、最低限度のラインを付けられた均等の重み付けであるといえる。評価項目で不十分な結果が出た場合、その結果がフィードバックされ、プロジェクトの改善に用いられる。事前評価も含めた政策評価は、プロジェクトの改善に焦点が置かれており、資源配分の意思決定に利用する意図をそれほど持っていないことが理由であると考えられる。

JBIC

JBIC（国際協力銀行）は、プロジェクト毎に DAC 5 基準に基づいた業績評価を実施している。JBIC では DAC 5 基準を若干修正し、インパクトと有効性をひとつの指標として判断している。図 1 は JBIC における総合評価の手法をまとめたものである。JBIC においては総合判断の際に、図のように階層的に判断している。重要度は左にある指

図 1 JBIC における総合評価の概要



出所：JBIC 2003

標ほど高くなっている。しかしこの場合、総合判断への影響力である重み付けは、評価の採点の数によって異なっている。例えば、aが3つ、bが3つの場合は、いずれの場合においても総合評価はAになる。これは重み付けが同じであるということと同じ意味である。しかし、aが2つ、bが2つの場合、(a, a, b, b)=B, (a, b, b, a)=B, (a, b, a, b)=B, (b, a, a, b)=B, (b, a, b, a)=C, (b, b, a, a)=D となり、妥当性が有効性よりも重要であり、有効性は効率性と持続可能性よりも重要であることがわかる。また、妥当性と有効性に関しては、cが付くとその時点でDとなることから、重視されていることが分かる。フローチャートが公開されてい

るため、明確ではあるが、場合によって重みが変わっている部分があり、変則的な重み付けの方法を採用していると言える。

KfW

ドイツの国際協力銀行である KfW は、評価の基準として有効性、効率性、妥当性を採用しており、それらを重み付けにより統合し、総合採点を行っている。KfW は、有効性、効率性、妥当性の3基準に対して、1から6までの6段階で採点を行っており、1が最高、6が最低得点になっている。大きく分けると、1～3が成功 (Successful)、4～6は不成功 (Unsuccessful) とみなされる。KfW は重み付けをして総合評価を行うことを規定として定めているが、固定さ

表 1 EBRD の総合採点の基準

Table 1 GUIDELINES FOR ASSIGNING THE OVERALL PERFORMANCE RATING FOR COMBINATIONS OF RATINGS ON FOUR MAJOR PERFORMANCE INDICATORS				
OVERALL PERFORMANCE RATING	Transition Impact rating	Project/Company Financial Performance rating	Fulfilment of Project Objectives	Environmental Performance
HIGHLY SUCCESSFUL	Excellent	Excellent	Excellent	Excellent
	Excellent	Good	Excellent	Good
	Excellent	Excellent	Good	Good
SUCCESSFUL	Good	Excellent	Excellent	Excellent
	Excellent	Good	Good	Good
	Excellent	Marginal	Satisfactory	Good
	Good	Good	Excellent	Good
	Good	Good	Good	Good
	Good	Satisfactory	Good	Good
	Good	Good	Satisfactory	Good
	Good	Good	Satisfactory	Satisfactory
PARTILY SUCCESSFUL	Good	Marginal	Excellent	Good
	Satisfactory	Satisfactory	Good	Good
	Satisfactory	Satisfactory	Satisfactory	Excellent
	Good/Excellent	Marginal	Marginal	Satisfactory
	Good	Marginal	Good	Satisfactory
	Satisfactory	Satisfactory	Satisfactory	Satisfactory
UNSUCCESSFUL	Satisfactory	Marginal	Satisfactory	Satisfactory
	Marginal	Good	Good	Satisfactory
	Good	Good	Good	Marginal
	Marginal	Marginal	Good	Marginal
	Marginal	Marginal	Marginal	Marginal
	Unsatisfactory	All	All	All
	Negative	All	All	All

出所：EBRD 2002

「負の影響 (Negative)」の 6 段階で採点を行っている。各基準に関して、採点を行った後、図の基準に基づいて総合的な判定を行っている。総合的な判定は「非常に成功 (Highly Successful)」、「成功 (Successful)」、「部分的成功 (Partly Successful)」、「不成功 (Unsuccessful)」の 4 段階で行われている (EBRD)。この表は、すべての組み合わせを網羅していないが、EBRD のガイドラインによると評価者はこの表にしたがって総合評価を実施している。こういった図示が必要であるということは、総合評価にあたって単純な平均値ではなく、何らかの重み付けが行われ

ていることを示している。図から「変化へのインパクト」>「プロジェクト・企業のパフォーマンス」=「目的の達成」>「環境面でのパフォーマンス」という順番で重み付けが決定されていると考えられる。

3. 重み付けのための手法

重み付けの手法は、

- 1) 階層的か
 - 2) 最低ラインが設定されているか
 - 3) 重み付けは均等か
 - 4) 重み付けは方法論に従って定められているか
- というコンセプトで分類が可能である。

れた重み付けは定めていない。これは、固定した重み付けでは不公正な事例が生まれるという考え方に基づいている (KfW 2006)。状況によって、重視すべき指標を変えることで、評価が不公正にならないようにしている。唯一の基準は、4より下の採点がひとつでもあった場合全体も不成功とみなされるという点である。(しかし、実際のレポートを見ると、4が付いているにもかかわらず、3が付いている場合あった。)

ADB

ADB (アジア開発銀行) は、政策評価基準として、妥当性、有効性、効率性、持続可能性、制度発展の5指標を用いている。ADB(2000)によると重み付けは妥当性20%、有効性25%、効率性20%、持続可能性20%、制度発展15%となっている (ADB 2000・2002)。公共部門に関する評価では、妥当性、有効性、効率性、持続可能性の4指標で評価され、それぞれ重み付けは、20%、30%、30%、20%となっている (ADB 2006)。これらの重み付けは、ADBの評価ユニットが主観的に定めたもので、このような重み付けがなされている理由に関しては説明が行われていない。有効性を重視するのは、OECDや世界銀行の考え方と同じ方向性のものである。

WHO

WHO (世界保健機構) は、2000年度の「World Health Report」で、各国の医療制度の評価基準として、健康結果 (Health

Outcome)、応需性 (Responsiveness)、公正財源 (Fair financing) の3つの指標と健康結果・応需性の分布を用いることを提案している (WHO 2000)。これまで制度の評価は有効性 (Effectiveness)・効率性 (Efficiency)・公平性 (Equity) の3つのEで評価されてきた。しかし、これらの概念は曖昧であり、再整理するために新しい基準が設定された。各指標の絶対値が有効性を表し、分布が公平性を表す。そして、資源の投入量に対する絶対値の割合が公平性を表している (WHO 2000、長谷川 2004)。指標の重み付けは、以下のようになっている。

$$\begin{aligned} (\text{医療制度指標}) &= 0.25 (\text{健康結果}) + 0.25 \\ & (\text{健康結果の分布}) + 0.15 (\text{応需性}) + 0.15 \\ & (\text{応需性の分布}) + 0.25 (\text{公正財源}) \end{aligned}$$

この重み付けは、指標の重要性に関して各国で実施されたアンケートによって決定されている。実際のアンケートから作成された重み付けは、0.24、0.25、0.13、0.16、0.22であったが、定義を簡単なものにするために、上記のような値に変更された。

EBRD

EBRD (欧州復興開発銀行) は、政策評価基準として「変化へのインパクト」、「プロジェクト・企業のパフォーマンス」、「目的の達成」、「環境面でのパフォーマンス」の4つの基準を採用している。これらの基準に関しては、「優秀 (Excellent)」、「良好 (Good)」、「十分 (Satisfactory)」、「最低限 (Marginal)」、「不成功 (Unsatisfactory)」、

表2 重み付けの分類

	階層的か	最低ライン	均等か	方法論的か
①階層的重み付け	YES	NO	NO	NO
②均等重み付け	NO	NO	YES	NO
③最低ライン付均等重み付け	NO	YES	YES	NO
④主観的重み付け	NO	NO	NO	NO
⑤最低ライン付主観的重み付け	NO	YES	NO	NO
⑥方法論的重み付け	NO	NO	NO	YES
⑦最低ライン付方法論的重み付け	NO	YES	NO	YES

出所：筆者作成

階層的であるというのは、図1のJBICの総合評価に代表されるような指標を一定の順番で判断していく重み付けである。最低ラインの設定というのは、ある指標が一定程度を下回っている場合、他の指標が高得点でも不成功であるとする評価体系のことである。重み付けが均等であるというのは、すべての指標の重み付けが同じであることである。重み付けが方法論に従って定められているというのは、重み付けの決定を統計的な解析やアンケートなどの具体的な方法に基づいて決定していることを示している。これらのコンセプトに対する答えの違いにより具体的には、以下の7種類に重み付けを分類することが可能である。

- ① 階層的重み付け
- ② 均等重み付け
- ③ 最低ライン付均等重み付け
- ④ 主観的重み付け
- ⑤ 最低ライン付主観的重み付け
- ⑥ 方法論的重み付け
- ⑦ 最低ライン付方法論的重み付け

この関係をまとめたものが、表2である。

階層的重み付けは、最低ライン付主観的重み付けに類似しているが、最低ラインの設定が指標ごとに異なっているなどの特性があるため、独自の分類とした。

表3 各重み付けの例

	具体例
①階層的重み付け	JBIC
②均等重み付け	UNDP
③最低ライン付均等重み付け	JICA
④主観的重み付け	EBRD,ADB
⑤最低ライン付主観的重み付け	KfW
⑥方法論的重み付け	WHO
⑦最低ライン付方法論的重み付け	

出所：筆者作成

4. まとめ

表3は、各分類の重み付けを用いている機関の例である。最低ライン付の重み付けを含めて援助機関で最も頻繁に用いられているのは、主観的に重み付けを決定する方法である。KfWはこの重み付けを採用している理由として、重み付けを一義的に定義することは難しく、ケースごとに調整する必要があるため、この重み付けを採用しているとしている(KfW 2006)。この指摘は非

常に重要である。しかし、一方で何らかの方法論を用いることを試みることも重要であると思われる。方法論的な分析によって正しい重み付けが明らかになるわけではないが、重み付けをどのようにしていくかという議論を進めるためには、こういったプロセスも必要である。実際、World Health Report において、指標の重み付けが発表された後、方法論や重み付けの値に関して活発な議論がなされた（詳細は、Murray 2002などを参照のこと）。現状は多くの援助機関は、重み付けに関して具体的に検証していない。しかし、最終的に総合的な評価を行っている以上、指標に何らかの重み付けをして、総合評価を行っていることになるのである。そして、それは多くの場合において、無自覚な均等の重み付けであることが多い。そういった観点から考えれば、何らかの方法論を用いて重み付けを決定してみることが必要ではないかと思われる。

4. EBRD 2002 Evaluation Policy
5. JICA(2004)「プロジェクト評価の手引き 改訂版 JICA 事業評価ガイドライン」
6. JBIC (2003)「円借款事業評価研修テキスト」
7. KfW(2006) KfW 評価部ホームページ
http://www.kfw-entwicklungsbank.de/EN_Home/Ex-post_Evaluation_at_KfW/Methodology/The_Three_Key_Criteria.jsp
(アクセス 2007 年 3 月 12 日)
8. KfW (2003) FC Projects and Sustainability. Discussion Papers 33.
9. WHO 2000 World Health Report: WHO

参考文献

1. ADB 2001 Guidelines for the preparation of project performance audit report
2. ADB 2002 Annual Review of Evaluation Activities.
3. ADB 2006 Guidelines for Preparing Performance Evaluation Reports for Public Sector Operations.

総合評価における指標の重み付けの決定手法に関する研究

小塩 篤史

1. はじめに

政策評価に際してどのような指標を用いるかということに関しては合意が進みつつあるが、その指標を用いてどのように総合的に評価を行うかということに関しては、研究の蓄積が未だ少ない状況である。本稿においては、総合的な評価の実施のために必要な指標の統合手法に関して、議論を行う。総合的な評価というのは、複数の評価基準や指標がある時に、各基準や指標を全て考慮に入れながら最終的な評価を行うことである。最終的な評価がどのような形式になるかは機関によって異なっており、総合採点（Overall Rating）を行って総合評価を一つの指標として採点する場合と質的に良好・不良と判断するだけのものもある。総合採点を実施することに関しては、利点と欠点が指摘されており、意思決定の道具としては賛否両論あるが、提示すること自身は各機関の方向性を示す上で有益な情報である。本稿では、この総合採点を実施するうえでどのような手法が存在するかを議論する。総合採点そのものに関する研

究は非常に少ないため、類似の手法である統合指標（Composite index）の考え方を援用している。

まず、議論の前提として、総合採点を実施する利点と欠点を考察する。次に、類似の手法である統合指標に注目し、既存の統合指標を概観する。そして、総合採点を実施する際に用いることの出来る手法を取り上げる。総合採点は現状では、主観的な価値判断に基づいた決定でなされている場合が多いが、本稿では指標の価値判断に関する議論は行わずに方法論に絞った議論を行う。最後に、手法の選択が総合採点の結果に及ぼす影響を具体例とあげながら考察し、手法決定の過程を分析する。

2. 総合採点・統合指標の利点と欠点

総合採点と統合指標は、共に複数の基準で判断されたものを一つの採点・指標で判断する手法である。統合指標は、新しい政策目標が出来た時に、それに相応しい指標を構築するため、複数の指標を統合する場合などにも使われるため、統

合指標の方が広い概念である。総合採点は、統合指標という概念の下位概念であると言える。そうであるなら、統合指標で議論される利点・欠点は、総合採点においても同様であると考えられる。本節では、統合指標の利点・欠点の議論を基礎として、総合採点の利点・欠点を議論し、注意すべき点を指摘したい。

Micklewright は、数多くの評価軸を含む複雑な事象を観察するときに、なんらかの総合採点や統合指標がない場合、結果としてどれか一つの評価軸に注目が集まってしまう、あるいはどの評価軸にも注目が集まらないという危険性があることを指摘している (Micklewright 1991)。これは、非常に重要な指摘であり、指標が複数ある場合、意思決定の情報として用いることは難しいものとなる。指標が複数ある状況で、政策の優先順位を決定することは非常に困難な作業であるが、総合的な採点を実施されていれば、複数の政策間の比較が容易であり、意思決定も容易になる。また総合採点に直接関係ないが、統合指標の利点としては以下のものがある。

第一は、データによって直接観測できない指標を数量化できる点である。例えば、人間開発や持続可能性というような新しい種類の指標を作成し、数量化する場合に、一つの指標で代替すると既存の政策目標との相違を明確にすることが出来ない。複数の指標を統合した指標で表現することで、本来の政策目標により近

い指標が作成可能になる。

第二は、複数の分野の政策目標を同時に考慮することが出来る点である。例えば、人間開発指標では、経済、教育、医療の3分野の政策目標を統合したことで、3分野の政策目標を同時に考慮することが可能になり、ランク付けも可能になる。そして、理論的には経済プロジェクトと教育プロジェクトの比較などのような分野を横断した政策比較が可能になる。

第三は、指標相互間の波及効果が統合指標の内部に組み込まれるため、適切に考慮できることである。良い波及効果のある政策は、統合指標においては、波及効果のない政策と比較して、大きな変化が見られるからである。医療サービスの場合では、小児保健への支援が小児保健分野の改善と同時に、基礎的なヘルスケアのインフラを整え、他分野に波及効果をもたらすことがある。統合指標によって評価すれば、波及効果の大きな政策を分析することも可能である。

この中で総合採点に関連してくるのは、一つ目の利点である。直接的に観察できない総合評価を可視化できる点が特徴である。実際、複数の指標が並んでいる場合、総合評価は各指標を均等に眺めることになる可能性が高い。例えば、指標1、指標2、指標3という3基準があり、重要度は1が最も高いとする。そして、5点を最高として採点を実施した時に、(3、5、5)という採点と(5、4、3)という採点を比較すると、点数上は前者の

方が良い評価をされているように見えてしまう。しかし、実際の評価は指標1をどの程度重視するか依存している。総合的な採点を行うことで、より直感的な新しい総合評価の指標が作成される。

しかし、総合採点・統合指標は、複数の指標を一つに統合し、情報量を減少させるので、限界点、注意点もある。まず、統合する際に方法論を誤ると、誤解や誤った政策的なメッセージにつながる点である (Saisana 2004)。また、総合採点・統合指標だけに注目が集まり、統合指標を構成している下位指標への注目が薄れ、その結果、総合採点・統合評価を高めるような、言い換えると最も重視されている指標を改善するような単純な政策に繋がる可能性が高まる。また、総合採点や統合指標を作成する際の手法の選択は、特定の価値判断に結びついてしまう場合がある。さらに、統合指標は、統合される下位指標の全ての値を必要とするので、採点により豊富なデータが必要となり、より時間がかかる。一つの下位指標のデータが欠損していてもそのケースに関しては統合指標を作成することが出来ない。この限界点は、総合採点・統合指標の欠点であるが、言い換えると使用する際の注意事項であるとも言える。これらに注意することで、より適切に総合採点を利用することが出来る。

3. 既存の統合指標のレビュー

既存の統合指標の代表的なものとして、UNDP の人間開発指標 (Human development index) や WHO の医療制度指標があげられる。本節では、これらの指標の概略を述べる。

・ 人間開発指標 (Human development index)

人間開発指標 (以下 HDI) は、開発政策の目標として GDP を代替する物を目指して作成された (UNDP 1999, Sen 2001)。HDI は、所得、教育、健康の3分野の指標を統合して作成される指標であり、GDP だけでは測定できない生活の質を捉えることを目的としており、開発政策の新しい目標として一定の支持を得ている。計算方法は

しかし、一方で厳しい批判にもさらされている。一つは、環境、政治、所得分配など生活レベルの測定に重要な変数が省かれていることへの批判である (Mahlberg 2001, Dasgupta 2001)。二つ目は、指標作成の方法論に関する批判である (Neumayer 2001)。HDI は後述するように各指標の単純平均で構成されており、その単純すぎる方法論が批判されている。

UNDP はこれらの批判に対しては、HDI は学術的、包括的な指標ではなく、明解さと生活の質の測定における GDP 以外の指標の強調が目的であるとしている (UNDP 2000)。

- ・ ヘルスシステム指標(Health system index)

ヘルスシステム指標は、各国のヘルスシステムを評価するために作成された指標である (WHO 2000)。ヘルスシステムを評価するために、5つの指標に関して調査をし、それを統合して咲く瀬際された。この指標の特徴は、絶対的な水準だけでなく、分配面にも気を配っていることである(WHO 2000)。

指標は以下の5つから成っている。

- ・ 健康状況
- ・ 良好な健康状態の分布
- ・ 感応性
- ・ 感応性の分布
- ・ 支払いの公正さ

統合指標作成の際の重み付けは、アンケート調査によって決定された (Murray 2002)。重み付けは、それぞれ 0.25、0.25、0.125、0.125、0.25 である。

上記2例以外の主要な統合指標としては、環境持続可能性指標 (イェール大学)、持続可能な発展指標 (UN) などがある。主に政策形成の分野で多くの統合指標が形成されている。しかし、有益であることは認識されているが、方法論に関する議論はそれほど進んでおらず、各機関は独自の判断で選択している (Nardo 2005)。次節では、方法論に関する概説を行う。

4. 指標統合のための手法

本節では総合評価を可能とするための統合指標の作成手法に関して概説する。統合指標の作成手法に関しては、Nardo や Saisana などが概観しているが、日本語で成されているものはほとんどない (Nardo 2005、Saisana 2005)。本節では、その作成手法に関して概観する。

統合指標は端的に定義すると複数の指標を一つに統合する手法である。しかし、その作成手法は様々である。指標を全部足しあわせる際に、各指標の得点にかけるものが重み付けであり、その重み付けをどのように定めるかということが重み付けの問題である。現在、統合指標の重み付けに用いられている主な手法は以下のものである。

- ・ 単純平均

単純平均は全ての変数の重みを等しいとする方法である。この方法の利点として、特定の指標を重視するという価値判断が不要である点や直感的に理解しやすい点を上げられる。しかし、重み付けが等しいので、どの指標を選択するか慎重に行う必要がある。指標の選択を誤ると、特定の分野が重視されたり、軽視されたりする偏りをもたらす可能性があり、望ましい方法ではない。人間開発指標は、下位指標の統合にこの手法を用いている。

・アンケートに基づいた選好による重み付け

どのように重み付けを行うのが良いかをアンケート調査などによって決定する方法である。各指標にどの程度の重み付けを付けるのかを直接質問し、その集計によって重み付けを決定する。人々の選好に基づいているため、感覚的に理解しやすい利点がある。ただし、誰の選好に基づくかという調査対象の選定の問題があり、専門家に限定するか、一般に開放するかで意味づけは大きく異なることになる。WHO が作成したヘルスシステム指標はこの方法で作成されている。

・予算配分方式 (Budget allocation)

この方式もアンケートによる決定に類似している。予算配分方式では、専門家に対して、各指標に予算を配分してもらうことによって決定する参加型決定手法である(Moldan and Billharz, 1997)。

この方式は、4つの段階を経て行われる。

- ① 専門家の選定
- ② 下位指標への予算の分配
- ③ 重みの計算
- ④ 必要な場合は全員の決定が収束するまで繰り返す。

この方法を用いた統合指標としては、以下のようなものがある。

- ✓ Employment Outlook (OECD, 1999)
- ✓ Composite Indicator on e-Business Readiness (EC-JRC, 2004b).
- ✓ National Health Care System

Performance (King's Fund., 2001)

- ✓ Eco-indicator 99 (Pré-Consultants NL, 2000)
- ✓ Overall Health System Attainment (WHO, 2000)

・主成分分析

統計的な解析手法である主成分分析を用いて統合することも可能である。主成分分析は、統計的な計算に基づき複数の指標の情報を最も表現している主成分を発見する手法である。主成分分析を使う最大の利点は、下位指標間に相関関係がある場合に適切に処理できることである。モデルに組み込まれた下位指標 A と B の間に高い共通性がある場合、両者に等しいウェイトを付けることは、特定の分野の指標を必要以上に重く計算してしまうことになる。主成分分析ではこれを回避することが可能である。主成分分析の欠点は、第一主成分が全ての変数に対して中立である可能性は低く、重要な意味を持つ変数の影響を落とす可能性があることやデータが少ないときに無意味な得点を計算してしまう可能性があるなどである。しかし、それでも全変数を表現するのに最適な重み付けを計算するので、統計的には、前2者よりも強い意味を持った指標が作成できる (Fay 2005)。また、主成分分析はパラメトリックな手法であり、分布が正規分布に近い場合は良好な推計結果が得られるが、分布が正規分布に近似していない場合は対数変換などに

よってデータを変換するか、あるいは用いない方が良い。また、サンプル数が少ない時は用いない方が良い。

この方法で統合された指標は、以下の通りである。

- ✓ Indicators of product market regulation (Nicoletti et al., OECD, 2000)
- ✓ Internal Market Index (EC-DG MARKT, 2001b)
- ✓ Business Climate Indicator (EC-DG ECFIN, 2000)
- ✓ General Indicator of S&T (NISTEP, 1995)
- ✓ Success of software process Improvement (Emam et al. 1998)

・DEA (Data Envelopment Analysis)

この手法は、オペレーションズリサーチで使われている手法である。最もパフォーマンスの良いケースを特定することで、最高パフォーマンスフロンティアを定義し、そのフロンティアからの距離によって、それぞれのケースの統合指標の数値を求めるものである（詳細に関しては、Charnes et al 1994などを参照のこと）。DEAの指標の計算方法に関しては、様々なモデルが存在するが、特徴としてはケースごとに重みを変更することである。そのケースにとって最も統合後の指標の値が大きくなる重み付けを設定することが出来る。この方法では、指標の改善に相対的に費用が必要な指標や困難な指標を特定でき、その指標が良いケース

に高い得点を付けられる利点がある。また、重み付けは、ケースにとって最適な重み付けであるので、評価の対象となっている集団からも重み付けに対する理解を得られやすいという利点もある。しかし、定義されたパフォーマンスフロンティアは現状の最善のケースに基づくもののため、特定の指標だけが大きなケースも評価されてしまう。また、外れ値の影響を大きく受けるという欠点もある。これらの欠点はあるが、精緻に定義づけを行えば、理論的に意味がある統合指標を形成でき、政策的含意も大きい。この方法を用いた統合指標としては、以下のようなものがある。

- ✓ Human Development Index (Mahlberg and Obersteiner, 2001)
- ✓ Sustainable development (Cherchye and Kuosmanen, 2002)
- ✓ Social Inclusion (Cherchye, Mosen, Van Puyenbroeck, 2004)
- ✓ Macro-economic performance evaluation (Melynand and Moesen, 1991, and Cherchye 2001)
- ✓ Unemployment (Storrie and Bjurek, 1999, and 2000)

・階層化意思決定手法 (AHP: Analytical Hierarchy Process)

階層化意思決定手法 (以下AHP) は、T.L.Sattyにより提唱された意思決定手法である。問題を分析する過程で、主観的な判断とシステムアプローチを組み合わ

せた問題解決型の意思決定手法の一つである。

解析の手順は、まず、問題の要素を[最終目標]—[評価基準]—[代替案]の関係でとらえて、意思決定に必要な要素に分割し、階層構造を作成する。次に、最終目標から見た評価基準の重要性を求める。この時、AHP に特徴的なのは、一対比較法という手法を用いてこの重要性を定める点である。各評価項目を一対一で比較し、その選好度合いを、1：同じくらい、3：少し選好する、5：かなり選好する、7：非常に選好する、9：圧倒的に選好する、の9段階で表現する。その一対比較行列を計算し、重み付けを計算する。そして、最終目標からみて評価基準の重要性を求め、各評価基準からみた代替案の重要度を評価する（中井 2005）。

留意点としては、評価基準の重要性を求める係数を設定する際に主観的な判断が伴うことである（経済的視点と環境的視点のそれぞれにあてはめる係数の値の大きさなど）。専門家と地元の住民など、立場が異なれば必ずと異なる値となるはずであり、結果として出てくる優先順位は違ってくることが予想される。従って、カウンターパートやステークホルダーとの協議が必要となる。

・デルファイ法

デルファイ法は、専門家の意見を集約して、優先度合いを決定する方法である。本手法は、絶滅の危機に瀕した種の保護

価値、競合する諸分野への限られた予算の振り分け、開発と保全の適切な事業調整といった評価に活用されている。

解析の手順は、まず各専門家が個別に評価対象の価値を設定する。各専門家の評価は、専門家を1箇所に集めて意見聴取を行うのではなく文書により収集する。次に、各専門家の設定した価値とその設定理由をメンバー全員に通知する。各専門家は与えられた情報に基づき最初につけた価値を再考し、新たな価値を設定する。さらに、全専門家の値が平均的になるまで、上記作業を繰り返す。この作業によって収斂した値を重み付けの値とみなす方法である。

留意点としては、参加する専門家の質や作業の進め方により作業結果が左右される点がある。

5. 手法の比較

これまでの議論において、全ての場合に通用する手法は存在していないことが分かる。手法はそれぞれ利点・欠点を持っており、適切に使用し、組み合わせることが重要である。本章では、各手法で作成された指標を比較しながら、手法の選択に関して議論を行う。

図1は、各手法で計算された重み付けを比較したものである。一番上の段は、8個の基準全てに均等な重み付けを行ったものである。二番目の段は、主成分分析

表 1 各手法で計算された重み付けの比較

Table 6.6. Weights for the sub-indicators obtained using 4 different methods: equal weighting (EW), factor analysis (FA), budget allocation (BAL), and analytic hierarchy process (AHP)

	Patents	Royalties	Internet	Tech exports	Telephones	Electricity	Schooling	University
EW	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
FA	0.17	0.15	0.11	0.06	0.08	0.13	0.13	0.17
BAL	0.11	0.11	0.11	0.18	0.10	0.06	0.15	0.18
AHP	0.09	0.10	0.07	0.21	0.05	0.06	0.18	0.25

出所：Saisana 2005

表 2 各手法で計算された重み付けによる順位の変動

Table 6.7. Countries rank according to five different weighting methods*

	EW	FA	BOD	BAL	AHP
Finland	1	1	1	1	1
United States	2	2	1	2	3
Sweden	3	3	1	3	4
Japan	4	4	4	5	5
Korea, Rep. of	5	5	6	4	2
Netherlands	6	6	7	8	11

出所：Saisana 2005

(ここでは因子分析となっているが、統計的には主成分分析と因子分析は類似の手法であるので、ここでは主成分分析とみなす。)であり、三番目は予算配分方式、四番目はAHPを用いたものである。一見したところ、値は非常に異なっているが、予算配分方式とAHPは比較的類似している。これは両者が共に意見集約型の手法であるからであると考えられる。

また図2は、図1の重み付けに基づいて計算された統合指標の順位を表している。図1の重み付けに加えて、DEAの手

法を用いて計算されたBODという手法が入っている。

これによれば、均一の重み付けと主成分分析は非常に類似性が高い。BODは、1位の数が増えているだけで、順位としての変動は少ない。AHPは、一番大きく変化している。別の研究では、Mahlburgが均等の重み付けで算出されている人間開発指標をDEAによって計算しなおしているが、大きな順位の変動が見られた(Mahlburg 2005)。