ISSN 1346-6151

日本評価研究

Japanese Journal of Evaluation Studies

Vol. 14, No. 1, November 2014

研究論文

政権交代を超えた行政事業レビュー～改変過程と課題～

田中 弥生

蓋然論理とその評価方法論的含意

西村 君平

Article

Rethinking the Human Development Index
：A Path toward Comprehensive Development Evaluation

Aki Yonehara

第14回全国大会開催報告

春季第11回全国大会開催報告

第15回全国大会のご案内

委員会活動報告

日本評価学会
Japan Evaluation Society

『日本評価研究』編集委員会
Editorial Board

編集委員長
Editor-in-chief
山谷　清志（同志社大学）
Kiyoshi YAMAYA

副委員長
Vice-Editor-in-chief
西野　桂子（関西学院大学）
Keiko NISHINO

常任編集委員
Standing Editors
三好　皓一（立命館アジア太平洋大学）
Koichi MIYOSHI

牟田　博光（東京工業大学）
Hiromitsu MUTA

編集委員
Editors
青山　温子（名古屋大学）
Atsuko AOYAMA

岩渕　公二（NPO政策21）
Koji IWABUCHI

大島　巌（日本社会事業大学）
Iwao OSHIMA

岡本　義朗（新日本有限責任監査法人）
Yoshiaki OKAMOTO

小野　達也（鳥取大学）
Tatsuya ONO

窪田　好男（京都府立大学）
Yoshio KUBOTA

佐々木　亮（国際開発センター）
Ryo SASAKI

渋谷　和久（内閣府）
Kazuhisa SHIBUYA

田中　弥生（大学評価・学位授与機構）
Yayoi TANAKA

西出　順郎（岩手県立大学）
Junro NISHIDE

林　薫（文教大学）
Kaoru HAYASHI

松岡　俊二（早稲田大学）
Shunji MATSUOKA

源　由理子（明治大学）
Yuriko MINAMOTO

事務局
Office
〒140-0002　東京都品川区東品川4-12-6　日立ソリューションズタワーB 22F
一般財団法人国際開発センター内
特定非営利活動法人日本評価学会

TEL: 03-6718-5931, FAX: 03-6718-1651
E-mail: jes.info@evaluationjp.org

『日本評価研究』は、和文年2回、英文年1回を刊行します。

日 本 評 価 研 究

第14巻　第１号　2014年10月

目　次

研究論文

田中 弥生

政権交代を超えた行政事業レビュー～改変過程と課題～ ……1

西村 君平

蓋然論理とその評価方法論的含意 ……17

Article

Aki Yonehara

Rethinking the Human Development Index

: A Path toward Comprehensive Development Evaluation ……31

第14回全国大会開催報告

第14回全国大会プログラム（実績）……48

共通論題セッション報告 ……50

自由論題セッション報告 ……54

春季第11回全国大会開催報告

春季第11回全国大会プログラム（実績）……59

共通論題セッション報告 ……61

自由論題セッション報告 ……66

委員会活動

企画委員会 ……69

国際交流委員会 ……70

広報委員会 ……71

第15回全国大会のご案内 ……72

日本評価研究刊行規定 ……73
日本評価研究投稿規定 ……75
日本評価研究執筆要領 ……77
日本評価研究査読要領 ……80

Publication Policy of the Japanese Journal of Evaluation Studies ……82
Information for Contributors (For English Papers) ……84
Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers) ……86
Referee-Reading Guideline ……88

【研究論文】

政権交代を超えた行政事業レビュー
～改変過程と課題～

田中　弥生
大学評価・学位授与機構
ytanaka@niad.ac.jp

要　約

行政事業レビューとは、前民主党政権下で、事業仕分けの府省における内性化を目的に、行政事業を対象とした資金の使途と事業効果を点検し、その結果を予算に反映する制度である。2012年12月の政権交代を経て、事業仕分けは廃止されたが、行政事業レビューは歳出改革の視点から継続されることになった。行政改革推進会議では、その運営の在り方と改変について議論されたが、政策評価との連携の在り方も視野に入っている。本論は、行政事業レビューの特徴と課題を整理した上で政権交代に伴う改変とその背景を説明し、政策評価との関係や定量化などの課題について考察する。

キーワード

行政事業レビュー、予算の流れ、定量化、政策体系、政策評価

1．はじめに

行政事業レビューとは各府省が概算要求前の段階で、国が実施する事業について、外部識者を加えた省内監視チームのもと、公開の場で事業効果等を点検し、その結果を次年度予算に反映する制度で、前民主党政権の行政刷新会議が実施した「事業仕分け」を参考に施行された。政権交代後の自民党政権下において、事業仕分けは廃止されたが、行政事業レビューについては、いくつかの修正を加えた上で継承されている。

筆者は漠然と2つの疑問を抱いていた。

ひとつは劇場型という言葉に象徴されるように、論理的な根拠や検証が希薄なまま直観的に判断している印象があったことである。

もうひとつは、政策評価制度に関連した制度の乱立が懸念された。民主党政権下の行政刷新会議が打ち出した事業仕分けや行政事業レビューは政策評価制度とは別に実施されており、屋上屋を重ねているようにみえた。さらにいえば、政策評価制度とは別のところで、政策や事業の廃止や見直しが命じられれば、政策のPDCAサイクルが壊れるのではないかと懸念したのだ。

2012年12月の政権交代後、自民党政権は事業仕分けを廃止した。だが、行政事業レビューは無駄の削減や歳出改革の一環として引き継がれることになった。稲田朋美前行政改革担当大臣は「府省が事業を見直し透明性を高めることは必要」[1]と述

日本評価学会『日本評価研究』第14巻第1号、2014年、pp.1-15

べており、自民党政権は無駄撲滅の観点から行政事業レビューが有用であると考えていることがわかる。

筆者は、現自民党政権の行政改革推進会議において行政事業レビューの改善案を提案する機会を得た。予想外だったのは、この作業を通じて、政策評価制度との関係で課題がよく見えてくることだった。行政事業レビューの仕組みには粗削りな部分もあるが、政策評価制度において未だに実現できていない課題に真っ向から挑んでいる点もある。

本論は政権交代を超えて継続されることになった行政事業レビューの特徴と課題を整理するが、そこから見える政策評価制度との関係の問題についても言及する。そして、政権交代を経て改変された内容がこれらの課題にどこまで対応しているのか、また残された課題は何であるのかを論ずることを目的とする。なお、本論は、行政改革推進会議における議論をベースに作成されているが、政権交代に伴う政策改変過程の記録の一助になれば幸いである。

2. 行政事業レビューの定義と成立過程

(1) 行政事業レビューとは

行政事業とは、各府省が施行・管理する事業をさすが、それが所掌する政策および施策の下位に位置するものであり、政策や施策の実行手段と言うことができる。その数は、平成24年度現在で、全府省あわせて4,906である。

行政事業レビューとは、無駄の撲滅などの歳出改革を目的に、行政事業の予算の執行状況や事業の効果や効率性を外部者の眼を入れながら、公開のかたちで確認する行為をさす。そして、その結果は次年度概算要求に反映することが求められる。なお、レビューの主体は、対象事業を所掌する府省の副大臣、政務官、会計課長、政策評価担当の職員であるが、外部の眼として民間有識者数名が参加する。

(2) 行政事業レビューの成立過程

行政事業レビューは前民主党政権が事業仕分けの「内性化・定常化」を図ることを目的に、府省が所掌する行政事業を自らチェックし、全省庁が自らの全所掌事業について仕分けの手法を用いて点検することとされていた。

この取り組みは2010年（平成23年）6月7日の閣議決定において決定され、毎年実施してゆくこととされた。しかし、2011年の東日本大震災を踏まえ、同年4月7日の行政刷新会議決定により公開プロセスやレビュー・シートの公開は要しないこととされた。

そして、2012年（平成24年）3月9日の行政刷新会議の決定により、平成24年度より施行された。

(3) 政権交代を超えた行政事業レビュー

2012年12月の衆議院選挙を経て、政権は民主党から自民党に渡った。自民党政権は事業仕分けの廃止を決定したが、行政事業レビューとして実質的に引き継がれることになった。

稲田朋美前行政改革担当大臣は、「国の事業の意義や効果を不断に総点検し、歳出改革をゆるぎなく進めてゆくことは有用な課題である。このため、行政事業レビューの今後の実施方法等について、行政改革推進本部の下に開催される行政改革推進会議において、改善に向けた検証・検討を行うこととしたい（2013年年1月29日）」と述べている。

すなわち、行政事業レビューは歳出改革の手段として継続するが、そのあり方については行政改革推進会議のもとで改善の検討がなされることになったのである。

では、行政事業レビューに対するスタンスや政策的な位置づけについて、自民・民主の両政権の間でどのような違いがあるのだろうか。以下は、両政権の説明を記すが、まず、民主党政権下の行政刷新会議は次のように説明している。

行政事業レビュー（以下「レビュー」という）は、各府省自らが、自律的に、概算要求前の段階において、原則すべての事業について、予算が最終的にどこに渡り（支出先）、何に使われたか（使途）といった実態を把握し、これを国民に明らかにした上で、事業仕分けの手法も用いながら事業の内容や効果の点検を行い、その結果を予算の概算要求や執行等に反映させる取組である

（2012年3月9日）。

自民党政権下の行政改革推進会議は、次のように説明している。

行政事業レビューは、各府省自らが、外部性・公開性を確保しながら、予算の執行状況を点検し、その結果を事業見直しに反映させる取組であり、予算のPDCAサイクルの具体化を図るものである。また、国の全事業についてレビュー・シートが公表されており、事業の実態に加え、予算の要求段階における検討過程が国民に明らかになった。

政府は、行政事業レビューを毎年実施することにより、事業のより効果的かつ効率的な実施、国民への説明責任の確保、透明性の確保を図り、もって国民に信頼される質の高い行政の実現を図るべきである（2013年4月2日）。

両政権の説明を比較すると、自民党政権下では、「仕分け」という言葉が用いられていないことなどの表現の差はあるが、その基本スタンスには大きな差異がないようにみえる。

では、政権交代を経て何が変更され、その根拠とはどのようなものだったのか。そこで、前民主政権下の行政事業レビューの特徴と課題を整理した上で、政権交代前後の変更の内容を見ることにする。

3. 民主党政権の行政事業レビューの特徴と課題

民主党政権下での行政事業レビューの制度と手続きについて説明し、その特徴と主な課題について論ずる。

3.1 行政事業レビュー制度の分析

(1) 行政事業レビューの工程

行政事業レビューの一連の作業工程は以下のとおりである。

①予算管理・効率化チームの編成

実行部隊として、府省内に予算管理・効率化チームを結成する。前述のようにチームは、副大臣、政務官以下、管理職に民間識者数名で構成される。このチームのもとで、レビュー担当者、地方支部局等の連携・協力体制を構築する。

②行動計画の策定

次に行政刷新会議が定める期日までに行動計画を策定するが、その内容は、レビュー・シートの作成と公表、公開プロセスの実施、概算要求や予算執行への反映、パブリックコメント等で、ここで資料作成から、概算要求への反映の一連の工程が決定される。

レビューの対象は約5000の全事業であるが、公開プロセスの対象は選択的に行われる。すなわち、1億円以上の規模、長期的に実施され改善の余地が大きいもの、過去に会計検査院や総務省行政評価局あるいはマスコミから指摘されたことのある事業などが対象になるが、その選定は当該府省自身が行う。

③公開プロセス

公開プロセスとは、先に選択された事業について各府省と行政刷新会議が指名した外部識者で行う審議のことであるが、レビュー・シートに記された情報を中心に評価を行い、「廃止」「抜本的改善」「一部改善」「現状通り」のいずれかで判断する。

④チームによる点検と概算要求への反映

その他の事業については、予算管理・効率化チームの管理下のもと外部識者を加えてレビューを行う。判定は公開プロセスと同様の4段階で行う。チームは、公開プロセスおよびチームレビュー結果をとりまとめ、次年度概算要求へ反映する。

⑤点検結果の公表と行政刷新会議による点検

点検結果はレビュー・シートに反映し公表する。また、概算要求への反映結果についても公開する。概算要求のタイミングで、行政刷新会議は各府省の点検が十分なものであるか、適格に概算要求に反映されているか否か、府省間の重複がないかなどを点検する。

このように、各府省が自ら所掌する全事業を自己評価し、その結果を概算要求に反映する。それを管理者の視点から行政刷新会議が点検するという仕組みである。

では、どのような視点で評価されているのか。そこで、次項では評価項目が記されている行政事業レビュー・シートの構成をみる。

(2) 行政事業レビュー・シートの構成と主な課題

行政事業レビュー・シートは事業単位で作成されることになっており、事務的経費と人件費を除く経費と事業実施状況がレビュー・シートに記される。行政事業レビュー・シート（以下、レビュー・シート）は、次のような項目で構成されている。

①行政事業の基礎情報
事業名、会計区分、根拠法令、事業目的、事業概要、予算状況と執行状況（複数年の場合には各年について記す）
②業績指標
成果目標と成果実績（アウトカム指標とその達成値、達成度（%）で示す）
活動指標と実績（アウトプット指標とその活動実施状況、算出コスト、単位あたりのコスト、算出根拠）
③事業担当部局による点検内容
目的予算の実施状況、資金の流れ、活動成果について所定の基準・視点[2]に基づき判定
④予算監視・効率化チームの所見
⑤所見を踏まえた担当部局による改善点
⑥補足情報
⑦資金の流れについてのフローチャート
樹形図のかたちで可視化して示すことが求められる。
⑧フローチャートで示したブロックごとの費目・使途・金額の代表例
⑨ブロック毎の支出先上位10者リスト

このフォーマットは5頁構成であるが、そのうちの3頁が資金の流れと使途に費やされ、その記載の仕方もより詳細に記すことが求められていることから、資金の流れにより記述の重点が置かれていることがわかる。

3.2　行政事業レビュー制度の3つの特徴

行政事業レビュー制度として大きく3つの特徴を挙げることができる。

第1に「公的資金の流れ」を明らかにしたことである。行政事業レビューでは、そのフォーマットに、府省から拠出された補助金や委託金の受け手について末端まで樹形図（ツリー）のかたちで記すことが求められており、資金の流れが可視化されて示されることになる。これによって府省から支出された補助金や委託金が、どのようなルートを通って都道府県や企業、公益法人に渡っているのかを一目して捉えることができる。実際不明確な資金の流れなどが明らかになったケースもある[3]。

第2に行政事業レビュー結果を次年度の予算に反映することが閣議決定されていることである。すなわち、「廃止」と判断された行政事業は、原則として次年度の概算要求の対象にしないというルールが定められていた。また、レビュー結果が概算要求に適切に反映されていることを点検・確認するための会議が設けられ、財務省主計局と連携が組まれている[4]。

第3に行政府機関の業績判断に民間人の参加機会が公開のかたちで設けられたことである。前述のように、予算監視・効率化チームに民間有識者が複数名参加している[5]。こうした事柄をもって、「裁判員裁判制度」に例えられることもあるが、広く国民から無作為に選ばれる裁判員制度と、行政府機関が指名した民間識者をチームに加えることを同一に扱うことは難しいだろう。

では、行政事業レビューと政策評価制度の相違は、評価の対象が行政事業と政策であるという違いの他にどこにあるのか。筆者は、第1の資金の流れと、第2の次年度予算への反映の2点が大きな相違点であると考える。政策評価において、資金の流れをその末端まで明らかにすることは行われておらず、予算総額の執行状況を確認するにとどまっている。また、評価結果の反映については、総務省から各府省に「勧告」がなされることになっているが、勧告は、罰則規定などの法的拘束力をもつものではない[6]。また、行政事業レビューのようにその結果を次年度の概算要求に反映することを求めるようなルールはない。評価結果に実効性を持たせることは、政策評価のみならず、評価一般において常に問題になるが、行政事業レビューはこの点にチャレンジしていると言えよう。

3.3　行政事業レビュー・シートの課題

他方で、行政事業レビューには課題もある。ここではレビュー・シートから捉えることのできた

3つの課題について説明する。

第1の課題は、「双頭の目的」ともいうべき問題である。すなわち、レビュー・シートは資金の流れと使途について詳細に記すように構成されており記述量の大半を占めている。しかしながら、最終判断は、相対的に記述の薄い事業の効果や効率性[7]にかかる評価項目を以て判断するようになっていた[8]。資金の流れについて重きをおくのか、事業効果に重きをおいて判断するのかが曖昧で、その目的が二つに分断されているようにみえる。

第2の課題は、評価項目に関する問題で抽象度が高いものが散見され、何をもってその基準を満たしているのかの判断は裁量の余地が大きく、実質的に任意になっていた。例えば、「国が実施すべきか、民間や自治体が担うべきではないか」、「財政資金投入の必要性があるか」という評価項目があるが、その判断の仕方は評価側の任意となっているため、解釈が多義的になったり、主観に陥る可能性がある。また、抽象度の高い評価項目から、具体的なものまで、ややランダムに並んでおり、また評価項目間の重複もみられた。

ちなみに、評価において判断の要となる基準や評価項目の論理整合性に問題がある場合、評価作業全体の論理体系性に支障をきたす可能性がある。「劇場型」と批判された背景には、仕分け人による議論のスタイルのみならず、こうした問題があったのではないか。

第3は成果の定量化の問題である。行政事業レビュー・シートでは、アウトプットおよびアウトカムレベルの成果を定量的に示すことが求められている。また、同シートでは年度ごとの支出規模や単位あたりのコストを記す欄があるが、これは投じた費用に比し、得られた成果が適当なものであるかどうかを判断するための参照情報を導こうとしたものである。

アウトプットレベルの目標達成状況は、事業計画に基づき実施した内容を記せばよいので比較的記しやすい。しかしながら、アウトカムレベルの成果については、事業実施による対象者や社会への影響を測定し、定量的に示すことが求められるので、より専門的な技術が求められるだけでなく、作業量が増えることを意味する。また、投じた費用と成果の比較をより精度を上げて行おうとした場合、アウトカム成果の定量化作業をした上で、それを金銭換算して投じた費用と比較をするため、さらに高度な知識や技術が求められる。

平成24年度の行政事業レビュー・シートをみると、「単位あたりコストの削減に努めているか。その水準は妥当か。」「他の手段と比較して実効性の高い手段となっているか」の評価項目がある。本来なら、アウトカム成果を定量的に示し、費用との比較や効果の大きさを比較した上で判断されるべきものであろう。しかし、アウトカム成果については、記述の不充分なものが目立ち、費用との比較においては先のような分析というよりも感覚的な記述が散見された。成果の定量化について十分に対応できているとは言い難く、行政事業レビューの特徴である費用との比較においては、そうした試みをしようとする意図は伝わるものの、道半ばであるという印象は否めなかった。

だが、この問題は行政事業レビューのみで解決できるものではなく、行政事業のアウトカム成果の測定は、その上位にある施策や政策評価との関連で解決せねばならない問題でもある。

そこで、次項では政策体系図を用いて、政策評価との関係をみながら行政事業レビューの課題を分析する。

4. 政策体系図から捉えた行政事業レビューの特徴と課題

前述のように、行政事業レビューのアウトカム成果の定量化の課題は、その上位にある施策・政策評価の課題に通じる。そもそも、行政事業は府省の政策の構成要素であるので、政策評価の準備状況が行政事業レビューに直接的に影響してくる。例えば、行政事業レビュー・シート欄にアウトカム成果を指標によって記す欄があるが、少なくとも成果指標のひとつは、政策評価の事前分析表から転記することが可能なはずだ。さらに事前分析表には、政策の関連事業を記す欄もあるので、これを参照すれば、どの政策からアウトカム指標を転記すればよいかはわかるようになっている。しかしながら、行政事業レビュー・シートにおいて、アウトカム成果の記載がないものが3割以上

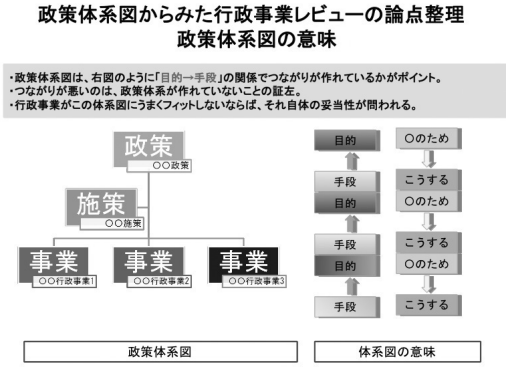
に達している（小野 2013）。この事実は、政策評価と行政事業レビューのいずれか、あるいは両者間において課題があることが示唆されている。

また、行政事業レビューという、より実施現場に近いところにある事業の評価から政策評価を捉え直すことで、政策評価のみでは明らかにならない問題を把握する機会にもなる。

そこで、政策体系図を用いて行政事業レビュー・シートの課題を整理、分析してみることにした。

(1) 政策体系図とその論理構造

図1　政策体系図



（出所）第2回行政改革推進会議筆者提出資料（2013.4.2）

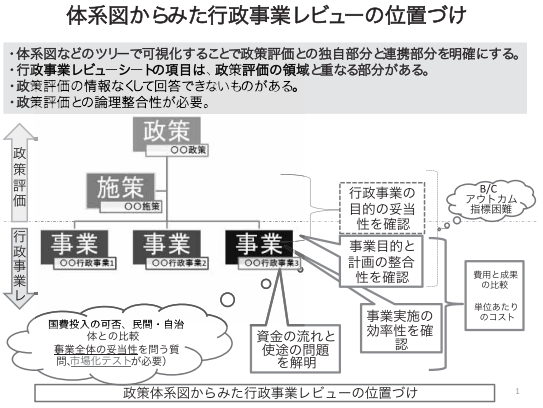
図1は政策体系図の論理構造を示したものである。体系図の構成要素は、いずれも「目的-手段」の関係で繋がっていることを示している。すなわち、政策目的を達成する手段が施策であり、施策目的を達成する手段は事業である。なお事業目的を達成する手段は計画である。政策を構成する施策、事業はこの論理構造に基づきひとつの体系を成さねばならない。

仮にある事業がこの論理体系にうまくフィットしないのであれば、事業の妥当性が問われることになる。

(2) 政策体系図から捉えた行政事業レビューの課題

この政策体系図を用いて行政事業レビュー・シートを見直すことによって得れらた特徴や課題を図示したのが図2である。

図2　政策体系図から捉えた行政事業レビューの特徴と課題



（出所）第2回行政改革推進会議　筆者提出資料（2013.4.2）

図2に基づいて、行政事業レビューの特徴や課題を以下のように大きく3つに分けて捉えてみる。すなわち、第1は行政事業レビュー作業の中で自己完結している特徴である（実線四角枠）。第2は政策評価との関係の中で解決されるものでなおかつ解決の道筋がある程度みえる課題である（点線四角枠）。第3は解決が困難あるいは解決の道筋がみえていない課題である（雲形枠）。以下、これらの3つについて詳述する。

第1の特徴では以下の4点を挙げることができる。第1点は資金フロー図を用いることによって、資金の流れと使途を明確にしようとしたことで一定程度それを可能にした。第2点は事業目的と計画の整合性に関するもので、事業目的と事業概要、実施方法を記させることでこの点を確認しやすくなった。第3点は事業実施の効率性について確認していることである。これは、事業実施状況や予算執行状況を確認することで可能になる。また、事業結果と予算額との比較によってアウトプット・ベースの成果と費用を比較しようとしている。また、興味深いのは第4点の特徴で、単位あたりのコストの算出を求めているという点である。具体的には予算を一日あたりに換算するもので、これによって類似事業に投じているコストの差異が明らかになるので比較可能性が高まった。

第2の課題、すなわち政策評価との関係の中で解決されるべきで、なおかつ解決の道筋がみえて

いる課題とは、事業の妥当性の判断にかかる問題である。行政事業レビュー・シートには、「広く国民のニーズがあり、優先度が高い事業であるか。」という評価項目がある。これは当該事業の妥当性を問うものである。妥当性の判断の仕方は先に掲げられた事業へのニーズの確認など複数あるが、少なくとも中央省庁のもとで実施する事業については、その上位にある政策的な根拠が明確でなければならない。つまり、図2に即して述べれば、事業の上位にある施策との関係において「目的-手段」の関係で論理的につながっていることを確認する必要がある。だが、行政事業レビュー・シートにおいてはこうした判断の根拠や考え方が示されていない。

第3の課題、すなわち解決の道筋がみえていない課題として、以下の2点が挙げられる。第1点はアウトカム成果の定量化の問題である。行政事業レビューは、アウトカム・ベースの成果とそれに投じた費用の関係を明らかにすることを求めているが、各府省が対応しきれていない。事業、施策、政策が体系を成しているのであれば、事業のアウトカム成果は上位の施策や政策の成果と共通するものがあるはずである。またそうでなければ政策の論理体系が成立しないことになる。

本来、事業の上位にある政策の目的と成果がアウトカム・ベースで定義され、それに基づいて検証がなされていれば、行政事業レビュー・シート上では、それを転記すればよいことになる。しかし、実際には政策効果が定量的に検証できているものは限定的である（小野 2013）[9]。また、そもそも政策と事業の関係が不明確でうまく紐つけができないために、政策評価で記載されている指標を行政事業レビュー・シートに転記できないケースもある[10]。

第2点は「国が実施すべき事業であるか。地方自治体、民間等に委ねるべき事業となっていないか。」という評価項目に関するもので、判断方法や判断根拠が明確に定められていないという問題がある。仮に、国と同一事業あるいは類似事業を自治体や企業が実施している場合には、その効果や費用を比較することができるので判断しやすいだろう。しかし、その場合であっても経済・社会環境などの外部条件が異なると、同じ企業や自治体であっても担えなくなることがある。さらに自治体や企業による実施例が存在しない場合もあるが、安易な憶測や直観的判断は避けるべきと考える。

この課題に対してひとつの考え方を提示しているのが2006年より施行された市場化テストであると考える。これは、ある行政事業を実施するにあたりそのコストや成果について鍵要因となる指標（Key Performance Indicator）を用いて定量的に説明した上で、最も効率的に担える主体を選ぶ仕組みである（田中 2008）。当該事業の実施主体として国、自治体、企業のどの主体が適しているのかを選択するのであれば、市場化テストで用いられているような手法を用いて科学的に検証した上で判断すべきであろう。

(3) 行政事業レビュー改善の提案

以上のような課題を踏まえ、筆者は行政改革推進会議（2013年4月2日）で、行政レビューについて当面の改善案と中長期の改善案に分けて次のような提案を行った。

①当面の改善案

「視点・基準の改善」

行政事業レビュー・シートにある視点・基準は現行をいかしながらもより論理明確に整理する。

「“資金の流れ”の強みと課題」

本項目は、行政事業レビューにおける最大の特長であることをより明確に説明する。他方、問題の所在を明らかにするためには技術や経験が求められることから、民間有識者や行政関係者がまずチェックし、その結果をもって民間有識者が判断する。

「情報整理の上での民間識者の参加」

効果の判断は専門的な技術・知識が不可欠で、それが欠如すると主観的な判断に陥る可能性があるため、より整理された情報や説明のもとに民間有識者が判断する。

「わかりやすい公開」

政策と事業の関係を示す体系図など、図や表を用いるなどの工夫を凝らし、レビュー結果をわかりやすく公開する。

②中長期の改善案

「定量化のための専門性の向上」

定量化（指標デザイン、データ測定・分析）には専門知識が求められるものであり、安易な定量化は業務上の負荷や無駄を招くことから、専門家を活用する。

「政策評価との関係の明確化」

行政事業の目的の妥当性は、本来、上位概念である施策・政策との整合性に基づいて判断されるものであるから、これらの体系性をどう描くかが課題になる。

そこで政策体系ツリー図を導入してはどうか。ツリーの効果としては、政策、施策、事業間の整合性の確認、事業の重複や欠落の発見、成果の定義の明確化への寄与などが挙げられる。

「作業効率化のための電子化」

行政事業レビュー・シートと政策評価には重複した項目があり、手動で転記している状況。1回記入すれば自動的に別箇所に転記されるように、電子化システムを導入する。それは技術的に可能であり、業務上の負担やミスを軽減し、作業効率化に寄与するはずである。

（以上、行政改革推進会議　2013年4月2日　筆者発表資料より）

上記は主な論点のみ列挙しているので、以下に若干の補足を加えたい。まず、提言文書にある「視点・基準の改善」とは、行政事業レビュー・シートでは、「国費投入の必要性」「事業の効率性」「事業の有効性」のもとに記された12の評価項目のことをさしている。よりわかりやすくするために本論では評価項目という名称に代えて説明している。

また、政策体系ツリーの導入を提言書において勧めているが、これを政策評価制度のもとで作成するのか、あるいは行政事業レビューから作成するのかという議論はあろう。しかしこの提案の本質はロジック・モデルの考え方に基づく政策体系ツリーを行政事業レビューにおいても活用すべきという点にある。これによって、事業とその根拠となる政策との関係やその整合性を確認することができるからで、先に挙げた「妥当性」の説明に寄与するものと思われる。また、各府省には類似の目的を有する政策が現存しているが、その達成手段としての事業の策定の方法や組合せについて比較しやすくなるし、事業レベルにまでブレークダウンされていれば整理・統合をより進めやすくなる。

逆に事業数や種類が不足しているために政策目標を達成し難い状況にある場合も、体系図を用いることで発見しやすくなると思われる。

5. 政権交代後の行政事業レビューの改変状況

行政改革推進会議事務局は、同会議の識者の意見等を踏まえ行政事業レビューの改変案を提示した。

では、過去の実施や先の提案を踏まえどのような改変がなされようとしたのか。ここでは、民主党政権下の行政事業レビューと政権交代後の自民党政権下のそれとを新旧対比をするかたちで改変点を説明する。

(1) 行政事業レビューの改変

政権交代前後の改変状況を表1にまとめた。

まず「レビューの目的」であるが、政権交代前は無駄の撲滅と歳出改革に重点が置かれていたが、自民党政権下では効率性の高い事業にむけて改善することも目的として掲げられた。

また「レビューの主体」は民主党政権下では、大臣官房、担当省の職員、民間有識者から構成される「予算監視・効率化チーム」によって行われていたが、自民党政権下では民間有識者は府省の職員から独立したかたちでチェックを行うことになった。その背景には、行政事業に関する予備知識などの情報量について職員と民間有識者の間に差異があるため、より効率的に議論を進めるためには両者を独立させて各々の視点から議論を進めたほうが得策であるという判断があったと思われる。

また「チェックの対象」が民主党政権時代よりも絞られた。5000程ある行政事業の資金の流れや

表１　政権交代前後の行政事業レビュー対比表

	自民党政権下	民主党政権下
レビューの目的	・無駄も撲滅の推進だけでなく、より効果の高い事業に改善するため ・歳出改革	・無駄の撲滅 ・歳出改革
レビューの主体	・各府省による自律的な取り組み ・行政事業レビュー推進チーム（大臣官房、各部局などの職員から構成） ・民間識者はチームから独立したかたちで点検を行う	・各府省による自律的な取り組み ・予算監視・効率化チーム（大臣官房、各部局などの職員、民間識者数名）
レビュー作業工程	・チーム編成、レビューシート作成と中間発表、チームによる見直し、民間識者によるチェック、チームによる判定、予算策定への反映の一連の工程を計画にする	・チーム編成、レビューシート作成と中間発表、チームによる見直しと判定、予算策定への反映の一連の工程を計画にする
レビューの対象	・全事業（約5000） ・基金も対象に ・民間識者はチームとは独立にレビュー初年度と最終年度にあたる事業（5年間で全事業が民間識者のチェックにかかるようにする）	全事業（約5000）
レビュー方法	・行政事業レビューシートの基準・視点を改変	・行政事業レビューシート
段階判定の内容	「事業全体の抜本的改革」「事業内容の改善」「現状通り」の3段階であるが、廃止は抜本的改革の中に含まれる。	「廃止」「抜本的改善」「一部改善」「現状通り」の4段階
レビュー結果反映	次年度予算の概算要求に反映	・次年度予算の概算要求に反映 ・廃止と判断された事業は、次年度予算に計上できない
公開レビューの対象	・公開の対象は、1億円以上の事業、会計検査院、総務省行政評価局、マスコミから指摘があった事業、長期的に取り組む事業で改善の余地が大きいもの	・公開の対象は、1億円以上の事業、会計検査院、総務省行政評価局、マスコミから指摘があった事業、長期的に取り組む事業で改善の余地が大きいもの
第三者チェック	行政改革推進会議が、概算要求のタイミングで、各府省が行政事業レビュー結果をもとに予算に反映されているかどうか、各府省の重複がないかなどを確認する。	行政刷新会議が、概算要求のタイミングで、各府省が行政事業レビュー結果をもとに予算に反映されているかどうか、各府省の重複がないかなどを確認する。

（出所）行政改革推進会議提出資料より筆者作成

効果のひとつひとつを民間有識者が直接点検することには限界がある。そこで職員が点検し整理したものをもって、適格に点検がなされているかを民間有識者が確認することになった。

「レビュー方法」は行政事業レビュー・シートに基づき実施されるが、前述のように同シートの評価項目に修正が加えられた。

また「段階判定」の内容も改変された。民主党政権下では「廃止」という判断があり、これが注目を集めた点でもあった。しかし、自民党政権下では、判定基準の「廃止」が「事業全体の抜本的見直し」に統合されることになった。この点についてメディアは「廃止」が無くなったと報道しているが（朝日新聞　2013.4.3朝刊）、実際には、「廃止」が消滅したわけではなく「抜本的見直し」に統合されたのであり廃止の判定は現存している。

「廃止」判定の見直しには、行政改革推進会議や議員の意見が影響したと思われる。行政改革推進会議では、複数の民間議員から「無駄撲滅」だけではなく生産性や効率性を向上させるための考え方を入れてゆく必要性があるという意見が出されている（第1回行政改革推進会議　2013年2月27日議事要旨）。「廃止」判断を「抜本的に見直し」に統合させた背景には、先の行政改革推進会議の意見および政治的には前民主党政権との違いを表わそうとしたものと思われる。

「レビューの対象」には新たに基金が加わった。基金とは、会計年度を越え2年以上にわたり支出がみとめられている予算である[11]。基金がレビューの対象になったことは、あまり注目されていないが、大きな意味があったと考える。基金は補正予算で積まれることが多く、緊急雇用対策として創出された「雇用創出の基金」など社会的急務に対処するために組まれたものもある。

この基金は、国から拠出され、都道府県に基金を造成し地方自治体が実施するというように、複数段階を経て支出されており、その管理は複雑である。また基金にかかる計画が当初予算で組まれていない場合、概算要求時に作成された計画よりも粗くなる可能性もあるため、計画の質がどこまで担保されているのかを確認することも重要である。したがって、資金の流れを明確にすることを求める行政事業レビューはこうした基金の点検に適している。

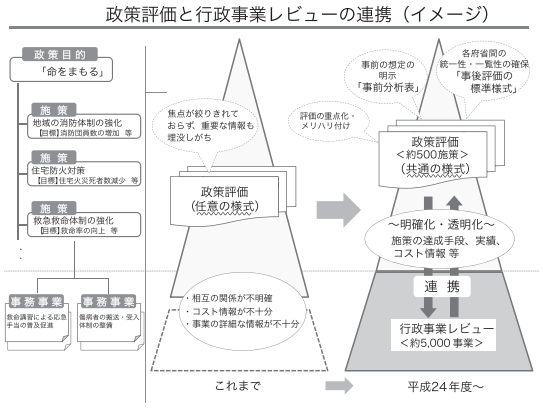
(2) 政策評価制度の対応　～行政事業レビューと政策評価の共通管理～

行政改革推進会議では、行政事業レビューとともに政策評価制度についても議論された。新藤義孝前総務大臣は行政事業レビューとの連携を念頭に次のような改善案を提示した。すなわち、全府省あわせて500程の政策とその下位にある5000の行政事業との関係をより明確にするというものである。図3は平成24年度以降とそれ以前の政策体系を比較したものである。新藤大臣は図3を用いて、これまで行政事業レビューとの関係が不明確であったものをより明確にすること、政策評価の対象を絞り込み、共通フォーマットを導入すると述べた[12]。ただし、これらの内容は、新政権下で

全く新しく作成されたというよりも、民主党政権時代に作成されたものと基本的には同じ内容となっている[13]。

また行政事業レビューとの連携強化の具体策として、事業名と事業番号を共通化することによって、一体的に管理するという案も示された。行政事業レビューと政策評価には類似の項目があるが、これを手動で転記していた状態であった。共通番号を導入することによって管理がしやすくなり、転記ミスの軽減につながると思われる。なお、電子化がより効果的な対応策であることは行政事業レビューにかかる事務局提出資料において言及されているが、具体案の提示には至っていない。

図3　政策評価と行政事業レビューの連携



（出所）行政改革推進会議　新藤義孝総務大臣提出資料（2013.4.2）

（3）未解決の課題

前述のように行政事業レビューは政権交代に伴い複数の改変が施された。だが、先に指摘した行政事業レビューの課題の全てが解決されたわけではない。以下、依然として残っている主な課題について述べる。

①双頭の目的

前述の行政事業レビューの「双頭の目的」の問題をどこまで解決できているのだろうか。「双頭の目的」の問題を解決するためには、行政事業レビューの判断を資金の流れによって行うのか、あるいは事業の効率性や妥当性評価によって行うのかを明確にする必要がある。ところが、評価項目に注目すると、「国費を投入しなければならないのか」「地方自治体、民間等に委ねることができない事業なのか」などの重要な判断を求める評価項目が列挙されているが、その割には記載欄が小さく、また判断方法に関する説明は記されていない。

他方、改定された行政事業レビュー・シートのレイアウトを見る限り「公的資金の流れ」のチェックにより重点がおかれたようにみえる。つまり「公的資金の流れ」をもって判断するのか、先の評価項目で判断するのかの優先度は明確ではなく依然として「双頭の目的」の問題は解消されていないようにみえる。

②定量化の問題

定量化問題については未だ大きな課題である。前述のように事業の上位にある政策評価作業で得られたアウトカム成果に関する定量情報を活用できれば、行政事業レビュー・シートで求められるアウトカム指標を記入することができるはずだが、現実には空欄になっているシートが散見された。

ただし、府省において、定量化についての経験の蓄積がないわけではなく、政策評価制度のもとで試行錯誤が続けられてきた。総務省が2011年に各府省に対して示した事前分析表や政策評価書の標準様式には、測定指標の記述欄があり、基準値および目標値、年度ごとの目標値の項目がある。具体的には、各府省は、施策にかかる事前分析表を作成し、それに基づき年度末に事後評価を行う。事前分析表には、施策名、担当部局名、作成責任者名、施策概要、政策体系上の位置づけ、達成すべき目標、目標設定の考え方、施策評価実施予定時期、測定指標（基準値、目標値、年度ごとの目標値、測定指標選定理由および目標値設定の根拠）の項目があり、以前にも増して定量的に記すことが求められるようになった。

しかし、その記述状況をみると測定指標の記述が不明確なものが少なくない。これは技術的な問題のみならず政策の性質に起因するものもある。例えば「ワークライフバランスの実現」という政策目標と「高速道路の建設」という政策目標では、その抽象度や社会的影響の範囲が明らかに異なる。後者はその成果の定量化が比較的容易だが、

前者はそれに比して容易ではない。また、仮に指標に基づいて目標値を設定できたとしてもデータ収集作業や作業コストの問題を克服しなければならない。

また、定量化に関連し形骸化の問題が指摘されて久しい。例えば、1970年代より実施されてきたODA評価では、既に評価の形骸化が指摘されており、平成13年の政策評価制度の施行時には既にこの問題が指摘されていたといえよう。こうした状況下で、特に着目したいのは規制の事前評価RIA（Regulatory Impact Analysis）である。なぜならば、比較的難易度の高い定量化作業に基づく評価を制度的に義務づけたのが本制度であるからだ。我が国では、規制の事前評価RIA（Regulatory Impact Analysis）が2005年より施行されたが、これによって、規制を閣議決定する前に、その社会的な影響を費用便益の手法を用いて分析することを義務付けられた。しかしながら、その実際には定量分析されたものは殆どなく、定性的な記述に留まっている。

行政事業レビュー・シートをみると定性的に成果を説明する欄は小さく、代わりに定量的に明記することが求められている。だが先の政策評価における試行錯誤の例が示すように安直な定量化は混乱を招く可能性がある。だが行政事業レビューの作業工程や時間、公開レビュー時間などの実際をみれば定量化の問題に技術論を踏まえじっくりと取り組む余裕はないように見える。その意味でも、定量化の問題は政策評価との連携の中で解決されるのが現実的であろう。

③段階判定の改変と概算要求への反映

「段階判定は、4段階から3段階になり、「廃止」判定が、「全事業の抜本的な見直し」判定に吸収されることになった背景には、前述のように、無駄撲滅ばかりでなく、生産性の視点を入れるべきという意見があったからである。

他方で、「廃止」には、廃止と判断されれば、次期概算要求には計上できないという強いメッセージ性があった。「廃止」がなくなることで、概算要求ルールが弱体化するのではないかと懸念されていた。実際こうした傾向があるようにみえる。2013年6月に実施された公開レビューは、レビューされる事業を所掌する府省が推薦した民間有識者と行政改革推進会議が推薦した民間有識者によって行われ、そのとりまとめは府省が推薦した民間有識者から選ばれた座長が行った。こうした状況下で、府省職員が「問題がある」と説明し、複数の民間議員が「廃止」や「抜本的見直し」と判断した事業にもかかわらず、座長が「現状維持」と結論を出すというケースが複数あった。

また同年11月に実施された公開レビューでは、6月に実施されたレビュー結果が概算要求に反映されているか否かを確認することになっているが、内容はおろか、名称も変更せず予算に計上されているケースもあった。

このようなケースが全体の何割程度を占めるのかについては検証が必要だが、「廃止」が「抜本的見直し」に統合されたことによって府省側の点検が甘くなったことが示唆されている[14]。

④省庁間の役割分担

前述の3点は、評価の技術的な課題である。しかし、その運営は微妙なバランスの中でなされているようにみえる。行政事業レビューは行政改革推進本部の、政策評価は総務省の所掌である。だが、行政事業も政策も本来ひとつの体系を構成するものであり、両者はこの体系の中で議論されることが望ましい。ならば、一か所で政策評価と行政事業レビューを管理すべきあるという考え方もあるが、所掌をめぐる議論はこれまでも繰り返されてきた。

新藤義孝前総務大臣が政策評価側から行政事業レビューとの連携強化に関する意見書を提出した背景には、自身の政策評価に対する強い関心[15]に加え、こうした府省間の微妙な問題があると思われる。

他方で、行政事業と政策評価を2つの省が別々に担うことによって、ある種のプラスの効果も生まれているようにみえる。行政事業レビューは評価結果を概算要求に反映させることが法的に定められているが、それは、現行の政策評価結果と概算要求が直接的に連動していないという重い問題提起をしているようにみえる。

またアウトカム成果にかかる指標と測定問題は、政策評価制度と行政事業レビューの双方から、

担当省庁に指導することができるが、そうしたプレッシャーが改善を促すかもしれない。

他方でひとたび歯車が狂うと大きな負担を生むことになる。この問題をどのように舵取りをするのかはまさに政権の手腕が問われるところであるが、それを省庁間の所掌をめぐる争いに留めないためにも、政策と事業をひとつの体系として捉え議論することが重要になる。

6. 結語　～さらなる課題へ～

本論は、行政事業レビューの特徴および課題に着目し、政権交代の前後の改変とその背景について、行政改革推進会議での議論を交え説明した。そして、複数の課題を指摘したが、特に政策評価制度との関係をどう整理するかは重要である。

2013年4月2日に開催された第2回行政改革推進会議ではこの問題が主要なアジェンダとなった。同会議で発表したのは私と新藤義孝前総務大臣であったが、二人の発表に対する推進会議メンバーの反応は象徴的であった。メンバーの発言の殆どが行政事業レビューというよりも、政策評価に関するものだったからだ。無論、百戦錬磨の新藤大臣と当方の未熟なプレゼンテーションでは他のメンバーの理解や関心を惹きつける上で比較にならない。だが、それ以上に、メンバーの関心は行政事業よりも政策にあることは明らかだった。そして、その発言をよく聞いてみると、府省別の政策というよりも、それらを束ねたより上位の政策にあることがわかる。あるメンバーはプログラム評価の必要性を述べた。プログラム評価は複数の意味で使われることがあるが、ここでは2つの含意があった。ひとつは社会保障や高齢者対策のように複数府省の政策を束ねるようなより上位にある政策評価という意味である。もうひとつは米国等で開発されたプログラム評価手法のことで、科学的に評価を行うべきという意味で述べられていた。

私は、プログラム評価は喫緊の課題であると考える。現行政策評価制度でも複数省庁にまたがる政策の評価が行われている。しかし、府省の上位にある政策を評価し、その結果について責任を持つ主体の所在が現行評価制度では曖昧である。例えば、府省の上位に位置する内閣官房および経済財政諮問会議や規制改革会議は、法的には政策評価の対象にならない。これらの機関は政策評価法第二条に定めるところの「行政機関」に当てはまらないためだ。内閣官房は内閣法第二十条に設置法によって定められたものであり、経済財政諮問会議などは内閣設置法に基づく調査審議機関であり、「行政機関」とは異にする主体である[16]。しかし、政権与党の主要政策は内閣官房や調査審議機関で検討され策定されたものが少なくない（田中2008）。だが、これらの評価とその責任の所在は、前述の法制度上の限界があるために曖昧にならざるを得ない。

実は、この問題を具体的な対案をもって提起していたのが、民主党政権が掲げた政策達成目標明示制度であった。これは、2009年に内閣官房に設置された国家戦略局が提示したもので、政権として最優先の政策で構成する大目標をトップダウンで作成する。各府省は大目標の達成状況について数値目標を立てて進捗を管理する。そして、これらのとりまとめを内閣官房が行うという新しい政策評価制度であった。だが、政策達成目標明示制度は閣議決定（平成21年10月23日）されたものの実行に至らず、野党から閣議決定違反という批判を受けながらも[17]、あいまいなまま消滅してしまった。しかしながら、先の行政改革推進会議でのプログラム評価の議論に鑑みれば、達成目標明示制度は、政策のプログラム化の方法や体制について、その失敗も含め重要な問題を提起している可能性がある。

自民党政権は、民主党政権が行ったものでも良いものは引き継いでゆくとして、行政事業レビューの続行を決めた。前政権が行ってきたことを頭ごなしに否定をせず良いと思われるものは活用するものを残すというのは合理的な判断である。だがその考え方を一歩進め、政策にまで視野を広げることでより多くの教訓を引き出すことができるのではないだろうか。

謝辞

本論は、冒頭で述べたように、政権交代に伴う制度の改変過程のについて記録として一助になればという思いで記している。しかし、迷いながらの執筆であり、不充分な点があることは否めなかった。そのような文章にもかかわらず、真摯に改善点を指摘してくださった査読者に感謝する。

（初稿投稿日　2013年7月24日）

脚注

1 2013年2月1日の稲田朋美大臣記者会<http://www.cao.go.jp/minister/1212_t_inada/kaiken/2013/0201kaiken.html>

2 視点・基準は以下のとおり
- 事業目的が妥当であるか。
- 財政資金投入の必要性があるか。
- 手段として有効であるか。
- 手段として効率的であるか。
- 他の事業と比べて緊要であるか。
- 国が実施すべき事業か。地方自治体、民間等に委ねるべき事業 ではないか。
- 支出先の選定は妥当か。競争性が確保されているか。
- コスト削減に努めているか。その水準は妥当か。
- 受益者との負担関係は妥当か。
- 具体的で適切な成果目標を設定し、成果や実績の検証が行われているか。達成度は着実に向上しているか。
- 類似事業があるか。他部局・他府省等と適切な役割分担・調整が図られているか。
- 透明化、情報開示の徹底が図られているか。 等

3 例えば、2013年6月に実施された行政事業レビューでは、外務省の「平和構築人材育成事業」において、外務省から受託を受けた団体は、団体の理事が勤務する大学に再委託しており、利益相反ではないかと指摘されている。

4 行政事業レビューは夏の時期にレビュー・シートをもとに事業を審査し、各府省から概算要求が出された秋に夏のレビュー結果が適切に予算に反映されたかを点検する。その議論は公開の場で行われるが、ここで財務省主計局の担当者は制度官庁として議論に参加している。

5 府省では政策評価に民間有識者に参加の機会を設けている。例えば、外務省では、法律に基づかない政策評価と法律に基づくそれとの双方を実施し、前者についてODA評価有識者委員会を設け、民間識者を中心に評価を行っていることもある。したがって、府省評価への民間人の参加が、行政事業レビューが最初であるわけではない。

6 総務省は政策評価結果に基づき、該当する府省に対して改善点などを勧告する。それを受けた府省は勧告に基づく対応案を提示することになっている。しかし、その内容は、総務省が改善点の検討を示し、府省はそれについての検討案を説明するものになっている。したがって、検討に留まり、改善を実現することを義務付けたり、それが実現しない場合に何らかのペナルティーが課せられるものではない。

7 最終的な判断基準は以下のとおり。
- **廃止**：事業全体として効果が薄い場合、国の事業として実施する必要性が認められない場合 等
- **抜本的改善**：効果の薄いメニューが多いなど、仮に国の事業としての必要性は認められても、事業内容について大幅な見直しが必要と判断される場合 等
- **一部改善**：国の事業としての必要性は認められるが、効果の薄いメニューがいくつか含まれるなど事業の一部に見直す点がある場合 等
- **現状通り**：特段見直す点が認められない場合 等

8 このような判断基準になっていた背景には、予算監視・効率化チームに副大臣や政務官などの政治家が参加していたことが考えられる。政治家は予算の流れや使途などの細かな点をもって判断することは物理的に困難でありこのような判断基準になったのではないかと推測する。

9 小野（2013）によれば政策評価書の8割には定量指標が記されているものの多くはアウトプット指標に限定されていること、また指標が政策目標をカバーできていないものがあると指摘している。

10 たとえば、総務省の「情報通信技術利用環境の整備」政策では多くの施策が盛り込まれ過ぎていて、事業との関連がよくわからず、結果的に政策と事業の紐つけがうまくできないケースがある。他にも「恩給行政の推進」政策なども類似の問題を包含しているようにみえた。

11 基金とは複数年にわたる事業であって、各年度の所

要額を弾力的な支出があることその他の特段の事情があり、あらかじめ当該複数年度にわたる財源を確保しておくことがその安定的かつ効率的な実施に必要であると認められるもの（補助金等適正化法施行の一部改正　平成26年10月17日）

12 共通フォーマットの導入については、既に平成23年度より、全省庁において試行的に行われてきた。これによって、わかりやすさや、比較可能性が向上したという利点がある。だが、同時にその問題も明らかになっている。制度官庁、事業官庁という言葉があるが、前者は、財政規律などの抽象度の高い政策を扱う省庁と、高速道路の施行など具体性の高い政策を扱う省庁のことをさす。共通フォーマットは国交省のフォーマットを参考したと言われているが、それは事業型の政策には適している。しかし、制度官庁が扱う政策にはうまく適合しない項目がある。

13 政策評価制度は、平成22年4月より、「事業仕分け」の施行に伴い、その連携を念頭に、「目標管理型の政策評価」として、事前分析表および評価書の標準様式を検討した。

14 平成26年度には「廃止」枠が独立した判定枠として復抬した。

15 新藤 義孝前総務大臣は、政策評価法が制定された平成23年時、総務大臣政務官であり、政策評価法創設を担ったという自負を持っている。また、平成23年度には衆議院決算行政監視委員会委員長に就任するが、そこでの議論が政策評価に基づくというよりも、「仕分け」によるものばかりで、疑問を抱いたという。こうした中で、政策評価の役割をより明確にすべきという考えを持っている（新藤前大臣との面談　2013年9月3日）

16 内閣府設置法第十九条第一項第一号及び第二号

17 平成23年3月9日に行われた、衆議院内閣委員会で、小泉進次郎議員が達成目標明示制度が施行されておらず、閣議決定違反ではないかと質問している。

参考文献

小野達也（2013）「政策評価と実績測定～府省の実績測定における計量・計数の現状～」、『日本評価研究』、13（2）: 21-36

大迫丈志（2012）「事業仕分けと行政事業レビューー意義と課題ー」、『調査と情報』、757 : 1-12

兼村高文（2012）「「事業仕分け」の意義を再考する」、『月刊自治研』、54（632）: 18-19.

田中弥生（2008）「市場化テストのPDCAサイクルに見る評価手法の考察～マニフェスト評価体系からのアプローチ～」、『日本評価研究』、8（1）: 95-114

南島和久（2011）「府省における政策評価と行政事業レビュー　ー政策管理・評価基準・評価階層ー」、『会計検査研究』、43 : 57-71

林和喜（2004）「実績評価の現状と有効性検査」、『会計検査研究』、29 : 185-202

東田親司（1999）「政策評価制度の導入をめぐる論点」、『季刊行政管理研究』、86 : 7

朝日新聞　2013年4月3日朝刊

稲田朋美行政改革担当大臣発言要旨　平成25年1月29日 <http://www.kantei.go.jp/jp/singi/gskaigi/dai1/siryou03_1sankou.pdf>

第2回行政改革推進会議　新藤義孝総務大臣提出資料　平成25年4月2日<http://www.kantei.go.jp/jp/singi/gskaigi/dai2/siryou4.pdf>

第2回行政改革推進会議　田中弥生議員提出資料　平成25年4月2日<http://www.kantei.go.jp/jp/singi/gskaigi/dai2/siryou4.pdf>

第1回行政改革推進会議　配布資料3-1「行政事業レビューについて」<http://www.kantei.go.jp/jp/singi/gskaigi/dai1/siryou03_1.pdf>

第1回行政改革推進会議　配布資料3-1　g行政事業レビュー参考資料<http://www.kantei.go.jp/jp/singi/gskaigi/dai1/siryou03_1sankou.pdf>

第1回行政改革推進会議　議事要旨<http://www.kantei.go.jp/jp/singi/gskaigi/dai1/gijiyousi.pdf>

第2回行政改革推進会議　配布資料1　「行政事業レビューの実施等について」<http://www.kantei.go.jp/jp/singi/gskaigi/dai2/siryou1.pdf>

「行政事業レビューについて（案）」（行政刷新会議（第6回）資料2-1）（平成22 年3 月11 日）行政刷新会議ホームページ <http://www.cao.go.jp/sasshin/kaigi/honkaigi/d6/pdf/s2-1.pdf>

内閣官房国家戦略室「予算監視・効率化チームに関する指針」（平成22 年3 月31 日）国家戦略室ホームページ<http://www.npu.go.jp/policy/pdf/20100401_yosank-ansi_sisin.pdf >

「行政事業レビューの実施について」（平成23 年6 月7

日閣議決定）
総務省　事前評価の拡充等<http://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/zizenhyouka.html>
総務省　規制の事前評価に関するガイドライン　平成19年<http://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/pdf/070824_2.pdf>
総務省　規制の事前評価http://www.soumu.go.jp/menu_seisakuhyouka/jizenhyouka.html
総務省政策評価<http://www.soumu.go.jp/menu_seisakuhyouka/kekka.html>

（2013.12.26 受理）

Administrative Project Review
～Beyond Change of Governmental Regime～

Yayoi Tanaka

National Institution for Academic Degrees and University Evaluation
ytanaka@niad.ac.jpp

Abstract

Administrative project review is newly started assessment system by the former cabinet governed by the Democratic Party. This review focuses on the effectiveness of the individual projects and on the appropriateness of the budget expenditures of all ministries. In fact, they reviewed about 5000 project a year. After a changeover of the government from Democratic Party to Liberal Democratic Party in 2012, the Prime Minister Abe decided to carry out administrative project review for the purpose of administrative reform. Then Administrative Reform Promotion Council chaired by Prime Minister has started in February 2013. The Council deliberated the policy and procedure of this review system and also discussed division of roles between this review system and the policy assessment system that is operated by the different ministry. In this paper, I will overview the administrative project review system, and will explain what and how the Council revised this system under Prime Minister's initiative.

Keywords

Administrative Project Review, Budjeting, Quantification, Systematic Policy Desigh Policy Evaluation.

【研究論文】

蓋然論理とその評価方法論的含意

西村　君平
広島大学 高等教育研究開発センター
k-nishimura@hiroshima-u.ac.jp

要　約

本稿では, 蓋然論理（probative logic）とはなにか, 特にその評価方法論的含意に着目して明らかにする。蓋然論理はマイケル・スクリヴェンの提唱する評価論理のキーワードの1つであり, 主に評価における専門的判断のあり方を分析・改善していく理論である。従って, 蓋然論理に関する理論的理解を獲得することは, スクリヴェンの評価論理に関するより詳細な理解や評価の中の専門的判断の方法論構築に資するものである。

本稿では, まず蓋然論理の問題設定を知るために, 蓋然論理の誕生の文脈である科学哲学, その中でも特に科学方法論に遡る。次に蓋然論理が着目する専門的判断とは, どのような特徴を持っているのか概観する。その上で科学方法論における専門的判断はどのような位置を占めているのか, 蓋然論理はその分析や改善にどのように役に立つのか検討する。最後に, 蓋然論理の観点から評価方法論について検討し, 特に構成概念妥当性という発想が持つ意義について議論する。

キーワード

蓋然論理, 評価論理, 価値自由原理, 蓋然性, 構成概念妥当性

1．研究の背景と目的

本論考の目的は, マイケル・スクリヴェンの提唱する評価論理のキーワードの1つであるprobative logicとは何か整理し, その評価方法論的含意の一端を明らかにすることである。

評価論理とは評価の定義, 評価の主要概念の定義（例：形成的評価と総括的評価）, 評価とそれ以外の社会調査との本質的な差異に関する考察, 評価に係る推論上のルールといったテーマを扱い, これらのテーマに関する論理学的な考察を行うものである（Scriven, 2005, p.235）。評価論理が扱うテーマは多岐にわたるが, 基本的に評価論理は評価を脱人格的で体系的な「手法」には還元できない, 評価者が行う複雑な認知的活動としての「推論」と見なす点に特徴があると言われる（Patton, 2012, p.105）。

probative logicは, この評価の手法には還元できない推論の側面に強く関わっている。probative logicは, 確からしい（prima facie）結論に至る推論の過程を分析するものである（Scriven, 1991, p.220）。確からしい結論に至る推論とは, 具体的に言えば, 裁判における司法判断のような専門的判断の中で行われる思考過程を指している(Scriven, 1987, pp.9-10)。そして評価において評価者が下す専門的判断, およびその中の推論のあり

日本評価学会『日本評価研究』第14巻第1号、2014年、pp.17-30

方もprobative logicの主な分析対象の1つである。

probative logicは, 専門的判断の推論の側面を分析するにあたって, その蓋然性に着目したアプローチをとる。論理学では知識の真偽が二値的に決定される場合その性質を確実性（certainty）と呼び, 真偽が多値的に決定される場合その性質を蓋然性（probability）と呼ぶ（渡辺, 1987）。スクリヴェンによれば, 専門的判断では, 蓋然性の中でも特に定性的蓋然性（qualitative probability）が問題となる。定性的蓋然性とは推論において用いる「前提の意味の確かさ」[1]である。これは定量的蓋然性と対になる概念で, 定量的蓋然性は概念を経験的に検証する過程で問題になる「前提の計量的な確かさ」を意味している。定量的蓋然性については確率論や統計学が扱ってきたが, 定性的蓋然性については看過されており, ゆえにその向上に資する知見の産出を眼目とする蓋然論理が必要になる（Scriven, 1987, p.29）。このことから本稿ではprobative logicを蓋然論理と意訳する。

このように蓋然論理は, 評価における専門的判断のあり方を論理学的に分析するものであり, 評価論理のキーワードの1つである。よって, 蓋然論理の内実を明らかにする作業は, 評価論理に関する理解を深化し, スクリヴェンが提案する評価方法論の改善に寄与するものである。またスクリヴェン以外の評価方法論に潜む専門的判断を析出し, それを改善していくためにも有益だと期待できるだろう。

しかしそれにも関わらず, 評価研究において蓋然論理はこれまで注目されて来なかった（例外的にMigotsky et. al., 1997）。蓋然論理が看過されてきた一因として, 蓋然論理に関するスクリヴェンの議論が, 専ら科学哲学の文脈で行われており, 評価研究にとってそれがどのような意味を持つのか必ずしも明確ではないことが挙げられるのではないだろうか。

このような着想から, 本稿ではスクリヴェンが科学哲学の文脈で行った蓋然論理についての議論を掘り起こし, スクリヴェン自身ですら必ずしも明確に出来ていない蓋然論理の要諦を明らかにするものである。

2．先行研究の検討と本稿の位置づけ

ここでは先行研究を手がかりとして, 蓋然論理が評価研究において従来どのように位置づけられ（1項）, 解釈されてきたのか（2項）検討し, 本稿の位置づけを明確化する。

（1）評価論理の中の蓋然論理

アメリカの評価研究では, 評価論理を扱った教科書的な解説がいくつか存在する（Shadish et. al., 1993； Alkin, 2004）。我が国では評価論理に関する第一人者である佐々木（2010）が, 評価論理に焦点を絞り, その知見を活用して評価の一般的方法論について解説している。

これらの解説ではスクリヴェンの評価論理は, 評価とそれ以外の社会調査との違いを明示し, それにより評価の専門職化や評価研究の焦点化の基礎を築いた理論として, 評価研究に広く受け入れられていることが指摘される。論者によって, スクリヴェンの評価観に関する説明は若干異なるが, スクリヴェンの原典に忠実で最も理解しやすい説明は佐々木（2010）による「評価＝事実特定＋価値判断」（p.3）という模式的な整理であろう[2]。

このモデルは, いわゆる社会調査は対象の事実的な特性を明らかにする行為であり, 評価は事実特定と価値判断の二つの部分から構成された行為であることを示している。一般的な手続きとしては, まず価値前提（criteria, standards）を明確化し, 次に対象の事実前提（performance）を特定し, 最後に価値前提と事実前提を総合する（synthesis）という形で実施される（Scriven, 1982, p.18；2007, p.5）。スクリヴェンによれば, 蓋然論理は特に価値判断（価値前提を明らかにする過程）に関わるものである（Scriven, 2012）。

ただし蓋然論理は, 価値判断の方法論として誕生したものではない。そもそも蓋然論理は科学から価値判断および専門的判断を排除しようとした価値自由原理・論理実証主義に対する批判を通して彫琢されたものであり, 科学哲学の一環として成立した経緯を持っている（Scriven, 1987, pp. 26-27）。

この価値自由原理・論理実証主義への批判の結

果として, 価値判断および評価が科学的な実践として成立し得るという結論が得られ, そこから更に, 科学的実践としての価値判断や評価が持つ論理的な特性に関する方法論的な含意が得られることになった。

このようなやや複雑な経緯もあってか, 蓋然論理は評価研究の中で必ずしも共有されていない。この点についてマイケル・クイン・パットンはスクリヴェンの論稿「価値判断の論理」の重要性を示唆した上で, 次のように評価研究者に問いかけている。

「あなたは価値自由な社会科学という主張の論理的な錯誤について理解し, それを明確に説明することができますか。なぜこの事が評価の実践にとって重要な問題なのか理解し, それを明確に説明することができますか。また, あなたは価値判断の論理および蓋然論理について理解し, それを明確に説明することができますか。あなたは自らが普段行っている評価の利害関係者に対して, 蓋然論理に基づいて, 評価対象の価値や意義についての自らの判断を説明し, 報告することができますか。こうしたことを自分自身に対して説明することができますか」(Patton, 2012, p.105)。

パットンのコメントは, 評価研究において価値判断の正当性を担保するための仕組みや方法の開発が急がれていることを示すと共に, その価値判断の基本原理の1つである蓋然論理の概要が未だ共有されていないことを示唆している。こうした状況に鑑みて, 本稿の第一の課題を, 科学哲学的な文脈における蓋然論理に関する議論を掘り起こし, 蓋然論理の固有の問題意識やアプローチ, その理論史的な位置づけを明確にすることとする。

(2) 蓋然論理の批判的考察

これまであまり関心を払われてこなかった蓋然論理であるが, 例外として, ミゴツキーらは蓋然論理を批判的に考察した論稿を発表している(Migotky et. al., 1997)。

ここでは評価論理の中でも特に価値前提と事実前提を総合する過程が議論の遡上に上がっている。スクリヴェンは総合の過程には「臨床的な推論, 直感, 専門的判断に対する過信」が横行していることを問題視し, 形式的・明示的な手法での総合を提案している。これに対して, ミゴツキーらはスクリヴェンの提案は, 総合から専門的判断を排除するものであると非難している。スクリヴェンの議論は「社会的相互作用よりも技術合理性を重視するものであり, 評価者が社会的正義の問題を理解し, その問題に取り組むことを制限するものである」(Migotsky, et. al. 1997, p.459)。

詳細は後の蓋然論理に関する整理に譲る他ないが, 結論から言えば, ミゴツキーらの批判は誤解を含んでいる。上で触れたように, 蓋然論理はどちらかと言えば総合よりも価値判断の過程に重きをおいている。なにより, 専門的判断を重視し, その分析と改善を眼目とする理論である。確かにスクリヴェンは現在の評価における専門的判断のあり方を批判しているかもしれないが, それはあくまでも専門的判断の改善が目的であって専門的判断への信頼を消滅させるものではない。また蓋然論理は, 認知的行為と思われがちな推論に, 社会的相互作用の側面が含まれていることを強調するものである。

このような先行研究の蓋然論理に対する誤解をとくためにも, 本論稿は第二の課題として, 蓋然論理が評価の方法論, 特に評価おける推論や専門的判断とどのような関係にあるのか, そしてどのような方法論的提案が可能なのかを明らかにする。

(3) 本稿のアプローチ

以下では, 3節で蓋然論理の誕生の文脈である科学哲学に遡り, 特に蓋然論理の論敵, 価値自由原理についてスクリヴェンの議論に沿って整理する。4節で価値自由原理が捨象し, また蓋然論理が着目する専門的判断における推論のあり方について概観する。5節で科学方法論における専門的判断の位置づけについて検討し, その知見に基づき具体的にどのような評価方法論的な提言が可能なのか議論する。最後に結論として知見の要約と本稿の意義, 残された課題を述べる。

こうした理論的検討を通して, 我々は単に蓋然論理の内容を知るのみならず, その科学哲学的な背景が, どのような理路を通って評価の方法論に繋がっているのかを知ることができる。このアプ

ローチは, 比喩的に言えば, スクリヴェンが考案した理論のアウトプットだけでなく, そのスループットも理解することで, 蓋然論理や評価論理の批判や改善に資する知見を得ようというものである。

3．価値自由原理の科学観

(1) 定義

スクリヴェンによれば, 価値自由原理とはいわゆる経験主義や実証主義の中でも, 特に論理実証主義と密接に結びついた方法論的なルールである。それは「科学, 特に社会科学は『科学の中で』評価的な結論を導出するべきではない, あるいは導出することはできないという信念である。『科学の中で』の評価とは, 検証された真実や定義的真実を前提とした評価のことである。」(Scriven, 1991, p.373)。

まず価値自由原理の概要をつかむために, この定義が「価値自由原理が考える科学的な真実の特徴」を示唆している事に留意したい。定義の第一の部分では, 価値自由原理とは科学の中で物事の価値を決定することは出来ないという考え方であることが述べられている。次の部分では, 価値自由原理が想定する「科学」とは何かを示すキーワードとして「検証された真実と定義された真実」が提示されている。本稿ではこの価値自由原理が想定している科学的方法とは一体どのようなものなのかに着目する。この点を詳らかにすることが, 価値自由原理の主張内容をより正確に知ることにつながり, ひいては価値自由原理の論理的な錯誤を知ることにつながるからである。

(2) 分析的判断・総合的判断の二元論

価値自由原理・論理実証主義のように, 真実を「検証された真実」と「定義された真実」で捉える発想は, 我々の判断を分析的判断と総合的判断に区別する古典的な議論に由来している (Scriven, 1959b)。

分析的判断は, 例えば「すべての独身男性は結婚していない」といった命題で表現されるものである。この命題において「独身男性」と「結婚していない」という概念は同語反復的な関係になっている。両者の関係は言葉と言葉の関係であるため, 分析的判断が真であるか偽であるかの判定は, その言い換え方が正しいかどうかを規定する論理的あるいは言語的な規則によってのみ判定される (パトナム, 2006)。規則に合致していればスクリヴェンの言う定義的に真 (truth by definition) であり, 論理実証主義の言う規則的に真 (truth by rules) となる　(Carnap, 1936)。

次に総合的判断とは, 例えば「全てのレモンは黄色である」という命題で表現される。「レモン」「黄色」という概念はそれぞれ別の指示対象を持っていて, 両者を結びつける論理的・言語的規則は存在しない。よって, 総合的判断の真偽を判定するものは, 経験的検証である。原理的に言えば, 現実のレモンを全て観察してその全てが黄色であればこの命題は経験的に真となる (Carnap 1936)。

分析的判断と総合的判断に関する区別は少なくともエマニュエル・カントにまで遡ることができる古典的な議論である。しかし, 論理実証主義は分析的判断と総合的判断に関する古典的な議論を, 我々が真偽をはっきりと判断できるのは, 分析的判断と総合的判断のみであるという二元論へと敷衍する点で, 新しい主張を含んでいた (パトナム, 2006, pp.2-3)。

まず論理実証主義は, 有意味な知識とは真偽を検証できる知識であると仮定する[3]。そして我々が真偽を検証できる知識は分析的判断か総合的判断のいずれかであると考える。分析的判断を扱うのであれば, 概念および命題の真偽は論理的な規約によってのみ判定することができる。代表例は数学である。総合的判断を行うのであれば, 概念・命題の真偽を事実によってのみ判定できる。これは経験科学, とりわけ経験的検証の正確性において大きな成功を収めた物理学が行っていることである。それ以外の判断を含む知識は真偽が不定であり「認識的に無意味」である。

(3) DNモデル―観察文と理論文―

論理実証主義は, 分析・総合の二元論を用いて, 科学的知識の確かさ, すなわち蓋然性を高めるための科学改造運動を起こした。改造の基本方針

は, 科学における全ての命題を観察文とそれを補助する理論文のいずれかに置き換えていくというものであった（カルナップ, 2003）。

観察文は「観察できる事物または事象を記述するために, 観察できる性質および関係を指示する用語を用い」（カルナップ, 2003, p.192）た命題で「黄色のレモンが1つ見える」と言った文章がこれに該当する。論理実証主義における観察の意味は限定的で, 観察とは人間の五感を通した直接的知覚, いわゆる感覚所与を意味している。

理論文は「観察できない事象, 事象の観察できない側面または特徴」（カルナップ, 2003, p.192）を示すもので, 観察文に意味を与えるための規則や法則である。ただし論理実証主義のいう理論文は非常に限定的で, 理論は記号論理学によって表されるものでなければならない。すなわち, 理論とはそれ以上分割できない単純な原子命題（例：黄色とは, 波長が570-585nmの色である）と原子命題をつなぐ論理記号（例：∧や∨）によって表されるものである（ラッセル, 2007）。

この観察文と理論分を組み合わせた理想の科学的説明は, 演繹的・法則的モデル（Deductive Nomological model, 以下DNモデルと略称）と呼ばれる（図1）。DNモデルは, 一般法則（理論文）と特定の条件（観察文）が大前提・小前提としての役割を果たす演繹的な三段論法である[4]。

例えば, ある時刻tに日食が起こるのはなぜかということを説明する場合を考えてみたい。この説明のために一般法則として, 例えば恒星や惑星の運動に関する一般的な運動を記述する理論が用いられる（大前提）。次に観察によって明らかになった過去の時刻t'における恒星系の正確な位置と速度を明らかにし, これを条件として扱う（小前提）。この理論と前提から結論Eとして, 時刻tに日食が起こるという結論が得られる（Hempel and Oppenheim, 1948, p.138）。

このような形で科学的知識を観察文とそれを補助する理論文に変換にしてしまえば, 科学的な説明のために必要となる推論は, スクリヴェンの言葉で言えば「小学生でも間違わないような」（Scriven, 1959b, p.462）単純な三段論法に置き換えることができる。

図1　DNモデル

C1,C2, C3, … Cn	条件
L1, L2, L3, … Ln	法則

E	結論

DNモデルは複雑な推論や論理を伴わない。それは科学的知識とは観察に基礎づけられた知識であり, 観察による検証が不可能な知識は, そもそも科学の範疇に入り得ないということを含意するモデルである。

（4）価値判断の排除

観察のみによって科学的知識の蓋然性を高めるためには総合的判断（観察文）でも分析的判断（理論文）でもない「その他の判断」は科学から排除されなくてはならない。ここで倫理学のような価値に関わる学問が排除対象となった（パトナム, 2006, 20-22頁）。

例えば「殺人は罪である」という価値判断において「罪」とは一体なにか。「罪」を直接的に観察することは不可能である。また罪の意味する所を統一的に決定し, 観察可能な概念へと置き換えるための定義や規約は存在しない。よって「殺人は罪である」という判断は真偽不定であり, 科学の蓋然性の低下の要因に過ぎない。

このように価値自由原理は, 価値前提は知覚による観察が不可能であるという理由で, それを科学から排除しようとするものである。知覚による観察の可能性をこれほどまでに重視する理由は,「我々が科学の中で行い得る判断は分析判断と総合的判断のいずれかである」という二元論に由来する。

そしてスクリヴェンは, 価値自由原理の二元論を崩し, 価値自由原理が「その他」と退けた諸々の判断は実は科学の方法論の中に組み込まれるべきものであるということを示す中で, 蓋然論理を彫琢していく。

4．専門的判断の文脈性

本節では分析・総合の二元論に対する批判の前に, 議論のレベルをより実践に近いものに戻すためにも「その他」の判断の中でも蓋然論理にとり

わけ密接に関わる専門的判断の概要を整理しておきたい。

(1) 典型例としての司法判断

スクリヴェンによれば, そもそもprobativeは法学に由来する概念である（Scriven, 2012, p.22）。法学では, probativeはprobative powerといった形で用いられ, 事実や意見の証明に関する効果, すなわち証拠力を表すために用いられる[5]。

裁判における証拠力の推定の例が示すように, 様々な職業, 特に高度な認知的な活動が求められるような専門職において, 専門家は実務的で学術的な推論を日常的に行っている。専門的な判断は, 数学の証明のような明快な概念, 厳密な論証, 確かな結論を生むものではない。例えば裁判では「証拠力」のように意味が不十分にしか定まらないような多義的な概念を用いる。そして非常に複合的なルールに則って, 様々な角度から現実を検証していく過程をたどる。また一度下された結論も, 裁判の三審制の存在が示すように, あくまでも現状における最も確からしい結論（prima facie）として扱われる[6]。このように, 専門的な判断は定義の不十分性・過程の煩雑性・結論の不安定性を伴っている（Scriven, 1987, p.23）。こうした専門的な判断における推論を, スクリヴェンは蓋然的推論（probative inference）と呼ぶ（Scriven, 1987, pp.7-8）。

具体的な例としては, 料理, 外科手術, 工場での作業, 裁判, パドック, オフィス, 戦場など様々な場面での様々な推論, そして後述する科学者が実験中や観察中に行っている推論が想定されている（Scriven, 1987, p.7）。そしてもちろん, 評価もその一種に含まれる（Scriven, 1987, p.32）。

(2) 専門的判断への着目

専門的判断における推論は, 従来的な論理学, 特に記号論理学では取り扱われて来なかった。なぜなら論理学は特定の領域や内容を伴った推論の実像を扱う学問ではなく, 推論の構造によって結論の真偽が左右されるような, 脱領域的で抽象的な推論の規約を抽出し分析する学問だからである。とりわけ, 前提が真であれば結論が必ず真となる演繹的推論の論理構造が, 形式論理学や記号論理学の主な関心となってきた。

しかし, 現実の専門的判断は, 演繹のような「静態的な構造を持った完全な推論ではなく, むしろ結論を根拠付けるための前提を必ずしも全て明示化できないような, 不完全な推論」（Scriven, 1987, p.9）である。

スクリヴェンは, 当時の論理学で軽視されていた専門的判断にあえて着目することには, 論理学的必要を超えた社会的意義があるという。なぜなら, ますます高度化し, 複雑化していく専門的実践のあり方を分析, 改善していくためには, 専門家が実務的な場面でどのように思考しているのか明らかにし, その改善に資することが求められるからである（Scriven, 1987, p.18）。

(3) 文脈性：氷山のメタファー

スクリヴェンは, 演繹的推論と専門的判断における推論の違いをそれぞれ海底油田の採掘機と氷山に喩えて, その特徴を説明している（Scriven, 1987, pp.22-23）。

油田採掘機も氷山も安定しているように見える。採掘機は海上の機構を堅牢な柱で海底に直接連結し, 自らの安定を保っている。これは演繹的推論が導出する結論が, 全て既知の真なる前提と妥当な論理構造によって基礎づけられていることを表している。他方で, 氷山は海底に連結するのではなく, 自らを海に漂わせることで安定を保っている。採掘機にとって海は自らの安定性を脅かすネガティブな要因であるが, 氷山にとって海は自らの安定化に寄与するポジティブな要因である。この海にあたるものが, 文脈である。このように演繹的推論と専門的判断の顕著な違いは, 脱文脈的か文脈的かの違いに求められる。

また, 採掘機と氷山のメタファーを用いると, 従来の論理学と蓋然論理の違いは次のように説明できる。採掘機の改善に関心を持つ者は, その頑健性を高めるための手がかりを求めて, 採掘機の内的な構造の分析に焦点を置く。同様に, 演繹に対する論理学的な分析の焦点は, 推論の内的な構造に置かれることになる。他方で, 氷山の安定性に関心を持つ者は, 氷山の構造にももちろん関心をもつが, その構造を氷山の環境要因（水温や気温など）との対応関係で分析していく。更に, 氷山

の構造は, 水面上の見えやすい部分のみならず, 水中に隠された見えづらくしかし氷山の少なからぬ体積を占める部分についても分析のメスを入れる。同様に, 専門的判断に対する論理学的な分析の焦点は推論の内的構造と外的な文脈との対応関係に置かれる。その際には, 明示化された前提のみならず, 暗黙の前提にも分析のメスを入れていく（Scriven, 1987 , pp.22-23）。

５．定性的蓋然性：科学における専門的判断の位置

(1) 通常的判断

スクリヴェンは専門的推論の論理的構造を説明するために, 分析的判断・総合的判断につぐ新しい区分である通常的判断（normic statement）を導入する（Scriven 1959b）。

通常的判断とは「レモンは通常は黄色い」と言った形で「通常は」「典型的には」「特徴的には」「理念的には」と言った限定詞を伴う判断で, 日常の至る所で用いられている（Scriven, 1987, p.9）。

記号論理学が想定する判断は構成内容が有限である。これは「if X then Y1 and Y2 …Yn」という形式で表現される。他方で, 通常的判断では「if X normally Y1 or Y2 or …」というような形で表される。つまり通常的判断に含まれる概念の構成内容は無限である（Scriven, 1959b）。

例えば「レモンは通常, フルーツの一種で, 酸っぱく, 皮の色は黄色である」というような判断が, 通常的判断の具体例である[7]。しかし「表面がでこぼこしている」というような要素も, レモンの構成要素をなしている。ただし多くの場合, 誰かにレモンとは何かを解説するときにせよ, 目の前のレモンのようなものが何かを検討するときにせよ, 「表面がでこぼこしている」という要素はなかなかリストに上がってこない。このように, 通常的判断は構成要素が無限なので, それを用いた推論を行う場合には自明の理（truism）すなわち説明や検証の中に暗黙の前提を含むことになる（Scriven, 1959c, p.456）。これはちょうど氷山における海上部分と海中部分のような関係になっている。

(2) 導出的推論と選択的推論[8]：DNモデルの中の隠れた推論

通常的判断というカテゴリーが存在することは価値自由原理にとって何ら問題ではない。問題になるのは通常的判断が, 真偽検証可能なカテゴリーとして存在し得るかどうか, 言い換えれば通常的判断が科学の中に組み込めるかどうかである。

スクリヴェンは論理実証主義にとっての理想の科学的説明であるDNモデルにおいてさえ, 通常的判断が含まれていることを指摘し, この問題に対する回答を与えている[9]（Scriven, 1959c, pp. 459-463）。

まずDNモデルでは法則が条件項を1対1的に指示することが自明視されている[10]。しかし科学における法則は特定の条件項と直接的に結びついたものではない。よって法則の意味する所を理解した上で, その法則が期待する特定の条件を推定し, 法則から条件項を導出する過程が必要になる。スクリヴェンはこうした推論を導出的推論と呼ぶ。

例えば天体に関する一般法則は, それが一般法則と名付けられていることからも明らかだが, 地球・月・太陽だけを想定した理論ではないので, 日食に係る特定条件が地球・月・太陽の位置と運動であることを指示するものではない。地球・月・太陽の位置と運動を観察するべきであることは, 一般法則に対する適切な理解と複合的な推論の過程に基づいて導出されるのである。

このように何らかの理論を説明に用いることを決定したとしても, そのことによって我々が何を観察しなければならないか自動的に決まるわけではない。DNモデルのような簡潔なモデルの中にさえ「長大で複合的な推論の過程」（Scriven, 1959c, p.461）が含まれているのである。

またDNモデルでは説明に用いる理論の特定の過程が自明であるかのように見なされているが, 科学的説明においてどのような理論を用いて, どのような条件項を扱うか決定する過程は, 必ずしも自明ではない。

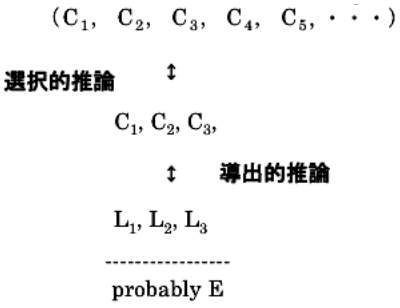
例えば日食の説明では, 天体に関する一般的な運動法則が理論として用いられているが, 厳密に言えばこの他にも光の直進性に関する法則が補助仮説として求められるし, また光の直進性に関する補助仮説は, ニュートン力学における様々な補

助仮説を求める。簡易な測量を行う場合には光の直進性は自明の理と見なされるが, 非常に厳密な天体観測においては光の直進性やニュートン力学に係る様々な要素ですら自明の理としては扱い得ないことになる。

こうした理論および条件の選択に関する過程を, スクリヴェンは選択的推論と呼ぶ。科学者は説明において用いる理論や条件項を調査によって明らかにするのではなく,「科学者はそれが適切なものかどうかを判断（強調原文）している」(Scriven, 1959c, p . 462)。

このように科学的説明は, 表面上は有限個の変数間の因果的効果について言明するものに見えたとしてもその背後には自明の理を持っている（図2)。実際の科学的説明には無限個の変数間の因果的効果が想定されていて, 導出的推論や選択的推論を通して, 説明で用いる概念を文脈に合わせて選択し, その文脈に応じて意味を解釈していく過程を不可避的に含んでいる。よって科学は直接的な観察や簡易な言語記号には還元できるものではない。

図２　DNモデルにおける導出的推論と選択的推論



「『証明』や『法則』,『説明』のみならず,『予測』『蓋然性』『観察』『情報』『理論』『反証』などといった論理実証主義者が形式的な論理分析に相応しいと考えた科学の概念のほとんど全ては, 本質的に主観的だったのである。ここでいう主観的とは, 論理実証主義の（注：蓋然論理のような）分析に対する蔑称を意味している。それはすなわち, プレゼンテーションや相互行為の実在的・仮想的な相手の知的水準や理解の程度に関する言及を含む推論を分析するものである」(Scriven 1987 p.26)。ゆえに我々は「推論が行われる際の環境的要因を推論の媒介変数として扱わなくてはならないのである。それは従来の論理学が分析の対象としてきた推論の言語的側面よりも多くの重要な情報を含んでいる。言語は文脈がなければ意味を持たないのだから」(Scriven, 1987, p.27)。

(3) 構成概念妥当性：定性的蓋然性と定量的蓋然性

論理実証主義はDNモデルに推論の過程が含まれていることを許容した上で, そのような推論は科学的知識の蓋然性を減じるものであって, あくまでも排除されるべきと反論するかもしれない。しかしスクリヴェンによれば, 導出的推論および選択的推論の方法論は「学習可能なものであり, 特定の者（注：科学者）は他の者よりもその能力に優れており, その能力はテスト可能なものである。このような能力が専門的判断と呼ばれる」(Scriven, 1959c, p.456)。

導出的推論, 選択的推論の存在は科学の蓋然性には, 2つの側面が存在するということを示唆している（Scriven, 1986, p.30)。1つめの側面は定量的蓋然性である。定量的蓋然性とは「前提の計量的な確かさ」である。科学的な知識を構成する前提は, 実験や調査によって検証されなければならない。2つめの側面は定性的蓋然性である。定性的蓋然性とは「前提の意味の確かさ」である。通常的判断は本来的に多義的であるため, 今回の推論の文脈において, なぜその前提を扱う必要があるのか, その前提はどのように解釈されるのか, 関連する理論や概念と関連づけて明確にしなければならない。定量的蓋然性は統計学の範疇であり, スクリヴェンが蓋然論理において主に取り扱うが定性的蓋然性である。

スクリヴェンは定性的蓋然性を向上させるための具体的な方法として, 構成概念妥当性（construct validity）の検討を示唆している(Scriven, 1986, p.38)。構成概念妥当性は, 評価研究者としても名高い心理学者リー・クロンバックらに由来するアイデアである（Cronbach and Meehl, 1955)。

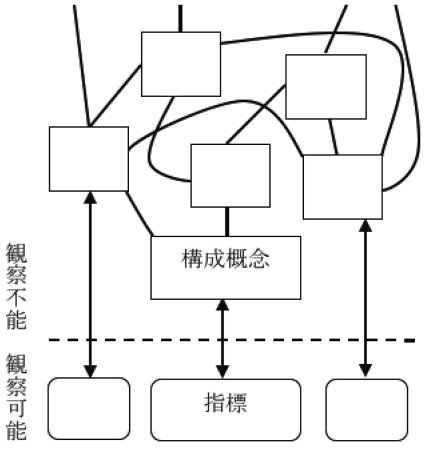
構成概念とは「動機付け」や「記憶」のような複合的な要素を持った概念で, その多くは直接的な観察が困難な概念を指している[11]。このように構成概念は通常的判断で用いられる概念の特徴を過不足なく持っている。

クロンバックらは, 心理測定は構成概念を用いた測定であると考える。構成概念は, 関連する構成概念や構成概念を観察可能なものに置換えた指標で構成された法則ネットワーク（nomological network）の中で意味を持つものである（図3）。

観察の背後に観察不能な理論や法則の存在を仮定する点では, 論理実証主義と近い。しかし両者の間には, 記号論理学で表現された要素還元的な理論文を仮定するのか, あるいは日常的な言語で表現された全体論的な構成概念を仮定するかという大きな違いがある（Scriven, 1986, p.40）。

科学者は, 法則ネットワークを「全面的に理解している。しかしだからといって, それを簡潔かつ網羅的に表現することはできない（強調原文）」(Scriven, 1982, p.24)。例えば「動機付け」という構成概念を単独で取り上げれば, その意味するところは多様である。しかし「動機付けが記憶に正の影響を与える」というように複数の構成概念と関係づけると, その意味は比較的定まる。更に動機付けに関する理論などを付与していけば, 今回の測定における動機付けの定義や指標に何を含めるべきなのか（選択的推論）, またその定義や指標をいかに解釈するべきなのか（導出的推論）について専門家同士の間では共有できるようになり, その妥当性をテストすることが可能となる。

図3　法則ネットワーク



このように構成概念妥当性を検討することで, 科学の中の通常的判断はその定性的蓋然性を向上させ, 担保することが可能であり, ゆえにただの個人的な見解ではない専門的判断として信頼されるのである。

そして評価の実践家や研究者は, 科学の中の価値判断の方法論を積極的に学んでいくことが可能である。構成概念妥当性はその1つである。スクリヴェンによれば構成概念妥当性は「科学的推論の語彙の中核を占めるである・・・それは科学の中の不明瞭で不定形な形而上学ではない。それは豊かで, 柔軟で, 本質的に科学の営為である。そしてこれこそが評価的概念（注：価値前提）の特性を端的に示すものでもある」（Scriven, 1986, p.39）。

（4）評価の中の構成概念妥当性

では蓋然的推論, 特に構成概念妥当性の検討は, 評価活動においてどのような形で現れるのか。ここではスクリヴェンが蓋然論理を説明する際に用いた事例を, 蓋然論理のキー概念を用いて分析し, この問いを検討する。

スクリヴェンは蓋然的推論の説明に際して, 災害救助事業に関する評価を示している。すなわちある国で大規模な地震があり, 数万人の住人が水や食料を必要としている（価値前提）。ここで水や食料を被災地住人に過不足なく水や食料を供給する事業が実施された（事実前提）。我々はこの事業が住人に十分な水や食料を提供したことをもって, 優れた事業だと認めた（評価結論）(Scriven, 2012, p.22)。

この推論では, 大前提である価値前提が求める要件を, 小前提である事実前提は満たしており, このことから事業に高い価値を認めている。この推論は一見するとDNモデルのような演繹的な構造を持つ妥当な推論であるように思われる。しかし蓋然論理の観点からは全く異なる議論が展開される。

蓋然論理は, 推論の文脈性に注意を払う。今回の場合, 地震により衛生環境が低下し, 疫病のリスクが高まっていることなど, 震災時の様々な状況

が想定される。そして, このような状況を踏まえると, 例えば評価対象事業が安易に水や食料を配布したせいで, かえって疫病が蔓延してしまった可能性などが浮上してくる。この可能性が現実のものであれば当該事業を高く評価することはできないため, 事業評価を行う際には疫病への対応状況も価値前提として選択される必要がある。

こうした推論の過程を看過し, 準実験計画法のような精緻な評価手法を採用しても, それは確かに事業の効果の一部を正確に測定することにはなるとしても（定量的蓋然性）, 事業の持つ様々な価値や弊害を明らかにすることにはつながらない（定性的蓋然性）。

このように推論の文脈性への注目は, 評価活動に即して言えば, 事業の副作用に最大限配慮するという事を意味している（Scriven, 2012, p.23）。副作用に配慮するということは, 推論としては「数万人の住人が水や食料を必要としている, また疫病への対策を必要している（・・・）」という通常的判断の形をとった, 価値前提に関する認知的なネットワークを想定するということを意味している。その上で, 自らの知識や経験を活かして, 今回の評価において重要性の高い価値前提を選択し, その意味を十分に解釈していくことである。この過程は「実験的な分析や数学的な分析というより, 経験的で論理的な分析である」（Scriven, 1982, p.27）。

こうした選択的推論・導出的推論が専門的判断の第一歩である。しかし評価者が独自に設定した価値前提に対しては, それが単なる個人的な見解ではないかという疑念がつきまとう。この疑念を払拭するために重要になるのがメタ評価である。ここで言うメタ評価は別の評価者による反対尋問(cross examination；Scriven, 2012, p.23)を理念型としている。すなわち自らの評価をあえて別の評価者による批判的検討に晒しそれをパスすることで, 評価者は自らの下した判断を「客観的で, 論理的で, 有能な専門家でさえも, 合理的な疑いを差し挟むことのできない根拠に裏付けられた, 専門的判断である」（Scriven, 2012, p.23）と言えるようになるのである。このようにメタ評価の方法論的意義の一つには, 評価の構成概念妥当性を検討し, 評価の定性的蓋然性を担保することが含まれるのである。

ただしメタ評価をどれだけ慎重に実施したとしても, 完璧な評価結果が得られることにはならない点に留意が必要である。それはあくまでも現時点で最も確からしい（prima facie）評価に過ぎない。評価の正当性を担保するためには, 専門家間の絶えざる相互検証が必要となる。「蓋然的推論は反復的あるいは潜在的に反復的な性質を持っている」（Scriven, 1991, p.220）のである。

6．結論

(1) 知見の要約と本稿の意義

本稿では蓋然論理に関する科学哲学的ルーツにまで遡り, それがどのような理路を通して評価方法論へとつながるのかを希求した。

これによりスクリヴェンが, 論理実証主義との議論を通して「科学的」とは一体どのようなことなのか問いなおす中で, 蓋然論理ひいては評価論理が構築されてきた事が明らかになった。科学は観察のような認知的過程や実験のような特定の手法には還元できない。科学には構成概念妥当性を検討する推論の過程が含まれ, この過程は専門家間の仮想的・実在的な対話という社会的相互作用の過程を内包しているのである。これを踏まえると, 評価の正当性を担保するためには, 社会調査の手法を応用して厳密に事実前提を特定するのみならず（定量的蓋然性の担保）, 構成概念妥当性を検討して妥当な価値前提を特定すること（定性的蓋然性の担保）が重要である。構成概念妥当性は専門家同士の反対尋問を通して進められるメタ評価として実践される。

本稿の意義は大きく2つあると言えよう。1つめはスクリヴェンの思考の結果のみならず, 思考過程の一端を描き出したことである。このような理路はスクリヴェン自身による説明でも必ずしも明確になっていなかった部分であり, また評価論理を批判的に検討して更に発展させていく手がかりとなる[12]。

2つめは, 評価論理の方法論的含意をより明確な形で提起したことである。本稿は蓋然論理に係る諸概念を再発見した上で精緻化した。特にスクリ

ヴェンがその重要性を示唆するにとどめていた構成概念妥当性について考察を深め, その観点から価値判断ならびに評価の正当性を理論的に説明した。

(2) 残された課題と展望

本稿の考察を踏まえて, 残された課題を2つ指摘したい。1つめは蓋然論理の知見を踏まえて, スクリヴェンの評価方法論全体の体系性を改めて精査するという課題である。蓋然論理はスクリヴェンの評価論理の基礎をなす科学哲学的な認識論および方法論で, スクリヴェンが提案してきた評価手法と密接な関係にある。例えば評価対象の副作用に焦点化したゴールフリー評価はとりわけ蓋然論理と関係が深いと考えられる。しかし本稿では, メタ評価については蓋然論理の観点から議論したが, それ以外のものについては扱っていない。この課題は, スクリヴェンの評価方法論に関する全体的な解説や評価の理論史的検討として改めて論じられる必要がある。

2つめの課題は, 専門性重視の評価における推論の過程を分析し, 改善していくという課題である。研究評価におけるピアレビューを例にとって考えてみよう。ピアレビューは, 論文や研究計画の価値を, その論文や研究計画の著者と同程度の専門性を持つものに判断させる手法である。一般的に, 科学論ではピアレビューは文脈から独立した認知的過程であるべきだと考えられてきた（西村2013）。この発想に基づくと, ピアレビューの過程で査読者同士が評価について協議するような行為は, 評価結果に学閥等のバイアスをもたらすと考えられる。しかし蓋然論理に基づくと, こうした過程はむしろ評価の定性的蓋然性を向上させる可能性を秘めていると理解される。このように蓋然論理は専門的判断が伴う社会的相互作用を発見し, その過程に評価方法論上の効用があることを示唆するとともに, 社会的相互作用のあり方を改善していくものである。専門性重視の評価における社会的相互作用について, 経験的・実践的に検討することが, 今後の課題として残されている。

注記

1 前提とは論理学の用語で命題・文のことである。

2 佐々木（2010）以外の解説は, フーニエに翻案されたスクリヴェンの評価観（Fournier 1995）に影響を受けている。ただしこの理解は, フーニエによるスクリヴェン解釈を否定するものではない。

3 なお分析的概念と総合的概念の二元論は, 論理実証主義の基本的な原理として知られる「意味の検証理論（verification principle）」の解説としても用いられる。パトナム（2006）参照。

4 DNモデルの説明については, ブラウン（2001）も合わせて参考にした。特に日食を例にとるアイデアはブラウンに拠っている。

5 “probative”の法学的文脈における意味の解釈については, Gale and Lehman（2004）を参考にした。

6 裁判における証拠力の認定の例示は, スクリヴェンによる不十分性・煩雑性・不安定性の説明（Scriven, 1987, pp.23-26）を参照している。

7 通常的判断に関する解釈は, スクリヴェン自身の議論のほか, Leibowitz（2011）, Schurz（1995）, Taye（2011）も一部参考にした。

8 導出的・選択的推論の原文はderivation explanation, selective explanation で, 直訳すれば導出的・選択的説明となる。しかしこの用法は, 現象の因果的効果に関する言明を説明と呼ぶ科学方法論の慣行から外れるものであり, また説明の中に説明が含まれるという入れ子構造は概念の混乱を生む。このことから, 本稿では導出的推論・選択的推論という訳語を与えた。

9 DNモデルの他にも, 通常的判断および自明の理は法則を用いない説明いわゆる経験則を用いた説明においても現れることが指摘されている（Scriven 1959a；b）。

10 後期のカルナップは「対応規則」という理論文から観察文を一対一関係で導出するための規則が存在すると主張する。しかしそのような規則の存在が説得的に論証されているとは言えないし, 何より, 対応規則のような検証不能なものを科学から排除することが論理実証主義の初発の問題意識であったはずである。対応規則についてはカルナップ（2003, pp.202-205）参照。

11 構成概念妥当性に関する解釈は, 村山（2012）も参

考にしている。

12　簡略化して言えば, 本稿は佐々木（2010）の言うスクリヴェン（およびハウス）の依拠する「現代哲学における議論」（p.44）を考察したものとも言える。この意味では本稿の理論的考察は佐々木（2010）の実践的提言と相補的な位置にある。

参考文献

カルナップ著・永井 成男・内田 種臣編（2003）『カルナップ哲学論集』、紀伊國屋書店

小西卓三（2006）「非形式論理学とディベート」日本コミュニケーション学会『スピーチコミュニケーション教育』、23-44

佐々木亮（2010）『評価論理―評価学の基礎―』、多賀出版

西村君平（2013）「研究評価の研究動向」広島大学 高等教育研究開発センター『大学論集』、227-242

村山航（2012）「妥当性―概念の歴史的変遷と心理測定学的観点からの考察」、『教育心理学年報』、51：118-130

渡辺博（1987）「蓋然性」、廣松渉・子安宣邦・三島憲一・宮本久雄・佐々木力・野家啓一・末木文美士編『哲学・思想辞典』岩波書店、207

Alkin, M. C. (Ed.). (2004) . *Evaluation Roots.* Sage Publication.

Brown, J. R. (2001). *Who Rules in Science.* Harvard University Press.（青木薫訳『なぜ科学を語ってすれ違うのか』みすず書房、2010年）

Carnap, R. (1936). Testability and Meaning. *Philosophy of Science,* 3, 419-471.

Cronbach, L, J. and Meehl, P. E. (1955). Construct Validity in Psychological tests. *Psychological Bulletin*, 52, 281-302.

Fournier, D. M. (1995). Establishing Evaluative Conclusions: a Distinction between General and Woking Logic. *New Directions for Evaluation*, 68, 15-32.

Gale, T. and Lehman, J. (2004). (eds). *West's Encyclopedia of American Law (2nd edition)*, Detroit : Thomson Gale.

Hempel, C. G. (1970). *Aspects of Scientific Explanation: and other essays in the philosophy of Science.* NY : Free Press.（長坂源一郎訳（1973）科学的説明の諸問題、岩波書店）

Hempel, C. G. and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15, 135-175.

Johnson, R. H. and Blair, J. A. (2000). *Informal Logic* 20, 2, 93-107.

Leibowitz, U. D. (2011). Scientific Explanation and Moral Explanation.*Noûs* 45 3, 472-503.

Migotsky, C., Stake, R., Davis, R., Williams, B., DePaul, G., Cisneros, E. J., Johnson, E. and Feltovich, J. (1997).Probative, Dialectic, and Moral Reasoning in Program Evaluation. *Qualitative Inquiry*, 3, 453-467.

Patton, M. Q. (2012). Contextual Pragmatics of Valuing. In G. Julnes (Ed.) *Promoting Valuation in the Public Interest: Informing Policies for Juding Value in Evaluation. New Directions for Evaluation*, 133, 97-108.

Putnum, H. (2002). *Collapse of the Fact and Value Dichotomy.* Harvard University Press.（藤田晋吾・中村正利訳（2006）『事実／価値二分法の崩壊』法政大学出版局。）

Russel, B.(1956). The Philosophy of Logical Atomism and Other Essays 1914-19, George Allen and Unwin Ltd.（高村夏輝訳（2007）『論理的原子論の哲学』筑摩書房）

Schurz, G. (1995). The Role of Negation in Nonmonotonic Logic and Defensible Reasoning. *Prepublication Series at the Departmet of Philosophy,* University of Salzburg, 1-44.

Shadish, W. R., Cook, T.D. and Leviton, L. C. (1993). *Foundations of Program Evaluation: Theories of Practice.* Sage Publication.

Scriven, M. (1959a). Explanation and Prediction in Evolutionary Theory. *Science* 130, 477-482.

Scriven, M. (1959b). The Logic of Criteria. In American Philosophical Association, *The Journal of Philosophy*, 56, 857-868.

Scriven, M. (1959c). Truism as the Grounds for Historical Explanations. In P. Gardiner (ed). *Theories of History.* Oxford Press.

Scriven, M. (1972). The Exact Role of Value Judgement in Science. In *Proceedings of the Binnial Meeting of the Philosophy of Science Association*, 219-247.

Scriven, M. (1982). *The Logic of Evaluation*, EdgePress.

Scriven, M.(1986). New Frontiers of Evaluation. In *American Journal of Evaluation*, 7, 7-44.

Scriven, M. (1987). Probative logic. F.H. van Eemeren, R.

Grootendorst, J.A. Blair & C.A. Willard. (Eds.), *Argumentation: Across the lines of discipline,* Foris Pubns, 7-32.

Scriven, M. (1991). *Evaluation Thesaurus (4th edition).* Sage Publication.

Scriven, M. (2005). Logic of Evaluation. Marthison, S. (Eds.), *Encyclopedia of Evaluation.* Sage Publication, 235-238.

Scriven, M. (2007). The Logic of Evaluation. Exact Role of Value Judgement in Science. Hansen, H. V. et. al., (Eds), Dissensus and Search for Common Ground, Windsor, 1-16

Scriven, M. (2012). The logic of valuing. G. Julnes (Ed.), Promoting Valuation in the Public Interest: Informing Policies for Juding Value in Evaluation. *New Directions for Evaluation*, 133, 17-28.

Taye, O. R. (2011). Normic Statement, and Normic Network: Scriven's Alternative Explanation for Histrical Explanation. *IUB Journal of Social Science and Humanities*, The Islamia University of Bahawalpur.

（2014.9.5 受理）

Probative Logic: its evaluation methodological implication

Kunpei Nishimura

Hiroshima University, Research Institute for Higher Education
k-nishimura@hiroshima-u.ac.jp

Abstract

The purpose of this study is to theoretically review “probative logic” by focusing on its philosophical background and its implication for evaluation methodology. Probative logic is a theory to analyze and improve judgment aspects of evaluation. This logic is developed by Michael Scriven and is one of the keywords of his "logic of evaluation". Thus, understanding probative logic lead better comprehension of logic of evaluation and meteorological implication about “judgment driven evaluation” such as research evaluation.

In this paper I will present three topics concerning this theme. First, I discuss the relationship between probative logic and “value free doctrine” to clarify its philosophical roots. Second, I illustrate some features of judgment from a perspective of probative logic. Third, I show some models of the probative logic as alternative to the value free doctrine. To conclude, I will discuss implication for evaluation methodology and limitation of this paper.

Keywords

probative logic, logic of evaluation, value free doctrine, probability, construct validity

【Article】

Rethinking the Human Development Index: A Path toward Comprehensive Development Evaluation

Aki Yonehara
Meiji University
yonehara@meiji.ac.jp

Abstract

The Human Development Index (HDI) is one of the most internationally recognized development indices. Unlike the conventional economic indices, the HDI considers not only economic but also educational and health aspects. The HDI is often used as a ranking index to assess the level of development of a nation. However, the theoretical foundation of the HDI, human development theory, does not aim at proposing a ranking index; rather, it provides a perspective for comprehensive evaluation to clarify the problems that each nation needs to tackle for its national development.

Therefore, reflecting the original idea of human development theory, this paper tries to develop a new method of development evaluation to assess the balance of development of each nation, rather than ranking them. 141 nations which have the necessary data for statistical analyses are sampled, and principal component analysis is used to decompose each of the 141 HDIs according to education, health and economic factors. The decomposed factors draw a chart of development balance and these 141 charts are categorized and utilized for a more comprehensive evaluation of development.

Keywords

Human Development Index, human development theory, development evaluation,
charts of development balance, principal component analysis

1. Introduction

Facing with the targeted date for achievement of the United Nations Millennium Development Goals (MDGs) in 2015 (World Bank 2010), global interest in the evaluation of development has been growing, as witnessed by the establishment of the MDG Achievement Fund (MDG-F) in 2007. Since 1990, the Human Development Index (HDI) has received heightened attention as a composite index that is used internationally as a comprehensive development index in place of conventional economic development indices. Unlike conventional indices, the HDI includes a health index (life expectancy at birth) and educational indices (mean years of schooling and expected years of schooling) in addition to an economic index (GNI per capita).[1] The HDI is calculated as a single index value for each

Japanese Journal of Evaluation Studies, Vol.14, No.1, 2014, pp.31-46 【Revised based on the translation from Japanese Journal of Evaluation Studies Vol.12, No.3, pp.91-105(Japanese Version).】

country based on individual country data and is used to measure its development level. This new index is applied in various fields, in both theoretical and practical ways, to provide a new perspective for international comparisons. In particular, since the 2000s, the HDI has gained influence as a political tool in policy responses in developing countries. For example, there was a case in Singapore where HDI ranking affected government policy (UNDP 2010, p. 14).

The HDI is highly regarded as a development index that takes a comprehensive perspective. Sen's human development theory (1981; 1987; 1992; 1997; 1999a; 1999b; 2000; 2002), which forms the theoretical basis for the HDI, is highly respected as well. With such worldwide support, the HDI has been utilized in many leading academic research projects and in major reports by international organizations. Although the HDI enables the comparison of countries using a more comprehensive index than conventional economic indices, it is difficult to say, however, whether using the HDI as a ranking index is the most appropriate application of it in view of human development theory. Esho (1997, p. 216) articulates this point as follows:

> Ranking countries by the HDI is not the best way to take advantage of Sen's idea. ... Instead, we must analyze specifically what capabilities are insufficient at the level of nation, region, social class, and gender, in order to investigate the reasons for those problems. This is a way to open the path toward effective development policies. *(translation by author from the Japanese)*

This paper, therefore, attempts to propose a new method of evaluation of the development balance with reconsidering the meaning of human development theory. The proposed method does not simply aim at country rankings but utilizes the HDI to achieve a more comprehensive evaluation of development. To achieve this goal, this paper is structured as follows. First, after clarifying the definition and characteristics of the HDI, the problems with its current applications are identified (section 2). Second, charts portraying the development balance are shown based on the component matrices obtained by disaggregating the HDI data of 141 countries using principal component analysis (PCA) (section 3). Finally, through the categorization of these 141 charts of development balance, the issues of mutual complementarity with the conventional HDI ranking information are discussed (section 4).

2. The HDI: Limitations and issues

2.1 What is the HDI?

In the field of international development in the postwar period, the meaning of development has been understood to be almost equal to economic development. According to Meier's definition (2001), the history of international development, which is considered to have started with the Colombo Plan for Cooperative Economic and Social Development in Asia and the Pacific in 1950, can be divided into the first generation from 1950 to 1975 and the second generation from 1975 to 2000. While the first generation of development adopted a development strategy using macroeconomic development theories[2], the second generation promoted a development strategy using a microeconomic development approach based on neoclassical economic theories[3]. Although there are differences between the macro and micro approaches, both the first and second generations understand the meaning of development as economic development. In contrast, a new concept of development, known as the third generation, is the human development theory proposed by Sen (1981; 1987; 1992; 1997; 1999a; 1999b; 2000; 2002), the 1998 laureate of the Nobel Prize in Economics. According to Sen (2000, p. 3) :

Development can be seen, it is argued here, as a process of expanding the real freedoms that people enjoy. Focusing on human freedoms contrasts with narrower views of development, such as identifying development with the growth of gross national product, or with the rise in personal incomes, or with industrialization, or with technological advance, or with social modernization.

Criticizing such "narrower views of development," Sen provides a more comprehensive view of development (Yonehara 2009, pp. 35-52). According to Sen, the ultimate purpose of development is "constitutive freedom" — that is, the freedom to actually realize the life which the individual wants (Sen 1992; 1999a; 2002)—while economic growth or economic wealth is considered as a means for this ultimate purpose of development. From this perspective, conventional approaches that measure the development level of a country solely in terms of GNP, GDP, GNI, or other economic indices are insufficient. The United Nations Development Programme (UNDP) thus came up with the more comprehensive index, HDI. Adopting Sen's theory, the HDI has been used in the annual *Human Development Report (HDR)* since the first issue in 1990. In *HDR*, international rankings and classifications (low, medium, high, and very high HDIs[4]) are reported.

A detailed definition of the HDI is given in the HDR technical notes (UNDP 2010, pp. 215-219).[5] In short, though, the HDI is a composite index calculated as a geometric mean of three indices—life expectancy, education, and GNI—that are compiled from four indicators—life expectancy at birth, mean years of schooling, expected years of schooling, and GNI per capita (see table 1).

Table 1 Definition of the Human Development Index

	Human Development Index = Geometric mean of the three indices		
Dimension	Long and healthy life	Knowledge	Decent standard of living
Index	Life expectancy	Education	GNI
Indicator	Life expectancy at birth	Mean years of schooling, Expected years of schooling	GNI per capita (PPP US$)

Source: Created by author based on UNDP (2010).

While questioning the validity of defining the HDI using only four indicators, Jolly (2002, p. 266) acknowledges the significance of the HDI as a statistical index employed in the context of human development and notes that "statisticians must abandon the common misconception that statistics have little to contribute [to human rights] ... statistics can play a critical role in raising awareness and generating pressure for action."

2. 2 Applications of the HDI

First announced in 1990, the HDI has since become a widely used tool for providing a comprehensive view of development.[6] For example, Haq (2000) uses the HDI to conduct a comparative study of the human development level in South Asian countries. Ahsan (2004) also performs an analysis of human development conditions in Islamic countries using HDI comparisons. In addition to these comparative studies using the HDI ranking, there are different kinds of applications of the HDI seen in the *HDR National Reports*. Published by each UNDP regional office, these reports mainly use the information of the three HDI indices to analyze the particular human development issues in each country (UNDP 2012a). Furthermore, the UNDP provides an online database that enables us to understand the

trends in, and the current status of, the various indices for each country, by generating graphs and charts on demand (UNDP 2012b). In another kind of study using the HDI, it is reported that the HDI for Costa Rica, Cuba, and Sri Lanka are higher than those for other countries with economic development conditions similar to these three (UNDP 2010, pp. 49-50), thereby proving that economic development and human development do not necessarily coincide.

As seen above, the introduction of the HDI has spurred debate on human development from a more comprehensive perspective, one that incorporates multiple aspects of human life, not merely the economic aspect. However, in view of the current situation where the HDI is often used as a ranking index for international comparisons, it does not seem that the problems noted in the earlier citation from Esho (1997) have been resolved. As mentioned above, there are some studies that seek to analyze human development issues in individual countries rather than to rank countries, but methods to analyze the balance of development in each country have not yet been introduced. In other words, methods have yet to be invented for evaluating the comprehensiveness and uniqueness of development in each country.

Furthermore, many of the criticisms directed at the HDI relate to its technical precision as an index, such as the selection and calculation of the variables that compose the indicators. In response to these criticisms, the UNDP has proposed various development indices including the Human Poverty Index (HPI), the Gender Inequality Index (GII), and the Multidimensional Poverty Index (MPI). A series of technical efforts to improve the HDI itself also have been made (Panigrahi and Sivramkrishna 2000; Herrero et al. 2010). However, these efforts do not change the fundamental fact that all these new indices keep the same purpose of making a raking evaluation by using a single value.

The ultimate purpose of Sen's human development theory is not to rank countries but to raise awareness of the multidimensional and complex nature of development and to open the way forward to create better development policies from a more comprehensive perspective.[7] Therefore, this paper will look into new possibilities for the HDI beyond a ranking index and try to propose a new development evaluation method that reflects the essence of human development theory.

3. From ranking to development balance

3.1 Charts of development balance

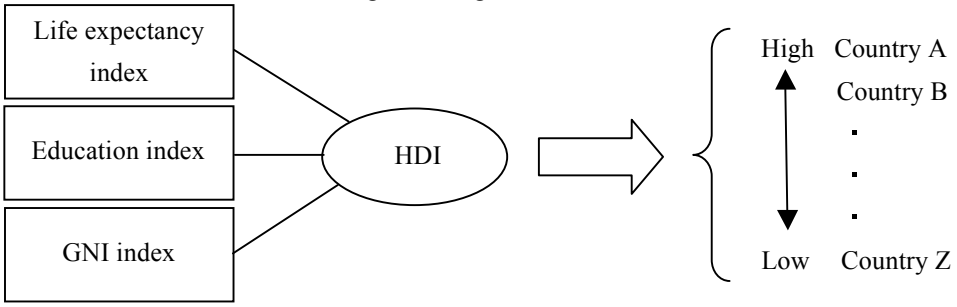
As shown above, the conventional HDI is often used to rank individual countries based on their HDI value which is calculated from the three indices (see fig. 1). In this paper, the balance among the indices is used to offer a visualization of imbalances in human development policies and thereby clarify the issues of development policies in individual countries (see fig. 2).[8] Through this model that expresses the balance/imbalance in development from multiple angles, the concept of a "comprehensive view of development" established by Sen can be better embodied as an evaluation method for human development.[9]

Hidalgo (2010) and Despotis (2005) also consider the balance among the indices. Hidalgo (2000) proposes a visual representation of the HDI by analogizing the HDI of each country as a tree, where the value of the HDI is represented by the size of the trunk while the three indices are each represented by branches extending from the main trunk. The HDI Tree can be intuitively understood by general readers who may not be familiar with statistics. However, as detailed later in this paper, this method is not capable of classification for more in-depth analyses. While possessing similar characteristics to the HDI Tree, the chart of development balance proposed in this paper offers the additional advantage of enabling classification.

The data envelopment analysis (DEA) of Despotis (2005) takes the comprehensiveness of the HDI into serious consideration and proposes a new model of international comparisons based on the efficiency defined by the three

Figure 1 Conventional Concept of the HDI

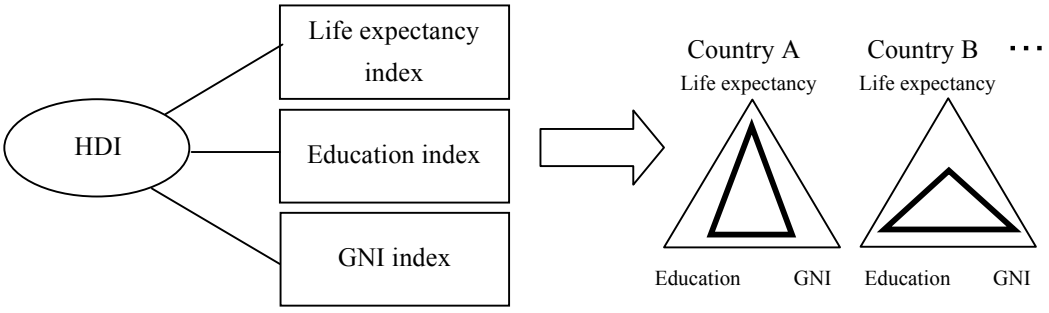
Calculates a single ranking index from the three indices



Source: Created by author.

Figure 2 Possible Applications of the HDI

Analyzes the balance of the three indices



Source: Created by author.

indices, rather than a simple ranking. Despotis's research addresses the same issue as this paper: it questions the use of the HDI as a simple ranking index. However, the DEA framework is originated in the study of business administration where it seeks to maximize investment efficiency from multifaceted viewpoints (Tone 1993). As Despotis himself points out, his research therefore takes the approach of reevaluating the HDI from an efficiency perspective. It thus ends in a different focus from the evaluation of balance among the indices sought by this research.

As described in the next section, this paper proposes a method to comprehensively evaluate the balance among the three HDI indices using principal component analysis (PCA). This method provides a platform for the discussion of human development policy issues, including which areas deserve priority. Classifying the charts of development balance, as described later, facilitates international comparisons from a different viewpoint from rankings. The originality of this research lies in these points.

3.2 Methods

To clarify the structure of the HDI or the balance among the three indices, PCA is conducted (Tabachnick and Fidell 2001; Grimm and Yarnold 2004). Through PCA, the principal component that constitutes the HDI of a country (or the first principal component, i.e., the eigen vector that expresses the major trend of human development for that country) is extracted. Then the levels of contribution of the three indices to the first principal component (or their impact on the first principal component, i.e., the contribution levels of each index to the human development vector

for that country) are derived as a component matrix. The development balance is evaluated based on the three components in the component matrix.

In this analysis, the component matrix obtained from PCA is used to examine the balance of the three indices; the raw values of each index are not used directly. This is because the main purpose of this analysis is not to compare the magnitude of the respective indices individually, but to examine the balance of the three indices as influences on human development in each country. The component matrix derived by PCA shows the correlations of the three indices with the first principal component. In other words, it can answer the question of how much each component (in this case, the three individual indices) contributes to the first principal component (in this analysis, the eigen vector showing the major trend of human development for each country). The difference between the charts of development balance created with the raw values of each index and the charts created with the component matrix obtained from PCA is shown in the next section.

The data used in the analyses is collected from 141 countries (18 countries from the low human development group, 59 from the medium human development group, 31 from the high human development group, and 33 from the very high human development group),[10] which have adequate data for the analyses, out of 177 countries in the ranking list of *HDR 2007/2008* (UNDP 2007). Using the panel data on the three indices for each country (1970-2010: including estimated values), PCA is conducted for each of the 141 countries.

When interpreting the results of the analyses, it is necessary to pay attention to the fact that this analysis cannot be a tool for relative evaluation. In other words, even when all three components in the component matrix show high values (i.e., the development balance is satisfactory), it does not directly mean that the original raw values of these indices are high as well. Even when the development balance looks good, the individual index values of life expectancy, education, and GDP may be low (the case of "low-ranking balanced"); conversely, the relative HDI ranking may be high for countries with an unsatisfactory development balance (the case of "high-ranking imbalanced"). Therefore, the model of development balance obtained from this analysis is not a replacement for the conventional HDI, but is a complement to the HDI ranking information. As will be discussed later, combining the conventional HDI ranking with the idea of development balance improves the potential of the evaluation method for a more substantive and comprehensive development evaluation.

3.3 Results

Although it is not possible to provide the detailed results of all 141 countries due to space constraints, the results for three countries are shown below as typical examples, one each from Latin America, Central Asia, and East Africa (see table 2). The values of the component matrices express the degree of contribution from each component (LEX, EDU, GDP) to the first principal component or the eigen vector that expresses the trend of human development of that country. The index values within parentheses in the second line in each cell are the average of the raw values of each index during the covered period (1970-2010).

In the outputs of PCA, the powers of explanation of the first principal components exceed 70-90% for each country, and thus the respective first principal components can be interpreted as the major trend of human development in each country. The balances of HDI components for these three countries are shown in fig. 3 as radar charts.

These radar charts represent the balances and levels of contribution of the three indices to the human development trend in each country (or the eigen vector expressed as the first principal component). Such modeling enables to examine how much balance is achieved by the HDI components in each country, beyond just looking at a single index value for HDI ranking. Ideally, in addition to a high HDI value, it is preferable that the three components of HDI contribute to the human development vector in a well-balanced manner. When a significant imbalance is

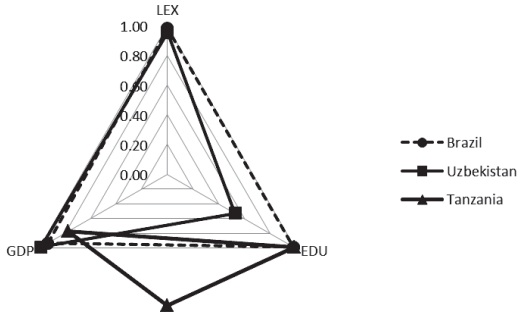
Table 2 Results of the Analyses: HDI Component Matrices

	Latin America	Central Asia	Eastern Africa
	Brazil	Uzbekistan	Tanzania
LEX[Note2] (LEX index ave.)	0.984 (0.766)	0.957 (0.680)	-0.886 (0.439)
EDU[Note2] (EDU index ave.)	0.983 (0.630)	0.528 (0.463)	0.99 (0.288)
GDP[Note2] (GDP index ave.)	0.932 (0.840)	0.985 (0.910)	0.774 (0.560)
% of Variance[Note3]	93.447	72.14	78.764

Notes: 1. Analyzed by SPSS ver. 11.5.
2. LEX = Life expectancy at birth; EDU = Education; GDP = Gross domestic product
3. "% of Variance" shows the power of explanation by the first principal component.
Source: Created by author.

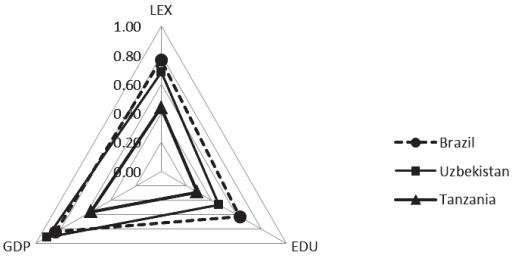
observed among the components, it indicates that the human development policies of that country require improvement. Radar charts of the same three countries directly using raw average values of the three indices are shown in fig. 4.

Figure 3 Results: HDI Balance Chart



Source: Created by author.

Figure 4 Results: HDI Raw Average Chart



Source: Created by author.

Compared to fig. 3, the charts of the three countries in fig. 4 look similar to one another and do not clearly present the characteristics of each country. Each average value in fig. 4 expresses the relative position between 0.00 and 1.00 that has been standardized by the given goalpost values (UNDP 2010, pp. 215-219). Because this standardized value possesses the characteristic like a deviant value, the countries with similar raw data and variance tend to have similar index values, and as a result, similar charts are drawn with these indices. In addition, the charts in fig. 4 show the average values for each index independently, so that the higher-level concept of human development that subsumes these three indices does not appear anywhere.

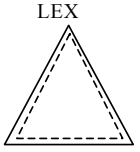
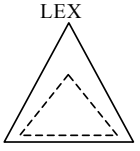
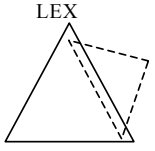
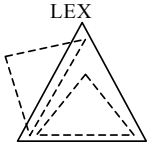
In contrast, the index values in fig. 3 express the relationships between each index and the eigen vector for human development or the higher-level concept of human development (i.e., the first principal component derived from PCA). Instead of the magnitude of the raw data for each index in fig. 4, the charts in fig. 3 show how the three

indices contribute to the first principal component, or human development, of that country, as well as the strength of their contribution toward the human development trend of that country.

In fig. 3, Brazil appears to have well-balanced human development. On the other hand, imbalances can be seen in the radar charts for Uzbekistan and Tanzania. In Uzbekistan, development imbalance can be seen in the education index, or the human development dimension of knowledge, while in Tanzania, a development imbalance can be seen in the life expectancy index, or the human development dimension of long and healthy life. When exploring human development policies, policymakers in these countries may need to take the existence of such imbalances into account.

Although imbalances need to be understood as signs that a certain problem exists, even when imbalances are discovered, there is also the possibility that no action needs to be taken beyond recognizing the fact that an imbalance exists within acceptable limits, if it is deemed to have low priority in that society. The important thing is that policymakers and the general public share a common awareness that an imbalance exists in a certain area and that public debate is held to seek for an appropriate policy response.[11] The role of the balance chart is to provide such a clue for awareness.

Table 3 Four Categories of Charts of Development Balance

Category	Definition	Interpretation	Shape of charts
I	**Equilateral triangle**: The levels of contribution of all three indices are above 0.800.[Note1]	A well-balanced development with regard to the three HDI dimensions, namely *long & healthy life*, *knowledge* and *a decent standard of living*. However, note that this does not guarantee that the raw data of each dimension is sufficiently high.	
II	**Isosceles triangle**: Although one of the indices is less than 0.800, the shape is maintained within the chart.	The contribution from one of the three indices to human development in that country is low. That is, imbalance of human development is observed in a certain dimension.	
III	**Deviant triangle**: Negative correlation is observed, and the triangle deviates outside of the chart.	One index is negatively correlated with the human development vector in that country. That is, a certain dimension has deteriorated during the covered data period (1970 to 2012).	
IV	**Plural triangles**: Two principal components are derived, with characteristics of both categories II and III.	There are two principal factors to explain the human development vector of that country. Having two or more principal components means that it contains the problems of categories II and/or III since it cannot shape an equilateral triangle for a statistical reason.	

Notes: 1. The UNDP HDI ranking defines "high human development" as above 0.800.
2. In *HDR* prior to the 2010 edition, GDP (per capita, PPP in US$) is used instead of GNI.
Source: Created by author.

4. Inquiry of charts of development balance

4.1 Four categories of balance charts

As described above, 141 countries have been analyzed and 141 balance charts produced. These 141 charts are classified into the four categories as below (see also the appendix).

Categories II and III are subclassified. Category II is subclassified into those with an EDU index of below 0.800 as II-1 (education-imbalanced cases), those with a LEX below 0.800 as II-2 (health-imbalanced cases), and those with a GDP below 0.800 as II-3 (economic-imbalanced cases); and Category III is also subclassified into those with a negative EDU as III-1 (education-deviated cases), those with a negative LEX as III-2 (health-deviated cases), and those with a negative GDP as III-3 (economic-deviated cases). Those with both negative indices and indices of below 0.800 are subclassified as III-4 (mixed cases) (see table 4 and the appendix). The results of the analyses are organized according to the categories above and summarized in table 4 with the HDI ranking information.

Table 4 Cross-Reference Table between HDI Balance and HDI Ranking

Categories / Ranking	I (Balanced)	II (Imbalanced)			III (Deviated)				IV (Plural)
		II-1	II-2	II-3	III-1	III-2	III-3	III-4	
Low	BEN, ETH, GMB		BFA	GNB, MLI	BDI, CAF, CIV, ZMB	MOZ	ERI, NER	MWI, RWA, TGO	TCD, SEN
Medium	BGD, CPV, CHN, DOM, EGY, FJI, GTM, GUY, IND, IDN, LAO, MAR, NPL, PAK, WSM, LKA, SDN, SYR, THA, TUN, VNM, YEM	KHM, MNG, UZB	ARM, BLZ, BWA, TJK	DZA, BOL, SLV, GHA, HND, IRN, MRT, PNG, PRY, PHL	HTI, MDA, UKR	COG, GVQ, LSO, SWZ, UGA	COM, DJI, JOR, MDG, NIC	CMR, NAM, TZA	JAM, KEN, NGA, ZAF
High	ALB, BRA, CHL, COL, CRI, ECU, EST, HUN, LBN, MYS, MUS, MEX, OMN, PAN, TUR, URY	BHS		ARG, BHR, BGR, HRV, LVA, LTU, ROM			PER, SAU, VEN		BLR, KAZ, RUS, TTO
Very High	GRC, HKG, ISL,IRL,ISR, ITA, JPN, KOR, LUX, MLT, NLD, NZL, NOR, PRT, SGP, SVN, ESP. SWE, CHE, GBR, USA						KWT, ARE		

Categories / Ranking	I	II			III				IV
		II-1	II-2	II-3	III-1	III-2	III-3	III-4	
Low	Low-ranking balanced	Education-imbalanced	Health-imbalanced	Economic-imbalanced	Education-deviated	Health-deviated	Economic-deviated	Mixed	Plural triangles
Medium									
High	Well-balanced								
Very High									

Note: Abbreviations of country names are based on *HDR* (UNDP 2010). Refer to the appendix for full names.
Source: Created by author.

The cells *I-Low* and *I-Medium* in table 4 include countries of "low-ranking balanced," while the I-High and *I-Very-High* cells include countries of "well-balanced" (see the lower section of table 4). Countries in the very high human development group in the HDI are almost all classified under the "well-balanced," and two exceptions, Kuwait (KWT) and United Arab Emirates (ARE), are classified as economic-deviated cases (III-3).

Looking at the low human development group in the HDI, more than half of the countries are classified as deviated cases (10 out of 18, or about 56%). Compared to those in the medium human development group (16 out of 59, or about 27%) and the high human development group (3 out of 31, or about 10%), this proportion is clearly higher. The deviated cases indicate that a certain kind of aggravation in one of the three dimensions has occurred in the past 40 years. The subcategories (III-1 to -4) in table 4 articulate where deviations have occurred in those countries. Policy responses are called for in these deviated areas.

Many countries in the medium human development group in the HDI show signs of imbalance (17 out of 59 countries, or about 29%). Considering the fact that 16 of those 59 countries (about 27%) are classified as deviated, more than half of the countries in the medium human development group suffer from instabilities (imbalances and/or deviations) in human development. Here again, the subcategories of II (II-1 to -3) and III (III-1 to -4) in table 4 clarify the dimension of instability. Policy responses for these instabilities are expected.

Although many countries in the high human development group are classified as well-balanced, it can be seen that the majority of those categorized as unstable (imbalanced and/or deviated) show economic instability (II-3 and III-3).

4.2 Mutual complementarity with HDI rankings

Based on the analyses above, focusing on balance in addition to ranking clarifies the human development issues in each country, which cannot be identified by using ranking information alone. The charts of development balance also capture visually national diversity within the same ranking group, which tends to be overlooked when using ranking information alone. However, as already discussed, the chart of development balance does not aim to refute the significance of conventional HDI rankings but instead aims to strengthen the mutual complementarity between balance and ranking.

It has been pointed out already that balance charts cannot be used as tools for relative evaluations. In other words, even if the development balance of one country looks satisfactory, the concurrent use of HDI ranking information is useful to avoid misinterpretations of countries that have balanced development but are in the low human development group, where the raw values of the indices are low. At the same time, the evaluation by balance charts serves as a means of complementing the ranking information. For example, even when a country is listed in the high human development group, it might be necessary to rethink critically whether the development balance is also good (well-balanced) or there is a problem in the development balance (imbalanced/deviated).

4.3 Limitations and issues of balance charts

In order to improve the balance chart, several issues remain to be addressed. First, to ensure mutual complementarity with the HDI, this paper had to limit the indices to only three, and it resulted in restricting the comprehensiveness of the evaluation model. To build a more comprehensive evaluation model to address Esho's expectation and Sen's assumption, the analysis needs to be conducted using more indices in addition to the three dimensions of the HDI.

Next, it is necessary to reconsider the standard for interpretation of the radar charts. There is still a question regarding how to determine whether a case is well-balanced or imbalanced and how to interpret the degree of imbalance or deviation when evaluating the charts of development balance. Although this paper adopted a provisional HDI value of 0.800, it is necessary to generalize the standards for evaluation with a theoretical backing.

Finally, while this paper categorizes the shapes of the charts into four basic categories (I to IV), further consideration is required whether this categorization is sufficiently rational. In particular, when the chart shape becomes a polygon of a higher order than a triangle, the definition of unstable cases (imbalanced/deviated cases) will become even more complex. More convincing categorization may support more relevant analyses to produce more substantive suggestions.

5. Conclusion

This paper gives a visual form to the three HDI components using a radar chart and proposes an evaluation model of human development from the perspective of development balance. While this paper has not been able to include all the factors listed by Esho (1997), the evaluation method suggested here provides a perspective for a comprehensive evaluation that has not been possible with ranking-based evaluation methods. The trial in this paper reveals that evaluation methods of this kind can help to visually express the adequacy of the development balance in individual countries, something which has been overlooked by using the HDI rankings alone. These evaluation models may stimulate policymakers' interest in formulating better-balanced human development policies.

Evaluation methods that consider balance among the various factors can be effective tools in comprehensive evaluations, such as the evaluation of MDGs or SDGs (Sustainable Development Goals). As comprehensive approaches are becoming the mainstream in international development today, it is an urgent task to develop and improve multidimensional evaluation models.

Appendix: Mutual Complementarity of HDI Ranking and HDI Balance

COUNTRY	CODE	HDI*1	REGION*2	CATEGORY*3	COMPONENT_1*4				COMPONENT_2*4			
					%var1*5	comp1EDU	comp1LEX	comp1GDP	%var2*5	comp2EDU	comp2LEX	comp2GDP
Benin	BEN	Low	Sub-Saharan Africa	1	84.449	0.985	0.910	0.857				
Burkina Faso	BFA	Low	Sub-Saharan Africa	2.2	70.034	0.957	0.463	0.986				
Burundi	BDI	Low	Sub-Saharan Africa	3.1	76.949	−0.848	0.916	0.866				
Central African Republic	CAF	Low	Sub-Saharan Africa	3.1	87.940	−0.978	0.886	0.946				
Chad	TCD	Low	Sub-Saharan Africa	4	40.192	0.769	−0.019	0.784	38.164	0.424	0.900	−0.394
Côte d'Ivoire	CIV	Low	Sub-Saharan Africa	3.1	91.199	−0.975	0.918	0.972				
Eritrea	ERI	Low	Sub-Saharan Africa	3.3	81.714	0.981	0.947	−0.769				
Ethiopia	ETH	Low	Sub-Saharan Africa	1	79.947	0.948	0.904	0.826				
Gambia	GMB	Low	Sub-Saharan Africa	1	84.048	0.975	0.959	0.807				
Guinea-Bissau	GNB	Low	Sub-Saharan Africa	2.3	67.202	0.961	0.991	0.333				
Malawi	MWI	Low	Sub-Saharan Africa	3.4	62.765	−0.927	0.879	0.500				
Mali	MLI	Low	Sub-Saharan Africa	2.3	70.664	0.989	0.873	0.615				
Mozambique	MOZ	Low	Sub-Saharan Africa	3.2	72.922	0.941	−0.656	0.933				
Niger	NER	Low	Sub-Saharan Africa	3.3	93.506	0.984	0.973	−0.944				
Rwanda	RWA	Low	Sub-Saharan Africa	3.4	51.068	−0.550	0.822	0.744				
Senegal	SEN	Low	Sub-Saharan Africa	4	65.545	0.984	0.995	−0.095	33.967	0.175	−0.060	0.995
Togo	TGO	Low	Sub-Saharan Africa	3.4	65.897	0.927	0.496	−0.934				
Zambia	ZMB	Low	Sub-Saharan Africa	3.1	94.701	−0.966	0.963	0.990				
Algeria	DZA	Medium	Middle East & North Africa	2.3	74.844	0.962	0.971	0.615				
Armenia	ARM	Medium	Europe & Central Asia	2.2	57.796	0.905	0.308	0.905				
Bangladesh	BGD	Medium	South Asia	1	97.641	0.984	0.990	0.991				
Belize	BLZ	Medium	Latin America & Caribbean	2.2	81.480	0.979	0.785	0.932				
Bolivia	BOL	Medium	Latin America & Caribbean	2.3	66.786	0.998	0.981	0.214				
Botswana	BWA	Medium	Sub-Saharan Africa	2.2	86.126	0.942	0.702	0.939				
Cambodia	KHM	Medium	East Asia & Pacific	2.1	79.055	0.776	0.991	0.887				
Cameroon	CMR	Medium	Sub-Saharan Africa	3.4	57.440	−0.725	0.849	0.691				
Cabo Verde	CPV	Medium	Sub-Saharan Africa	1	98.050	0.982	0.998	0.990				
China	CHN	Medium	East Asia & Pacific	1	99.392	0.997	0.995	0.998				
Comoros	COM	Medium	Sub-Saharan Africa	3.3	78.335	0.823	0.951	−0.876				
Congo	COG	Medium	Sub-Saharan Africa	3.2	73.043	0.923	−0.696	0.924				
Djibouti	DJI	Medium	Middle East & North Africa	3.3	96.292	0.975	0.999	−0.969				
Dominican Republic	DOM	Medium	Latin America & Caribbean	1	95.919	0.990	0.972	0.976				
Egypt	EGY	Medium	Middle East & North Africa	1	96.983	0.995	0.989	0.970				
El Salvador	SLV	Medium	Latin America & Caribbean	2.3	75.384	0.951	0.964	0.654				
Equatorial Guinea	GNQ	Medium	..	3.2	69.019	0.805	−0.658	0.995				
Fiji	FJI	Medium	East Asia & Pacific	1	93.191	0.984	0.962	0.950				
Ghana	GHA	Medium	Sub-Saharan Africa	2.3	73.731	0.991	0.906	0.639				
Guatemala	GTM	Medium	Latin America & Caribbean	1	88.569	0.979	0.971	0.870				
Guyana	GUY	Medium	Latin America & Caribbean	1	85.202	0.977	0.868	0.920				
Haiti	HTI	Medium	Latin America & Caribbean	3.1	93.585	−0.982	0.934	0.986				
Honduras	HND	Medium	Latin America & Caribbean	2.3	82.628	0.967	0.965	0.783				
India	IND	Medium	South Asia	1	99.408	0.998	0.999	0.995				
Indonesia	IDN	Medium	East Asia & Pacific	1	99.244	0.995	0.998	0.996				
Iran, Islamic Rep. of	IRN	Medium	Middle East & North Africa	2.3	68.355	0.987	0.990	0.311				
Jamaica	JAM	Medium	Latin America & Caribbean	4	55.613	0.917	0.303	0.858	33.668	0.017	0.942	−0.351
Jordan	JOR	Medium	Middle East & North Africa	3.3	82.206	0.952	0.962	−0.797				
Kenya	KEN	Medium	Sub-Saharan Africa	4	60.423	−0.795	0.975	0.479	36.281	0.578	0.045	0.867
Lao People's Dem. Rep.	LAO	Medium	East Asia & Pacific	1	99.208	0.999	0.993	0.996				
Lesotho	LSO	Medium	Sub-Saharan Africa	3.2	86.040	0.952	−0.857	0.970				
Madagascar	MDG	Medium	Sub-Saharan Africa	3.3	95.744	0.972	0.989	−0.975				
Mauritania	MRT	Medium	Sub-Saharan Africa	2.3	75.363	0.976	0.952	0.635				
Moldova, Rep. of	MDA	Medium	Europe & Central Asia	3.1	60.570	−0.633	0.828	0.855				
Mongolia	MNG	Medium	East Asia & Pacific	2.1	70.498	0.623	0.874	0.981				
Morocco	MAR	Medium	Middle East & North Africa	1	98.205	0.994	0.993	0.986				
Namibia	NAM	Medium	Sub-Saharan Africa	3.4	77.982	0.723	0.937	−0.969				
Nepal	NPL	Medium	South Asia	1	99.123	0.994	0.999	0.993				
Nicaragua	NIC	Medium	Latin America & Caribbean	3.3	90.924	0.972	0.976	−0.911				
Nigeria	NGA	Medium	Sub-Saharan Africa	4	50.436	0.318	−0.880	0.799	34.221	0.922	−0.045	−0.417
Pakistan	PAK	Medium	South Asia	1	99.529	0.997	0.998	0.999				
Papua New Guinea	PNG	Medium	East Asia & Pacific	2.3	74.981	0.982	0.945	0.627				
Paraguay	PRY	Medium	Latin America & Caribbean	2.3	69.914	0.988	0.967	0.431				
Philippines	PHL	Medium	East Asia & Pacific	2.3	80.171	0.973	0.953	0.742				
Samoa (Western)	WSM	Medium	East Asia & Pacific	1	92.239	0.986	0.948	0.947				
South Africa	ZAF	Medium	Sub-Saharan Africa	4	61.399	−0.960	0.010	0.959	34.926	−0.158	0.997	−0.169
Sri Lanka	LKA	Medium	South Asia	1	96.746	0.973	0.999	0.978				
Sudan	SDN	Medium	Sub-Saharan Africa	1	94.368	0.990	0.981	0.943				
Swaziland	SWZ	Medium	Sub-Saharan Africa	3.2	84.283	0.878	−0.900	0.973				
Syrian Arab Republic	SYR	Medium	Middle East & North Africa	1	93.904	0.968	0.995	0.943				
Tajikistan	TJK	Medium	Europe & Central Asia	2.2	67.076	0.987	0.276	0.981				
Tanzania, U. Rep. of	TZA	Medium	Sub-Saharan Africa	3.4	78.764	0.990	−0.886	0.774				
Thailand	THA	Medium	East Asia & Pacific	1	94.718	0.950	0.983	0.987				
Tunisia	TUN	Medium	Middle East & North Africa	1	98.296	0.998	0.991	0.985				
Uganda	UGA	Medium	Sub-Saharan Africa	3.2	72.417	0.959	−0.543	0.979				
Ukraine	UKR	Medium	Europe & Central Asia	3.1	75.303	−0.744	0.969	0.875				
Uzbekistan	UZB	Medium	Europe & Central Asia	2.1	72.140	0.528	0.957	0.985				
Viet Nam	VNM	Medium	East Asia & Pacific	1	98.504	0.989	0.999	0.989				
Yemen	YEM	Medium	Middle East & North Africa	1	97.049	0.995	0.990	0.971				

Albania	ALB	High	Europe & Central Asia	1	*88.570*	0.951	0.991	0.878				
Argentina	ARG	High	Latin America & Caribbean	2.3	*71.963*	0.959	0.970	0.547				
Bahamas	BHS	High	..	2.1	*54.848*	0.336	0.887	0.864				
Bahrain	BHR	High	..	2.3	*78.426*	0.927	0.980	0.730				
Belarus	BLR	High	Europe & Central Asia	4	*65.082*	0.999	−0.755	0.620	*34.723*	0.008	0.654	0.784
Brazil	BRA	High	Latin America & Caribbean	1	*93.447*	0.984	0.983	0.932				
Bulgaria	BGR	High	Europe & Central Asia	2.3	*74.970*	0.877	0.929	0.785				
Chile	CHL	High	Latin America & Caribbean	1	*97.482*	0.981	0.987	0.994				
Colombia	COL	High	Latin America & Caribbean	1	*96.036*	0.984	0.990	0.966				
Costa Rica	CRI	High	Latin America & Caribbean	1	*93.824*	0.983	0.961	0.962				
Croatia	HRV	High	..	2.3	*76.513*	0.972	0.956	0.661				
Ecuador	ECU	High	Latin America & Caribbean	1	*88.447*	0.987	0.914	0.919				
Estonia	EST	High	..	1	*97.000*	0.991	0.990	0.974				
Hungary	HUN	High	..	1	*94.260*	0.987	0.959	0.966				
Kazakhstan	KAZ	High	Europe & Central Asia	4	*62.936*	0.982	−0.105	0.955	*36.577*	−0.170	0.994	0.284
Latvia	LVA	High	Europe & Central Asia	2.3	*81.680*	0.962	0.966	0.769				
Lebanon	LBN	High	Middle East & North Africa	1	*95.879*	0.984	0.995	0.959				
Lithuania	LTU	High	Europe & Central Asia	2.3	*78.348*	0.947	0.979	0.704				
Malaysia	MYS	High	East Asia & Pacific	1	*99.356*	0.997	0.996	0.997				
Mauritius	MUS	High	Sub-Saharan Africa	1	*99.500*	0.997	0.997	0.998				
Mexico	MEX	High	Latin America & Caribbean	1	*95.245*	0.995	0.972	0.961				
Oman	OMN	High	..	1	*96.922*	0.992	0.988	0.973				
Panama	PAN	High	Latin America & Caribbean	1	*93.537*	0.982	0.974	0.945				
Peru	PER	High	Latin America & Caribbean	3.3	*73.987*	0.962	0.973	−0.588				
Romania	ROM	High	..	2.3	*76.910*	0.973	0.956	0.669				
Russian Federation	RUS	High	Europe & Central Asia	4	*56.204*	0.365	0.790	0.964	*39.592*	0.916	−0.576	0.125
Saudi Arabia	SAU	High	..	3.3	*93.615*	0.980	0.988	−0.933				
Trinidad and Tobago	TTO	High	..	4	*56.120*	0.924	0.894	0.172	*34.211*	0.067	−0.258	0.977
Turkey	TUR	High	Europe & Central Asia	1	*98.989*	0.995	0.996	0.994				
Uruguay	URY	High	Latin America & Caribbean	1	*84.371*	0.945	0.984	0.818				
Venezuela	VEN	High	Latin America & Caribbean	3.3	*95.318*	0.995	0.976	−0.957				
Australia	AUS	Very High	..	1	*96.775*	0.968	0.990	0.993				
Austria	AUT	Very High	..	1	*97.548*	0.977	0.994	0.992				
Belgium	BEL	Very High	..	1	*97.487*	0.977	0.997	0.988				
Canada	CAN	Very High	..	1	*89.297*	0.882	0.985	0.965				
Cyprus	CYP	Very High	..	1	*97.282*	0.974	0.996	0.989				
Czech Republic	CZE	Very High	..	1	*87.825*	0.887	0.922	0.999				
Denmark	DNK	Very High	..	1	*96.233*	0.972	0.997	0.974				
Finland	FIN	Very High	..	1	*97.894*	0.986	0.997	0.986				
France	FRA	Very High	..	1	*98.425*	0.986	0.999	0.991				
Germany	DEU	Very High	..	1	*93.128*	0.930	0.981	0.984				
Greece	GRC	Very High	..	1	*91.556*	0.962	0.935	0.973				
Hong Kong, China (SAR)	HKG	Very High	..	1	*97.517*	0.991	0.991	0.980				
Iceland	ISL	Very High	..	1	*96.239*	0.974	0.998	0.971				
Ireland	IRL	Very High	..	1	*98.854*	0.993	0.994	0.996				
Israel	ISR	Very High	..	1	*98.418*	0.994	0.997	0.985				
Italy	ITA	Very High	..	1	*93.261*	0.934	0.994	0.968				
Japan	JPN	Very High	..	1	*92.227*	0.924	0.994	0.961				
Korea, Rep. of	KOR	Very High	East Asia & Pacific	1	*99.286*	0.994	0.999	0.996				
Kuwait	KWT	Very High	..	3.3	*81.925*	0.964	0.937	−0.806				
Luxembourg	LUX	Very High	..	1	*94.706*	0.972	0.984	0.963				
Malta	MLT	Very High	..	1	*98.794*	0.991	0.999	0.992				
Netherlands	NLD	Very High	..	1	*97.708*	0.985	0.991	0.989				
New Zealand	NZL	Very High	..	1	*97.615*	0.986	0.992	0.986				
Norway	NOR	Very High	..	1	*98.223*	0.988	0.997	0.988				
Portugal	PRT	Very High	..	1	*95.623*	0.960	0.985	0.989				
Singapore	SGP	Very High	..	1	*99.152*	0.998	0.996	0.993				
Slovenia	SVN	Very High	..	1	*99.028*	0.991	0.999	0.996				
Spain	ESP	Very High	..	1	*98.190*	0.990	0.995	0.987				
Sweden	SWE	Very High	..	1	*95.674*	0.957	0.995	0.982				
Switzerland	CHE	Very High	..	1	*90.813*	0.935	0.994	0.929				
United Arab Emirates	ARE	Very High	..	3.3	*97.674*	0.990	0.984	−0.990				
United Kingdom	GBR	Very High	..	1	*95.143*	0.950	0.990	0.986				
United States	USA	Very High	..	1	*96.461*	0.966	0.984	0.996				

*1 According to the classification defined by Human Development Report 2009 (UNDP 2009).

*2 World bank website http://go.worldbank.org/D7SN0B8YU0 (last accessed in March 28, 2012)

*3 1=Category I, 2.1=II-1, 2.2=II-2, 2.3=II-3, 3.1=III-1, 3.2=III-2, 3.3=III-3, 3.4=III-4, 4=IV.

The definition of each category and sub-category is as follows. Also refer to Table 4 in the main text for details.

Category I: All components are above 0.8.

Category II: One component is below 0.8 but the shape maintains a triangle within the chart. — II-1: EDU below 0.8, II-2: LEX below 0.8, II-3: GDP below 0.8.

Category III: One component is negative and the triangle deviates out of the chart. — III-1: EDU negative, III-2: LEX negative, III-3: GDP negative, III-4: Mixed with a category II trait.

Category IV: Two or more principal components derived, with traits of categories II and/or III.

*4 Component matrix of the first (or second) principal component: the components below 0.8 are in double-lined cells, the negative components are highlighted in gray.

*5 The power of explanation of the first (or second) principal component.

Source : Created by author.

Notes

1 Of the three indices comprising the HDI, the definitions of the economic and educational indices were amended in the *HDR 2010*. The economic index, which was formerly defined by GDP, was changed to GNI; and, the educational indices, formerly defined by school enrollment and literacy rates, were changed to mean years of schooling and expected years of schooling (UNDP 2010).

2 For major first-generation theories, see Nurkse (1953), Hirschman (1958), Rostow (1960), Gershenkron (1966), Bauer (1971), and Chenery (1979).

3 For major second-generation theories, see Becker (1967) and Schultz (1971) for the human capital theory and Streeten (1981) for the basic human needs approach (BHN), a humanitarian economic theory.

4 Due to the rapid increase of the HDI value in many countries, the "very high human development" group (HDI of 0.900 or above) was added to the existing definitions of the "low human development" group (HDI of 0.500 or less), the "medium human development" group (HDI of 0.500-0.799), and the "high human development" group (HDI of 0.800-0.899) since *HDR 2009* (UNDP 2009).

5 The definition and calculation of the HDI have been amended substantially since *HDR 2010* (UNDP 2010). Although the latest definition of the HDI is introduced in this section, the data used in the analysis section later in this paper was collated under the old definition. For the old definition, see the technical notes in *HDR 2009* (UNDP 2009).

6 On the other hand, since the HDI attempts to measure the level of human development by using only four indicators, it has been criticized for its conceptual accuracy (Jolly 2002). Moreover, as the rankings may possibly change depending on the goalpost values, there is another argument regarding technical issues in the HDI definition. For example, Panigrahi and Sivramkrishna (2000) have proposed a new type of HDI called Adjusted HDI (AHDI), which recalculates HDI using an additional adjustment term that is not influenced by the goalpost values. In addition, the UNDP (2010) has released the Inequality-adjusted HDI (IHDI) that adjusts the HDI with the degree of social inequality.

7 Sen has repeatedly insisted on the importance of taking a comprehensive perspective in development, one that is not limited to the economic perspective alone (see Sen 2000, etc.). He also has stated that he was initially unwilling to create a rough, single index, such as the HDI, which might not appropriately reflect the comprehensiveness of development (Sen 1999b, p. 23). From these statements, it is clear that using a single index value to evaluate development is not his original intention. Taking these background into consideration, this paper tries to create an approach towards comprehensive development evaluation, beyond evaluation using a single index value, by focusing on development balance.

8 The analyses in this paper were conducted following the definition of the HDI (three indices). There are two reasons why the existing definition was used despite its limitations. The first reason is, as discussed later, that the chart of development balance in this paper is assumed to be used to complement the conventional HDI. The second reason is that there is a strong tendency for discussions about a better index to be unrealistic in terms of the practical feasibility and availability in international secondary data and the fact that the process of index selection is inevitably an arbitrary exercise.

9 The comprehensiveness expressed in the chart of development balance in this paper is based on the definition of the HDI. Therefore, when the definition of the HDI itself is modified, the chart of development balance is also modified according to those changes. Although a triangle-shaped chart is proposed in this paper because the current HDI has three indices, the chart shape may possibly change into a rectangle, pentagon, or other shape, following the addition of more indices in the future.

10 Due to the increase in the number of countries in the "high human development" group, the *HDR 2009* (UNDP 2009) introduced a new group called the "very high human development" group (HDI of 0.900 or above). This paper follows the classification in the 2009 report.

11 On the importance of public deliberation and democracy in decision-making for public policy, see Sen (2000, chapter 6; 2009, chapters 15-18).

References

Ahsan, M. (2004). Human development in the Muslim world: From theory to practice. *The Muslim World*, 94, 181-200.

Bauer, P. (1971). *Dissent on Development*. London: Widenfeld and Nicolson.

Becker, G. (1967). *Human Capital and the Personal Distribution of Income: An Analytical Approach*. Ann Arbor, MI: University of Michigan.

Chenery, H. B. (1979). *Structural Change and Development Policy*. New York: Oxford University Press.

Despotis, D. K. (2005). A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society*, 56(8), 969-980.

Esho, H. (1997). *Kaihatsu no seiji-keizai-gaku* [The political economics of development]. Tokyo: Nihon Hyoronsha.

Gershenkron, A. (1966). *Economic Backwardness in Historical Perspective*. Cambridge, MA: Belknap Press.

Grimm, L., and Yarnold, P. (2004). *Reading and Understanding Multivariate Statistics*. Washington, DC: American Psychological Association.

Haq, K. (2000). Human development challenges in South Asia. *Journal of Human Development*. 1(1), 71-82.

Herrero, C., Martinez, R., and Villar, A. (2010). Improving the measurement of human development. *United Nations Development Programme Human Development Research Paper*, 2010(12).

Hidalgo, C. A. (2010). Graphical statistical methods for the representation of the human development index and its components. *United Nations Development Programme Human Development Research Paper*, 2010(39).

Hirschman, A. (1958). *The Strategy of Economic Development*. New Haven: Yale University Press.

Jolly, R. (2002). Statisticians of the world unite: The human development challenge awaits. *Journal of Human Developmen*t, 3(2), 263-272.

Meier, G. M., and Stiglitz, J. E. (Eds.). (2001). *Frontiers of development economics*: *The future in perspective*. New York: Oxford University Press.

Nurkse, R. (1953). *Problems of Capital Formation in Underdeveloped Countries*. Oxford: Basil Blackwell.

Panigrahi, R., and Sivramkrishna, S. (2002). An adjusted human development index: Robust country rankings with respect to the choice of fixed maximum and minimum indicator values. *Journal of Human Development*, 3(2), 301-311.

Rostow, W. W. (1960). *The Stages of Economic Growth*. New York: Cambridge University Press.

Schultz, T. (1971). *Investment in human capital*. New York: The Free Press.

Sen, A. (1981). *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Clarendon Press.

Sen, A. (1987). *On Ethics and Economics*. Oxford: Basil Blackwell.

Sen, A. (1992). *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

Sen, A. (1997). *On Economic Inequality*. Oxford: Clarendon Press.

Sen, A. (1999a). *Commodities and Capabilities* (Oxford India Paperbacks). New Delhi: Oxford University Press.

Sen, A. (1999b). Assessing human development. In *Human Development Report 1999: Globalization with a Human Face*. New York: Oxford University Press.

Sen, A. (2000). *Development as Freedom*. New York: Anchor Books.

Sen, A. (2002). Capability and well-being. In M. Nussbaum and A. Sen (Eds.), *The Quality of Life* (pp. 30-53). Oxford: Oxford University Press.

Sen, A. (2009). *The Idea of Justice*. London: Penguin Books.

Streeten, P. (1981). *First Things First: Basic Human Needs in Developing Countries*. Washington, DC: World Bank Publications.

Tabachnick, B., and Fidell, L. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn & Bacon.

Tone, K. (1993). *Keizai koritsusei no sokutei to kaizen: Horaku bunsekiho DEA ni yoru* [Measurement and improvement of management efficiency: Using the data envelopment analysis (DEA) method]. Tokyo: JUSE Press.

United Nations Development Programme. (2007). *Human Development Report 2007/2008: Fighting Climate Change: Human Solidarity in a Divided World*. New York: Oxford University Press.

United Nations Development Programme. (2009). *Human Development Report 2009: Overcoming Barriers*. New York: Oxford University Press.

United Nations Development Programme. (2010). *Human Development Report 2013: Rethinking Human Development*. New York: Oxford University Press.

United Nations Development Programme. (2012a). *Human Development Reports: National Reports*. Retrieved July 18, 2012, from http://hdr.undp.org/en/reports.

United Nations Development Programme. (2012b). *Human Development Reports: Indices & Data*. Retrieved July 18, 2012, from http://hdr.undp.org/en/statistics.

World Bank Independent Evaluation Group. (2010). *Priorities in Meeting the MDGs: Lessons from Evaluation*. Washington, DC: Author.

Yonehara, A. (2009). *Human Development Policy in the Global Era: A Proposal from an Educational View*. Okayama, Okayama: University Education Press.

(Accepted 5th August, 2013)

日本評価学会第14回全国大会
「地域活性化への評価の役割」開催の報告とお礼

2013年12月14-15日開催（於：米子コンベンションセンター）の日本評価学会第14回全国大会には112名の方々にご出席いただき誠にありがとうございました。会員各位の日頃の研究や実践活動の報告を基に、評価研究者、実務者の間の経験、情報、知識を共有化する場として、皆様にとって有益な機会となりましたら幸甚に存じます。多くの方のご尽力により本大会を開催することができましたこと、心より感謝申し上げます。今後とも当学会の活動にご高配賜りますようどうぞ宜しくお願い申し上げます。

実行委員長　小野達也（鳥取大学）
プログラム委員長　牟田博光（東京工業大学）

2013年12月14日（土）

13：00 - 13：30	受　付		
午後の部Ⅰ 13：30 - 15：30	**共通論題 1** SROI評価：自立支援団体の社会的投資リターンの計算の実例 （伊藤健／佐々木亮） 第4会議室	**共通論題 2** 行政評価局調査の近年の取組 （小島卓弥） 第5会議室	**共通論題 3** グローバル人材の育成・定着支援による地域の活性化と評価 （佐藤由利子） 第6会議室
午後の部Ⅱ 15：45 - 17：45	**共通論題4** 科学技術への社会的期待の可視化と研究開発の社会インパクト評価 （伊藤健） 第4会議室	**共通論題5** 公的法人のガバナンスと評価－独立行政法人等公的法人・自治体等の論点－ （南島和久） 第5会議室	**自由論題1** 東日本大震災・被災地での活動評価 （林薫） 第6会議室
18：00 - 18：30	総会　第７会議室		
18：45 - 20：00	懇親会（レストラン Le Port）		

2013年12月15日（日）

9：00 - 9：30	受　付		
午前の部 9：30 - 11：30	**共通論題6** 民間ビジネスの社会貢献評価の課題 (世界銀行グループとJICAの取り組みから) （畑島宏之） 第4会議室	**ＲＴＤ** 評価人材の教育について （田中啓） 第5会議室	**自由論題2** 教育・研修評価 （齊藤貴浩） 第6会議室
11：30 - 12：30	お昼休み		
午後の部 12：30 - 14：30	**共通論題7** エビデンスに基づく教育 －日本での実践に向けて－ （田辺智子） 第4会議室	**自由論題3** 評価手法 （廣野良吉） 第5会議室	**自由論題4** 行政評価 （小野達也） 第6会議室

（　　）は座長です。

日本評価学会第14回全国大会
「地域活性化への評価の役割」プログラム詳細

12月14日（土）　12:00-13:00　理事会

12月14日（土）午後の部Ⅰ　13:30-15:30

共通論題１　SROI評価：自立支援団体の社会的投資リターンの計算の実例　第４会議室

座長　伊藤　健　慶應義塾大学
座長　佐々木　亮　(株)国際開発センター
司会　大沢　望　早稲田大学
コメンテータ　池内　賢二　日本財団

K1-1	SROI評価の事例1：社会福祉法人オリーブの樹		伊藤　健	慶應義塾大学
K1-2	SROI評価の事例2：社会福祉法人はる		佐々木　亮	(株)国際開発センター
		○	伊藤　健	慶應義塾大学

共通論題２　行政評価局調査の近年の取組　第５会議室

座長　小島　卓弥　総務省
コメンテータ　西出　順朗　岩手県立大学
コメンテータ　南島　和久　神戸学院大学

K2-1	複数府省にまたがる政策の評価 「事例；法曹人口の拡大及び法曹養成制度の改革に関する政策評価」		松本　順	総務省
K2-2	行政評価・監視 「事例；社会資本の維持管理及び更新に関する行政評価・監視」		楠原　修	総務省
K2-3	地域計画調査 「事例；行政評価局が実施している地域計画調査 ～中国四国管区行政評価局の地域計画調査を中心に～」		城代　充郎	総務省

共通論題3　グローバル人材の育成・定着支援による地域の活性化と評価　第６会議室

座長　佐藤　由利子　東京工業大学

K3-1	地域における留学生の受入れと定着支援の評価と課題 ～広島県の事例を中心として～		尾本　勝昭	広島大学大学院
K3-2	島根県内のインターンシップによる 就職支援の効果測定の枠組み		青山　桃子	(株)シーズ総合政策研究所
K3-3	グローバル人材の育成・就職支援による地域の活性化 ～関連政策・事業の評価の枠組み～		佐藤　由利子	東京工業大学

12月14日（土）　午後の部Ⅱ　15:45-17:45

共通論題４　科学技術への社会的期待の可視化と研究開発の社会インパクト評価　第４会議室

座長　伊藤　健　慶應義塾大学
司会　玉村　雅敏　慶應義塾大学

K4-1	科学技術の予測と評価 ～事前評価としての予測と社会インパクト評価～		白川　展之	慶應義塾大学
K4-2	科学技術の社会的期待に対する統合評価モデル～スマートシティ構想に関する政策マーケティングとSROIの設計思想～		高橋　武俊	慶應義塾大学
			玉村　雅敏	慶應義塾大学
			白川　展之	慶應義塾大学
			伊藤　健	慶應義塾大学
K4-3	SROI法による定量評価を活用したテクノロジー・ソリューションとコミュニティ・ソリューションの最適施策形成		伊藤　健	慶應義塾大学

共通論題５　公的法人のガバナンスと評価ー独立行政法人等公的法人・自治体等の論点ー　第５会議室

座長・司会・コメンテータ　南島　和久　神戸学院大学

K5-1	行政改革における公的法人の制度的地位		新井　誠一	内閣府
K5-2	独立行政法人のガバナンスとコントロール		森田　弥生	総務省
K5-3	自治体等のマネジメントおける予算と評価の役割		佐藤　幹	広島大学

自由論題１　東日本大震災・被災地での活動評価　第６会議室

座長　林　薫　文教大学

J1-1	震災復興における施設整備事業の評価の視点～QFF・日本財団助成事業「多機能水産加工施設整備事業」より		岩渕　祐二	公共価値創造研究所
J1-2	ピーズ・ウィンズ・ジャパン東日本大震災心理社会的ケアプロジェクト～評価結果から導き出される日本のNGOが国内の被災地で活動する際の教訓～	○	小林　美紀	元NPO法人HANDS
			中村　安秀	大阪大学
			西井　栄美子	NPO法人HANDS
			和田　知代	NPO法人HANDS
			平野　志穂	NPO法人HANDS
			沼田　眸	NPO法人HANDS
J1-3	大学生を主体とした被災地支援活動の実践と評価		星野　敦子	十文字学園女子大学

12月14日（土）　総会　18:00-18:30

12月14日（土）　懇親会　18:45-20:00

12月15日（日）　午前の部　9:30-11:30

共通論題6　民間ビジネスの社会貢献評価の課題 ー世界銀行グループとJICAの取り組みからー

第４会議室

座長　畑島　宏之　世界銀行

K6-1	JICAのBOPビジネス連携促進事業と評価フレームワーク整備にむけての取組み		池田　則宏	国際協力機構
K6-2	民間ビジネスにおける社会貢献評価手法の課題　世界銀行グループの経験から		畑島　宏之	世界銀行

ＲＴＤ　評価人材の教育について

第５会議室

座長　田中　啓　　静岡文化芸術大学
司会　小島　卓弥　総務省

RTD-1	行政評価局における人材育成に関して		松田　綱児	総務省
RTD-2	自治体における評価教育に関して		佐藤　　徹	高崎経済大学
RTD-3	日本の政策系大学・大学院での評価教育に関して		窪田　好男	京都府立大学
RTD-4	米国の政策系大学・大学院での評価教育に関して		西出　順郎	岩手県立大学

自由論題2　教育・研修評価

第６会議室

座長　齊藤　貴浩　大阪大学

J2-1	技術・家庭科技術分野におけるポートフォリオ検討会の教育効果に関する研究Ｉ		磯部　征尊	新潟市立亀田小学校
		○	伊藤　大輔	金沢工業大学
			五十嵐　和義	上越市立上越特別支援学校
J2-2	発達障害のある生徒を対象にした学習支援員（LSA）普及事業の評価 ～助成事業評価にプログラム評価の継続的改善評価CD-TEPアプローチ法を用いた試み～	○	新藤　健太	日本社会事業大学
			大島　巌	日本社会事業大学
			鴨澤　小織	日本社会事業大学
			及川　博文	日本社会事業大学
			村里　優	日本社会事業大学
			池内　賢二	日本財団
			大田　直樹	日本財団
			藤堂　栄子	（特活）エッジ
J2-3	開発途上国支援のための研修の効果 ～JICA課題別研修を事例として～	○	岸本　昌子	（一財）日本国際協力センター
		○	菊田　怜子	（一財）日本国際協力センター
J2-4	インパクト評価研修プログラムの効果測定	○	青柳　恵太郎	国際協力機構
			越知　直哉	国際協力機構
			西野　宏	国際協力機構

12月15日（日）　お昼休み　11:30-12:30（編集委員会 11:30-12:30）

12月15日（日）　午後の部　12:30-14:30

共通論題７　エビデンスに基づく教育ー日本での実践に向けてー

第４会議室

座長　田辺　智子　国立国会図書館
コメンテータ　津富　宏　　静岡県立大学

K7-1	英国におけるエビデンスに基づく教育		惣脇　宏	京都大学
K7-2	エビデンスに基づく教育とは		森　俊郎	岐阜県養老町立東部中学校
K7-3	EBE研究会の活動について ～実践･研究･行政の三位一体を目指して～		中井　俊之	広島市立福木小学校
K7-4	生徒の主体的な学びを実現している教師の実践知～ナラティブとエビデンスを統合した研究アプローチの提案～	○	水野　正朗	名古屋市立桜台高等学校
			清道　亜都子	三重短期大学
			柴田　好章	名古屋大学

自由論題３　評価手法

第５会議室

座長　廣野　良吉　成蹊大学

J3-1	効果的プログラムモデル形成のための実践家参画型評価アプローチ法の開発（その2）： 「評価ファシリテータ」の機能と役割に注目して	○	大島　巌	日本社会事業大学
			新藤　健太	日本社会事業大学
			贄川　信幸	日本社会事業大学
			中越　章乃	神奈川県立保健福祉大学
			小佐々　典靖	国立障害者リハビリセンター
			廣瀬　圭子	目白大学
J3-2	政府予算案分析のための独立財政機関とそれらの評価基準		上野　宏	（一財）国際開発センター
J3-3	米国の「政策評価への事業費1パーセント保留」の歴史的意義		上野　真城子	アジア都市コミュニティー研究センター
J3-4	1人の父兄からの学校評価		石田　健一	東京大学

自由論題４　行政評価

第６会議室

座長　小野　達也　鳥取大学

J4-1	地方自治体における建築物の地球温暖化対策の制度と評価		中田　雅陽	法政大学
J4-2	図書館経営において評価活動が事業の改善に与える影響		下山　佳那子	筑波大学
J4-3	市民判定人と仕分け人の評価の関係について		佐藤　章	大津市役所
J4-4	参加型戦略プログラム評価の導入		真野　毅	京都産業大学

共通論題セッション報告

共通論題セッション１「SROI評価：自立支援団体の社会的投資リターンの計算の実例」

座長　伊藤 健（慶應義塾大学）
座長　佐々木 亮（(株)国際開発センター）
司会　大沢 望（早稲田大学）
コメンテータ　池内 賢二（日本財団）

共通論題セッション２「行政評価局調査の近年の取組」

座長　小島 卓弥（総務省）
コメンテータ　西出 順郎（岩手県立大学）
コメンテータ　南島 和久（神戸学院大学）

総務省行政評価局では、2011（平成23）年の第12回全国大会より行政評価局が所管し各府省が自ら実施する政策評価やその客観性担保評価、独立行政法人評価等の実践状況について報告してきた。本セッションでは、もう一つの行政評価局業務の大きな柱であり、政府におけるレビュー機能の一端を担う行政評価局調査の近年の取組について事例報告を行い、現状の成果と課題の共有を図ることを目的に企画したものである。

行政評価局調査とは、行政評価局が政府内における第三者的な評価専門機関として、必要性・有効性・効率性等の観点から、複数府省にまたがる政策や各府省の業務の実施状況について調査を実施し、各府省の課題や問題点を実証的かつ総合的に把握・分析し、改善方策を提示するものである。本セッションでは行政評価局調査の三類型である「複数府省にまたがる政策の評価（統一性・総合性確保評価）」「行政評価・監視」及び「地域計画調査」のそれぞれについて概要の整理と近年の実践例について報告した。

第一報告者の松本順氏（総務省行政評価局）からは「複数府省にまたがる政策の評価」の実践例として『法曹人口の拡大及び法曹養成制度の改革に関する政策評価』について、続いて楠原修氏（総務省行政評価局）より「行政評価・監視」の実践例として『社会資本の維持管理及び更新に関する行政評価・監視』について、最後に城代充郎氏（総務省中国四国管区行政評価局）より「地域計画調査」の実践例として『中国四国管区行政評価局の地域計画調査を中心に』についての報告が行われた。

報告を受けてコメンテータの南島和久会員（神戸学院大学）からは、政策判断まで踏み込むような勧告を行う際の留意点や自治事務・司法領域に絡む施策に対して調査を行う際の考え方等について質問があった。同じくコメンテータの西出順郎会員（岩手県立大学）からは近年会計検査院が実施している有効性を重視した検査との違い等について質問があった。

さらに、西出会員からは行政評価局調査を、政府におけるプログラム評価（program evaluation）の一種として捉え直す視点が提起された。これまで、この視点で行政評価局調査を研究した事例は乏しく、今後新たな研究の端緒となることを期待したい。

本学会において各府省や自治体が実施する政策評価制度に関する報告は多く報告され、実務家と研究者の交流による発展が図られてきたところである。今後、行政評価局調査に関しても継続的に報告を行っていき、同様の交流が進むことでさらなる発展向上に繋げていければと考えている。

共通論題セッション３「グローバル人材の育成・定着支援による地域活性化と評価」

座長　佐藤 由利子（東京工業大学）

大都市圏と地方の格差が拡大する中、グローバル人材の育成や定着においても、留学生が大都市圏に集中する、インターナショナルスクールや外国語で対応できる病院などの施設が不十分であるため優秀な外国人材が定着しないなど、地方は大都市圏に比べて不利な状況に置かれている。そのような中で、「アジア人財資金構想」「グローバル人材育成推進事業」「地域中小企業の海外人材確保・定着支援」などの施策を通じ、グローバル人材の育成・定着支援を目指す取り組みが、地方でも行われている。本セッションでは、広島、島根、鳥取における取組みについて紹介し、評価の枠組みと課題について論じた。

第一報告者の広島大学大学院国際協力研究科（広島県留学生活躍支援センター）尾本勝昭氏は、「地域における留学生受入れと定着支援の評価と課題～広島県の事例を中心として～」と題し、アジア人財資金構想事業終了後に自立、発展した広島県の二つ事例（「広島県ものづくりグローバル人材育成協議会」と「広島県留学生活躍支援センター」による人材育成プログラム）を紹介し、地域における産学官連携による留学生受入れと就職・定着支援の事業の取組みの成果と課題を、政策評価の視点から考察した。

第二報告者のシーズ総合政策研究所青山桃子氏は、「島根県内のインターンシップによる 就職支援の効果測定の枠組み」と題し、島根県内の大学生を対象に、県内の中小企業・小規模事業者において行ったインターンシップが、学生の就職活動や職業観の育成、また、地域の雇用にどのように寄与しうるかについて、評価測定方法を検討した結果を報告した。

第三報告者の佐藤（筆者）は、「グローバル人材の育成・就職支援による地域の活性化と評価－関連政策・事業と評価分析の枠組み」と題し、グローバル人材の育成・定着事業において、教育の国際化にかかるコストは、地方の方が大都市圏よりも高く、就職やアルバイト機会は、地方の方が大都市圏よりも少ないなど、地方が抱える課題を抽出した上で、評価の枠組みとしてのログフレームを示した。

これらの報告の後、上述の広島、島根、鳥取の事業について、次の5つの評価視点から、評価の方法と課題を整理し、議論を行った。

＜妥当性＞グローバル人材（日本人・留学生）の育成・定着にかかる事業は、地域経済・社会のニーズに合致しているか？県の政策の中にどのように位置づけられているか？

＜効果＞グローバル人材（日本人・留学生）の育成・定着にかかる事業の効果をモニタリング・評価するには、どのような方法が適当か？

＜インパクト＞グローバル人材（日本人・留学生）の育成・定着にかかる事業が地域社会経済に及ぼすインパクトをモニタリング・評価するには、どのような方法が適当か？

＜効率＞地方の限られた財源の中で、グローバル人材（日本人・留学生）の育成・定着にかかる事業を効率的に行うためには、どのような方法が最も適当か？

＜持続性＞財源が限られる中、グローバル人材（日本人・留学生）の育成・定着にかかる事業を継続していくには、どのような方法が適切で、どのような政策や支援が必要か？

共通論題セッション４「科学技術への社会的期待の可視化と研究開発の社会インパクト評価」

座長　伊藤 健（慶應義塾大学）

司会　玉村 雅敏（慶應義塾大学）

共通論題セッション５「公的法人のガバナンスと評価
ー独立行政法人等公的法人・自治体等の論点ー」

座長・司会・コメンテータ　南島 和久（神戸学院大学）

本セッションでは、公的法人のガバナンスと評価について、行政改革の文脈に位置付けながら検討を行った。ここでいう「公的法人」には、政府・自治体以外に、独立行政法人、公益法人、NPO法人などの公的資金が投入されている法人のことも含まれる。公的資金が投入されている以上、それに対する管理の目は必然的に厳しいものとなる。近年、「評価」や「アカウンタビリティ」が叫ばれているが、そこにはそうした「管理の目」あってのことである。そうである以上、公的資金の投入された公的法人の改革については、今後ともますます重要な行政改革の課題となるだろうし、それに直接的に絡む評価やアカウンタビリティのあり方もより厳しく問われることになるだろう。

以上のような問題意識を設定した上で、本セッションでは政策実施過程を担う法人の改革のあり方を議論することとした。また議論に供する目的で多彩な管理改革を実践している自治体のマネジメント改革も紹介していただいた。

第一報告は新井誠一氏（内閣官房行政改革推進本部事務局）による「行政改革における公的法人の制度的地位」であった。同報告では公的法人の特徴やそのコントロールに関する一覧的な整理が示された。この議論は、公的法人一般に対して、政府はそもそもどうコントロールすべきなのかという原理的な問いへの基礎となる議論である。

第二報告は森田弥生氏（総務省行政評価局）による「独立行政法人のガバナンスとコントロール」であった。政府では、独立行政法人制度10年を超え、法人の業務運営を改善する仕組みとして、監事の機能強化等による法人の内部ガバナンスの強化が議論されている。同報告ではこの監事の機能強化についての改革の意義を議論した。

第三報告は佐藤幹氏（広島大学マネジメント研究センター）による「自治体等のマネジメントおける予算と評価の役割」であった。同報告では、日本の地方自治体の運営の改善のため、地方自治体予算のマネジメント・コントロール・システム化への課題を議論するものであった。あわせて自治体における出資法人等のマネジメントについても紹介された。

本セッションの議論はまだまだ充実させていかなければならない。評価はいわば「行革の時代の産物」である。行政改革の文脈を踏まえつつ、評価やアカウンタビリティ、そして公的法人ガバナンスの課題について、もう一度洗い直す時期を迎えている。

共通論題セッション６「民間ビジネスの社会貢献評価の課題
ー世界銀行グループとJICAの取り組みからー」

座長　畑島 宏之（世界銀行）

途上国へ向かう民間資金は近年、増大している。途上国への資金の7割が、直接投資、証券投資、短期融資、送金、贈与などの民間資金であるとされている。また、民間ビジネスの手法で貧困削減、基本サービスの提供などを行う事業も盛んになっている。そのような途上国での民間活動を前に、2012年版「政府開発援助（ODA）白書」でも、民間資金とODAの連携が援助効果を一層大きくすると指摘している。本セッションではこうした状況をふまえ、民間ビジネス連携を通じた開発事業の事前評価の枠組みと意義、そして、民間投融資の開発効果についての包括的な事後評価手法の特色と課題の二点を発表し、民間ビジネスと開発に対する評価の位置づけを議論した。 第一報告者である池田則宏氏（国際協力機構）は、「JICAのBOPビジネス連携促進事業と評価フレームワーク整備にむけての取組み」の発表で、日本企業が

途上国において貧困層（Base of the Pyramid）のニーズを満たす製品・サービスの提供や貧困層の市場への参加促進が期待できる事業の振興を取り上げた。特に、BOP事業での開発効果を事前に検討するフレームワーク整備への取組みを示し、企業と共有できる事前評価の枠組みを提案した。最後に、この取り組みはまだ始まった段階であり、経験を蓄積していくことで、事業の評価のツールとして適用可能なものか検証し、改善を図っていきたいという抱負を述べた。次に、畑島は、「民間ビジネスにおける社会貢献評価手法の課題～世界銀行グループの経験から」の発表で、民間投融資を専門に行う国際金融公社（IFC）のプロジェクト事後評価手法の理論的背景を紹介した。そこで、民間ビジネスの包括的な開発効果を評価する際、援助の評価手法を応用することが困難であることを、財務指標や市場からの評価の重要性、プロジェクトサイクルの違いなどを挙げて説明した。会場からは多数の質問や意見が出され、池田氏の発表に対して、JICAのBOPビジネス連携促進についての期待と、事前評価枠組みという制度設計について好意的なコメントが寄せられた。また、世界銀行については、先だって発表された貧困削減への傾倒と組織改革の中での評価の占める役割、そして民間事業評価で下される判断の客観性の担保についてなどの議論が交わされた。民間ビジネスが開発のアクターとして強く認識されるようになったが、その活動の評価についてはまだ試行錯誤の段階にあり、継続して意見交換を続けていくことが求められる。

共通論題セッション７「エビデンスに基づく教育―日本での実践に向けて―」

座長　田辺 智子（国立国会図書館）
コメンテータ　津富 宏（静岡県立大学）

エビデンスに基づく実践については、日本評価学会でもしばしば取り上げられ、議論が行われてきた。本セッションでは特に教育分野を取り上げ、日本で今後エビデンスに基づく教育（Evidence Based Education: EBE）を実現していく上での課題について議論を行った。

第一報告者の惣脇宏氏は、国立教育政策研究所在職中に海外のEBEについての研究プロジェクトを実施している。同氏からは、英国におけるEBEを取り上げ、その経緯と性格、特に定量的研究と定性的研究を統合している点に焦点を当てた報告が行われた。第二報告者の森俊郎氏は、現役の中学校教諭であり、教育実践者や研究者等がエビデンス産出や活用に向け取組みを行っているEBE研究会の代表である。同氏からは、同研究会における問題意識と議論、これまでの活動で明らかになった課題等について報告があった。第三報告者の中井俊之氏もEBE研究会のメンバーであり、現役の小学校教諭である。同氏からは、エビデンスに基づく医療における臨床医の行動モデルをもとに、EBEにおける教員の行動モデルの定式化について報告があった。第四報告者の水野正朗氏は、教育学の研究者であると同時に高等学校の教諭である。同氏からは、現場の教員によるエビデンス産出に向けた取組みとして、教師の実践知について、定量的分析と定性的分析を統合して明らかにした研究事例が報告された。

コメンテータの津富宏氏からは、エビデンスの範囲を広く捉えることに懐疑的な立場から、エビデンスの定義と捉え方についてコメントが寄せられた。これを受け、会場を含めて活発な議論が展開された。議論を通じ、今後、日本でエビデンスに基づく実践を目指していく上では、ベースとなるエビデンスの定義と範囲について、さらなる議論と共通理解が必要であることが明らかとなった。今後も、社会福祉、教育、刑事司法など、それぞれの分野でのエビデンス実践や研究発表に加え、分野横断的な議論を行う場が必要といえる。

自由論題セッション報告

自由論題セッション１「東日本大震災・被災地での活動評価」

座長　林 薫（文教大学）

第一報告の岩渕祐二会員（公共価値創造研究所）による「震災復興における施設整備事業の評価の視点～QFF・日本財団助成事業「多機能水産加工施設整備事業」より」はカタールフレンドシップ基金と日本財団が宮城県女川町で実施した多機能水産加工施設の完了時評価について論じたものである。結論として良好な費用便益分析結果が得られていること、水産業施設集積、積極的に事業に取り組む土壌の育成、産業交流の場の発展などの点で、積極的な成果が見られたことなどが報告された。ただ、地域全体の復興事業と産業の発展の効果を見るためには長期的なフォローアップが必要であることが指摘された。

第二報告の小林美紀会員（元NPO法人HANDS）による「ピース・ウィンズ・ジャパン東日本大震災心理社会的ケアプロジェクト～評価結果から導き出される日本のNGOが国内の被災地で活動する際の教訓～」はピース・ウィンズ・ジャパン（PWJ）がComfort for Kids（C4K）の手法にて実施した被災地の子供と大人の心理的ケアのプロジェクトについて、NPOのHANDSが実施した第三者評価の結果をとりまとめたものである。評価はDACの5項目評価によって行われ、結果としてプロジェクト経験のあるスタッフの配置が必要であること、C4K手法の更なる発展や見当が必要であることなどの結論が報告された。あわせ、今後国際協力NGOが同様の支援を途上国で展開する際の示唆が示された。

第三報告の星野敦子会員（十文字学園女子大学）による「大学生を主体とした被災地支援活動の実践と評価」は大学生による福島への風評被害払拭を目的とした有機農業支援活動へ参加した学生の参加動機、事前事後の知識や意識の変化を調査したもので、特に参加の前後の知識や意識に有意差があり、活動の意義が確認されたことが報告の中心であった。しかし、報告者が言及しているように、ロジックモデルに基づく活動自体のアウトカム評価はまだ行われておらず、また調査手法については、事前事後に質問表に回答を求めるより、非質問者の事前事後の変化を答えさせる方が客観的なデータが得られるのではないか、との議論がなされた。

上記3報告は、震災後2年以上を経た時点での様々な復興支援の取り組みの評価に関するものであり、時宜を得たものであるが、長期的な効果の把握のためには更なる継続的な事業のその評価への取り組みが必要であると考える。

自由論題セッション２「教育・研修評価」

座長　齊藤 貴浩（大阪大学）

教育や研修の評価は難しい。その事実が容易に変わることはないが、だからこそ今、評価研究のフロンティアとしてさまざまな取組が試みられている。実際、今大会では共通論題「エビデンスに基づく教育－日本での実践に向けて－」、RTD「評価人材の教育について」も行われた。その中で、自由論題2「教育・研修評価」でも4つの優れた報告が行われ、活発な質疑応答が行われた。

磯部征尊（新潟市立亀田小学校）、伊藤大輔（金沢工業大学）、五十嵐和義（上越市立上越特別支援学校）による「技術・家庭科技術分野におけるポートフォリオ検討会の教育効果に関する研究Ⅰ」は、中学校技術・家庭科技術分野で行われた「ポートフォリオ検討会」が、一部の学生の教育効果を高める効果がある

ことを明らかにした。新藤健太、大島巌（日本社会事業大学）ほか６名による「発達障害のある生徒を対象にした学習支援員（LSA）普及事業の評価」は、プログラム理論、エビデンスの構築、効果的実践をサイクルとして継続的に行っていくCD-TEP評価アプローチ法を、LSA普及のための助成事業評価に用い、それが適用可能であり、かつ適切な評価を行うことが可能になったことを紹介した。岸本昌子（(一財)日本国際協力センター)、菊田怜子による「開発途上国支援のための研修の効果～JICA課題別研修を事例として～」は、国際協力機構が実施している課題別研修の効果を測定し、知識・技術の伝達によって費用効果を高めることができることを示した。青柳恵太郎、越知直哉、西野宏（国際協力機構）による「インパクト評価研修プログラムの効果測定」は、インパクト評価の実施を可能とする人材育成プログラムの紹介と、回顧的事前テストデザインを用いた研修の有効性評価の結果、研修が有効であることを示した。

各報告における質疑応答を通じて参加者はともに理解を深めたが、各研究発表が様々な方法論を駆使する中でも共通の問題として根底にあるのは、教育・研修の効果をどのように測定するのかという問題である。どの参加者も、多くの教育・研修の評価に関する知見と教訓とを集積しながら、学会としてよりよい方法へと近づけていく必要性と可能性を感じたものと考えられる。

自由論題セッション３「評価手法」

座長　廣野 良吉 （成蹊大学）

本セッションでは、4人の報告者を予定していたが、「1人の父兄からの学校評価」について報告者であった東京大学石田健一氏が個人的理由で報告中止となったため、上記3人の報告と質疑応答がなされた。参加者は総数26名。

最初に上野真城子氏による米国での政策分析・評価と政策産業の発展に多大な貢献を果たした米国政府のイニシャテイブについての開陳があった。1966年に健康教育福祉省（HEW現HHS）に設立された計画評価局（ASPE）による諸々のプログラム評価・分析についての考察に続いて、HEW省の事業の予算要求の中に、政策予算の1%を評価のために保留するという政策決定の歴史的意義を説いた。特に、米国の政策産業が資金的にも、雇用の面でも膨大な市場となっているだけでなく、「米国の取り組むべき課題とその解決に、より有効に機能する政策を生み出し、米国のデモクラシーの過程を強靭なものとすることに役立った」という氏の指摘は傾聴に値する。米国の政策産業発展史の考察を通じて、日本の政策形成力の強化と政策産業の育成・成長への示唆を得ようとした上野氏の熱のこもった発表に、参集した多くの日本評価学会の会員が共感を覚え、多大な知的刺激と今後の行動への端緒をえたことは、発表後の多くの質問から疑いない。最後に、「政策評価を行政組織内部の活動とするのは、行政組織自体にとっても実りある成果を期待できない。また外部に委託契約する政策評価研究への資金が準備されていないことは、政策評価産業を政府外に育成する契機をつくることが出来ない。」という上野氏の指摘は、我が国おける政策評価制度の今後の改革に関心ある評価専門家や市民一般の共感を得たであろう。

続いて、登壇した上野宏氏による発表は、政策の実施に不可欠な政府予算案を分析・評価して、広く国民の代表である国会議員の慎重な審議にゆだねるための独立予算分析機関（IFIs）の設置を強く訴えていた。この提案は既に3年前に日本評価学会総会で決議設置した「国会による予算審議・予算評価に関する提言書（案)」（筆者が委員長）の延長線上にあるが、上野氏が主張するように、IFIsの目的・機能は、もともと①財政規律の維持、②立法府支援、③透明性、負託責任の強化を通じた国民・市民支援にある。発表後の質疑応答で指摘されたように、予算制限法（米国）や独立財政機関（英国等IFIs）の導入や、上野氏が主張するIFIsの独立性、非党派性と有能なスタッフという条件をたとえ満たしたとしても、これら諸国で財政赤字削減が実現していのは事実である。しかし、IFIsの下で予算作成過程を通じて②と③が各国間で程度の差こそあれ、有効に果たされていることが重要である。さらに①についても、政策・施策・事

務事業の有効性を高めるための予算編成の担保という財政規律の維持こそがIFIsが狙っていることであり、財政赤字の削減は重要だが、これは政治の課題であるといえる。議会制民主主義体制の下では、再選を指向する政治家は既得権擁護を主張する選挙民・団体を無視することはできない。深刻な財政危機を解決する唯一残された法は、独裁体制の導入（これは望ましくないし、現実には共産党一党独裁の中国でも不可能）、もしくは高い経済成長による財政収入の拡大と行政府の強力なリーダーシップによる無駄な財政支出の大幅な削減に頼らざるを得ないであろう。特に、市民社会の成熟度が低い日本では、この意味から「3本の矢」で各種規制・保護という岩盤をドリルで崩す環境整備を進める「アベノミックス」の効果に期待したいが、その結果は不確定である。

大島氏と同僚は、2010－2011の2年間に亘って発表してきた社会福祉等対人サービスという実践プログラム領域で、効果モデルに関するデータベースを構築し、実践モデルのインパクトを有効性評価法で検証する中で、効果的なプログラム要素、プログラム実施方法やプログラム理論・エビデンス・実践間の円環的対話を通じたCD－TEP評価アプローチ法を提案してきた。また、2013年春季大会では、CD－TEP法を用いた実践家参画型評価アプローチ法開発の現状と課題を考察した。今回の発表では、これらの発表に基づき、CD－TEP評価法を基礎にした包括的評価アプローチ法を提示し、そのアプローチ法を活用できる評価人材を育成するための評価ファシリテーターの機能と役割について検討した。特に、障害者就労移行支援等各種実践家参画型ワークショップを通じて評価ファシリテーターが作成した実施・評価マニュアルと実践家評価担当者マニュアルの作成、評価ツールや実践家評価ファシリテーターの重要性が強調されたが、大島氏自身が本セッションで強調されたように、今後は実践家評価担当者の評価活動の在り方、評価支援の方法論、評価ファシリテーターの実施内容、評価教育の在り方等で、現在研究中の試行プロジェクトを通じて評価ファシリテーターの機能と役割を一層明確化し、「プログラム評価理論・方法論を用いた効果的な福祉実践モデル構築へのアプローチ法開発」研究の成果を期待したい。このような重要な役割をもつ実践家評価ファシリテーターが抱える問題点についてのフロアからの実践家による指摘は説得的であった。

自由論題セッション４「行政評価」

座長　小野達也（鳥取大学）

本セッションでは、「行政評価」というタイトルのもと、行政の制度・施設・事務事業・戦略に関する評価について様々な角度から論ずる4本の報告が行われた。

まず、最初の報告となる中田雅陽氏（法政大学大学院公共政策学専攻博士後期課程）の「地方自治体における建築物の地球温暖化対策の制度と評価」は、自治体において地球温暖化対策のため条例が制定され、実施計画等を定め複数の施策・事務事業等が実施される中、それらの評価は十分ではないという観点から、都道府県における建築物の地球温暖化対策について、その制度や施策の有効性が検証されていないことを指摘したものである。質疑では、国の制度・施策との関係に関する質問などがなされた。

次に、下山佳那子氏（筑波大学大学院博士後期課程）の「図書館経営において評価活動が事業の改善に与える影響」が報告された。本報告は、図書館評価が事業の改善に与える影響の有無を把握するため、公立図書館・大学図書館・学校図書館の評価報告書を対象に改善策が明確に示されているか、示された改善策は次年度以降実施されたかという観点から分析を行ったもので、公立図書館について改善策の実施が確認され、図書館評価が事業の改善に影響を与えている可能性があるとしている。会場からは今後の研究に向けて、改善策の記載やその実施の検証方法に関するコメントなどがなされた。

3番目の報告は、佐藤章氏（大津市役所）の「市民判定人と仕分け人の評価の関係について」であった。本報告は、自治体における事業仕分け等の外部評価においてしばしば設けられる市民判定人の評価と、仕

分け人の議論や意見との関係について、その現状と課題を論じている。前者は後者を基本としつつ、納税者よりも受益者としての判断を示す傾向にあり、納税者の視点による仕分けには事前の学習が必要であるとした。質疑においては、市民判定人には厳しい判定を下すための根拠が不足する傾向などの指摘がなされた。

最後の報告である、真野毅氏（京都産業大学マネジメント研究科博士後期課程）の「参加型戦略プログラム評価の導入」では、報告者が副市長を務める兵庫県豊岡市において、抜本的な経営改革を行う新しい戦略マネジメント手法として実践的に導入が進められている参加型戦略プログラム評価について、その理論的枠組と基本的考え方が紹介された。会場からは、この意欲的な取り組みについて、その実務の実際や今後の見通しなどを巡る質問が提出されるなどした。

本セッションの4本の報告は、何れも自治体における行政や公共サービスの現場に関わる評価について具体的に論じたものであった。これらは行政評価に係る研究と実践の広がりを感じさせるものであり、それぞれの進展が期待される。

日本評価学会春季第11回全国大会
「評価の国際交流」開催の報告とお礼

2014年5月31日開催（於：JICA地球ひろば）の日本評価学会春季第11回全国大会には88名の方々にご出席いただき誠にありがとうございました。会員各位の日頃の研究や実践活動の報告を基に、評価研究者、実務者の間の経験、情報、知識を共有化する場として、皆様にとって有益な機会となりましたら幸甚に存じます。多くの方のご尽力により本大会を開催することができましたこと、心より感謝申し上げます。今後とも当学会の活動にご高配賜りますようどうぞ宜しくお願い申し上げます。

実行委員長　和田義郎（国際協力機構）
プログラム委員長　牟田博光（東京工業大学）

2014年5月31日（土）

9：30 - 10：00	受　付		
午前の部 10：00 - 12：00	共通論題 1 学校評価の評価者育成プログラム （橋本昭彦） Rm. 600	共通論題 4 国際防災評価 （林薫／和田義郎） Rm.201AB	自由論題 1 自治体評価の諸相 （小野達也） Rm.202AB
12：00 - 13：15	お昼休み (12:05～13:00　理事会)		
午後の部Ⅰ 13：15 - 15：15	共通論題 2 自治体戦略の協働マネジメントとプログラム評価 ～自治体における政策評価の新たな試み～ （源由理子） Rm.600	自由論題 2 評価手法 （牟田博光） Rm.201AB	自由論題 3 参加型評価 （高千穂安長） Rm.202AB
午後の部Ⅱ 15：30 - 17：30	共通論題 3 ネパールおよびその他の開発途上国におけるモニタリング・評価体制と評価人材の育成の現状と課題 （廣野良吉） Rm.600	共通論題5 行政評価のアカウンタビリティー指定管理者評価の功罪 （山谷清志） Rm.201AB	

（　　）は座長です。

日本評価学会　春季第11回全国大会
「評価の国際交流」プログラム詳細

2014年5月31日（土）

9:30-10:00　受付

午前の部　10:00-12:00

共通論題１　学校評価の評価者育成プログラム～現場に必要な評価者を育てる取組の比較検討～　Rm. 600

座長　橋本　昭彦　国立教育政策研究所

K1-1	学校を経営する上で役に立つ学校評価を実施する為に必要なこと		石田　楓軒	日本評価学会　学校専門評価士
K1-2	学校評価の評価者育成プログラム ～教育委員会における研修の実施～		橋本　昭彦	国立教育政策研究所
K1-3	学校評価の評価者育成プログラム ～学校内での教員に対する研修の実施～		池田　琴恵	東京福祉大学
K1-4	学校評価の評価者育成プログラム ～英国のInspectorsの養成を中心に～		植田　みどり	国立教育政策研究所

共通論題４　国際防災評価　Rm. 201AB

座長　林　薫　　文教大学
座長　和田　義郎　国際協力機構

K4-1	ODA「防災協力イニシアティブ」の評価と今後の防災評価の課題		林　薫	文教大学
K4-2	バングラデシュのサイクロン被害と避難態様に関する研究		日下部　尚徳	大妻女子大学
K4-3	コミュニティ防災協力の効果と課題 ～ソロモン諸島での事例～		三村　悟	福島大学 うつくしまふくしま未来支援センター

自由論題１　自治体評価の諸相　Rm. 202AB

座長　小野　達也　鳥取大学

J1-1	地方議員の議会での質問に対する年間評価		本田　正美	東京大学大学院
J1-2	地方自治体における行政評価に関する提案		谷口　敏彦	行政経営アナリスト
J1-3	自治体評価の向上～評価担当者の視点から～		高千穂　安長	アジアサイエンスカフェ

午後の部Ⅰ　13:15-15:15

共通論題２　自治体戦略の協働マネジメントとプログラム評価～自治体における政策評価の新たな試み～　Rm. 600

座長　源　由理子　明治大学

K2-1	「自治体戦略の協働マネジメント」実践のねらい		北大路　信郷	明治大学
K2-2	協働マネジメントにおけるプログラム評価の導入 ～マネジメントと評価の融合を探る～		源　由理子	明治大学
K2-3	プログラム評価における社会調査の活用可能性 ～「協働型社会調査」の試み～		米原　あき	明治大学
K2-4	自治体戦略の協働マネジメントにおけるプログラム評価の実践		真野　毅	兵庫県豊岡市

自由論題２　評価手法　Rm. 201AB

座長　牟田　博光　東京工業大学

J2-1	道路の問題点は何か調査結果		大島　章嘉	市民満足学会
J2-2	大学のガバナンスと経営効率性	○	山﨑　その	京都外国語大学
			伊多波　良雄	同志社大学
			宮嶋　恒二	京都学園大学
J2-3	コミュニティ保健介入が妊産婦の健康希求行動に与える影響～ バングラデシュ母性保護サービス強化プロジェクトの事例～	○	青柳　恵太郎	国際協力機構
			神谷　祐介	龍谷大学
J2-4	研修の波及効果についての費用便益分析 ～JICA課題別研修を事例として～	○	菊田　怜子	フリーランスコンサルタント
			岸本　昌子	（一財）日本国際協力センター

自由論題３　参加型評価　Rm. 202AB

座長　高千穂　安長　アジアサイエンスカフェ

J3-1	効果的プログラムモデル形成のための実践家参画型評価アプローチ法の開発（その3） ～「実践家参画型ワークショップの活用」に注目して～	○	新藤　健太	日本社会事業大学
			大島　巌	日本社会事業大学
			大山　早紀子	日本社会事業大学
			方　真雅	日本社会事業大学
			高野　悟史	日本社会事業大学
			鈴木　真智子	日本社会事業大学
			宇野　耕司	目白大学
J3-2	モスト・シグニフィカント・チェンジ（MSC）の実践 指標を用いない参加型手法～バングラデシュでの実践からの教訓		田中　博	参加型評価ファシリテーター
J3-3	ロジック･モデルを活用した市民参加型意思決定システム ～東海市を事例として～		海川　能理子	名古屋大学

午後の部Ⅱ　15:30-17:30				
共通論題３　ネパールおよびその他の開発途上国におけるモニタリング・評価体制と評価人材の育成の現状と課題 座長　廣野　良吉　成蹊大学 パネリスト　湊　直信　国際大学 パネリスト　佐々木　亮　(株)国際開発センター				Rm. 600
K3-1	Capacity Building in M&E _ Challenges and Lessons within the International Development Context		大内　文香	国連開発計画
K3-2	Current Practice and Issues in the Development of Monitoring and Evaluation System and Evaluation　Professionals in Nepal		Dilip Kumar Chapagain	ネパール政府国計画委員会
共通論題５　行政評価のアカウンタビリティー指定管理者評価の功罪 座長・司会・コメンテータ　山谷　清志　同志社大学				Rm. 201AB
K5-1	「男女共同参画政策の推進に向けた評価に関する調査研究」の結果より	○	内藤　和美	芝浦工業大学
			山谷　清志	同志社大学
			高橋　由紀	国立女性教育会館
K5-2	公務との均等待遇と評価～NPO「活動者」を手がかりに～		渋谷　典子	NPO参画プラネット

共通論題セッション報告

共通論題セッション１「学校評価の評価者育成プログラム ～現場に必要な評価者を育てる取組の比較検討～」

座長　橋本 昭彦（国立教育政策研究所）

全国4万にも及ぶ学校や幼稚園で学校評価が義務化されているが、学校評価を学校の組織改革や運営改善のために実効性ある形で普及させる鍵として標記の論題を掲げた。本学会の「学校評価分科会」の2月例会においても同趣旨の勉強会を開催したが、本セッションはその議論を発展させたものであった。

セッションでは、最初に石田楓軒会員が公立学校における長年の経験や過年度の本学会での報告の蓄積をもとに、実効性ある学校評価を実施する為に必要な学校経営上の要素について提案を行った。ついで橋本が、教育委員会における教職員研修の内から学校評価の評価者育成に関わるプログラムの実施状況を説明した。3番目に池田会員が、学校内における評価者の育成例としてGTO（Getting to Outcomes）の考えに基づく自らの実践を紹介した。4番目に植田会員が、評価者育成の先行事例であるプログラム英国の学校評価におけるInspectorsの養成について、国際比較の視座を交えた報告を行った。

石田健一氏のコメントの一部を借りて各報告の特長を寸評すれば、第一報告は実際的で分かりやすく、第二報告は教委対象の珍しい調査であり、第三報告は報告者＋学校＋教職員の三位一体による貴重な事例、第四報告は認定要件が厳しいとみえる英国の注目すべき事例だ、と言えそうである。

フロアからは各報告に対する質問がたくさん出て、学校評価における評価者養成には多様な側面があることが浮き彫りになったほか、日本の制度作りや実践の確立に外国の事例をどのように生かすのかという重い関心がクローズアップされた。他方で、幅ひろい対象を検討するあまり、評価者養成プログラムの在り方の考察というセッション本来の目的は掘り下げる時間がなくなってしまった。得られた成果と課題を次回以降に生かしたい。

共通論題セッション２「自治体戦略の協働マネジメントとプログラム評価 ～自治体における政策評価の新たな試み～」

座長　源 由理子（明治大学）

自治体における事務事業評価の限界が指摘される中、重要政策が着実に成果をあげることができるように、戦略レベルのアウトカムを評価し、行政の政策とその遂行が持続的に改善される状態を実現する必要性があるのではないか。このような問題認識のもと、本共通論題セッションでは、自治体における政策評価の新たな試みについて、兵庫県豊岡市の事例とともに検討・報告を行った。取り分け、近年自治体では、市民など多くの地域主体との協働によるVFMの高い公的サービスの供給を目指し、協働推進政策が取られており、多様なアクターを巻き込んだ協働マネジメントのあり方が問われている。それら政策の改善改革を行っていくためには、プログラム評価の方法論を使って継続的に政策評価を行うこと、特にセオリー評価、プロセス評価といった政策の計画、実施段階における評価を関係者と協働で行う仕組みが必要になると考えられる。本報告では、その取組を「協働型プログラム評価」と呼び、自治体戦略の協働マネジメントとプログラム評価を組み合わせた新たな政策評価制度と位置付けている。

第一報告者の北大路信郷氏からは、公共経営の観点から、市民協働、公共サービスの生産性向上におけ

る市民協働によるプログラム評価の有効性について発表があった。豊岡市の協働型プログラム評価、協働型戦略マネジメントの実践は、ガバナンス・マネジメント論の進展における貢献への期待が大きい。第二報告者の源は、協働型プログラム評価を評価論の観点から論じ、市民とともにロジック・モデルを使いセオリー評価とプロセス評価を継続的に実施することによる効果について発表を行った。

第三報告者の米原あき会員からは、協働型プログラム評価におけるアウトカムの指標化と達成状況の把握について、市民協働で社会調査を行う際の可能性と課題について発表があった。第四報告者の真野毅会員からは、協働型プログラム評価の制度化を推進した豊岡市副市長としての立場から、実際の取り組みの状況とその特徴、課題について発表があった。

これらの報告に対し、フロアからは、事務事業評価との比較、市民協働で行うことの「代表性」の問題、社会調査のあり方など、協働型プログラム評価の実践の諸側面について質問・意見が提示された。協働型プログラム評価は、現場の創発を重視し、地域のガバナンスを活かした戦略マネジメントシステムと評価の融合の一形態であり、豊岡市で2年間の試行・準備過程を経て、今年度から本格実施に入ったものである。今後もそれらの実践を踏まえ、協働型プログラム評価の理論化について、学会での発表を重ねていきたいと考えている。

共通論題セッション３「ネパールおよびその他の開発途上国におけるモニタリング・評価体制と評価人材の育成の現状と課題」

座長　廣野 良吉 （成蹊大学）
パネリスト　湊 直信 （国際大学）
パネリスト　佐々木 亮 （(株)国際開発センター）

第一報告者 : 大内 文香　氏　国連開発計画独立評価室評価顧問
第二報告者 : Mr. Dilip Kumar Chapagain　ネパール政府国計画委員会事務局M&E課長

1．本セッション趣旨：本セッションでは、ネパールおよびその他の開発途上国における評価人材の育成と評価体制の現状をJICAやUNDPの支援プロジェクトの実例を基に考察し、その主要な課題の解明と今後の支援体制の在り方について討論が集中した。
2．評価の現状：多くの途上国では、インド、フィリピン等の一部の例外を除いては、政府や地方自治体の開発政策、環境政策、社会福祉政策等に対する評価活動は、官民ともに長い間低調であった。しかし、近年国際社会の動き、特にOECD、国連や世界銀行、地域開発銀行等による評価人材の育成や評価体制の強化への支援、さらに「アジア太平洋評価協会」（APEA）の誕生に見るように、世界各地域での評価意識の向上、評価人材の育成も手伝って、徐々に変化がみられつつある。
3．評価における具体的な課題：我が国でも観察されたように、途上国でも生産される物品については、年季奉公、品質改善・管理というような語句に代表されるように、「評価の文化」は、職人たちの間で昔から導入されてきた。しかし、政策やプロジェクト形成、予算執行等社会システムの構築や運用面では、「評価の文化」は定着しなかったし、現在でも甚だ不十分である。その背後には、多くの途上国が相変わらず、中央集権国家であり、地方自治体体制は弱体であったり、さらに民間のNGOの社会的批判勢力が未だ不十分であることが最大の要因であろう。途上国における評価の具体的な課題としては、基調講演者両人および討論者お二人も再三指摘しておられたように、①評価を律する法制化が不備のため、政府・公的機関による評価政策の欠如と評価需要の低迷、②国会による行政監視・評価体制が不備、③行政機関における評価体制が不備、④評価結果の政策形成等へのフィードバック・活用が不十分、⑤評価に不可欠なベースライン・データ、統計作成体制の不備、⑥国内評価人材が不備、⑦評価人材育成機関が不備、⑧政府

政策・施策等の実施を監視する市民社会組織、そのネットワークや独立マスコミが未熟なために、評価の独立性に不備、⑨プロジェクト評価が大半，⑩ドナーによる一方的評価が多く、合同評価が少ない、⑪評価報告書の内容が専門的すぎ、利用者にとって判りにくい、⑫評価報告書・情報公開が不十分等があげられる。

4．評価の将来展望：ネパールやその他の途上国でも、経済成長の中で一人当たりの国民所得や生活水準の向上、教育の普及と教育水準の上昇、市民社会意識・運動が盛んになるに従い、社会システムへの評価の導入の重要性が徐々に認識されてきており、共産党一党独裁国家でさえ、その深みや広がりで、いわゆる政治的民主主義体制下の国々に対比すると劣るが、評価人材の育成、評価体制の強化が徐徐であるが、着実に進んできている。この点で、日本評価学会による評価人材育成プログラムと評価士認定制度は、多くの途上国にとって学ぶところが大であろう。途上国の大半は、現在この評価体制の初歩的ないし中位的発展段階にあるが故に、DAV加盟援助国や国際機関は、途上国のこの評価体制の強化を側面から加速化して、この内発的発展を支援していくことが望ましい。今後、中産階級の飛躍的台頭の中で、納税者意識の強化に伴い、さらに急増する財政赤字削減のための税制の近代化、徴税制度の改善、無駄な政府支出の削減、政治腐敗と政官癒着への社会的批判の激化への動きの中で、この評価体制の強化への動きは一層高まるであろう。

5．途上国の評価体制強化支援における注意事項：我が国の評価体制の発展過程でも経験したように、それぞれの社会には、特異な歴史的・文化的態様があり、政治的体制も異なるので、誕生している評価体制も異なる。自国のODA評価を既に30年以上実施している我が国はもちろん、主要ODA供与国や国際機関は、開発途上国政府・機関の代表を招いてODA評価ワークショップを開催し、途上国の評価体制の進展や評価人材の育成を支援したが、そこで学んだ多くの教訓に基づき、今後の評価分野での支援を増大して欲しい。

6．特記すべき事項：今回の日本評価学会春季第11回全国大会では、一方でJICAのご協力を得て、ネパールからの多くの評価関係者を本セッションにお招きし、現場からの声として、ネパール国家計画委員会評価部責任者唐基調講演をいただき、さらにUNDPからは日本評価学会設立時から参加頂いている日本人評価アドバイザーをお招きし、これら重要な評価に関する課題について問題提起の機会を戴いた。この場を借りて、日本評価学会国際交流委員会は、本セッションの組織者として、JICAならびにUNDPに対して心から謝意を表したい。

共通論題セッション４「国際防災評価」

座長　林 薫 （文教大学）
座長　和田 義郎 （国際協力機構）

阪神および東日本大震災を経て、防災協力は日本の経験が活かせる分野として防災協力は日本の国際協力の重点分野の一つになっている。しかし、日本の経験を生かすためには、日本国内の、あるいは日本が協力した防災政策、プログラム、プロジェクトの評価は不可欠である。本セッションは、バングラデシュおよびソロモン諸島の防災協力を通じて、防災協力の評価、あるいは評価を行っていく上での課題を論点とたした。

第一報告の『ODA「防災協力イニシアティブ」の評価と今後の防災評価の課題』（林薫）は外務省の第三者評価として実施された「ODA防災協力イニシアティブの評価」をについて紹介しつつ、今後2015年に開催される第3回国連防災会議やそこでのアジェンダとなる新イニシアティブに向けて、防災評価の課題を考察した。日本の防災イニシアティブがより国際的な訴求力を持つためには、東日本大震災をはじめとする過去の災害に防災事業がどの程度効果があったかという科学的な議論の積み上げが必要であることを

指摘した。

第二報告の『バングラデシュのサイクロン被害と避難態様に関する研究』（日下部尚徳）はバングラデシュのサイクロン高度危険地域355世帯における質問紙調査をもとに、住民が避難行動に移れない原因となり得るファクターを分析し、住民にとって大切な財産である食糧や家畜へのダメージは、生活再建の大きな障害となり、サイクロン被害による生活水準の低下は地域の災害脆弱性を高めることから、人命だけに焦点を絞ったこれまでのシェルター支援では、十分な効果は見込めないとしつつ、被害の軽減策に関して問題提起を行った。具体的には、事前の避難率をあげるためには、これまで論じてきたような住民の避難意思の妨げとなる家財を保護するための施策が必要である。加えて、シェルター建設といったハードの構造物支援だけでなく、避難訓練や防災教育などを通じて、避難行動に円滑に移れるようにするためのソフト面でのサポートが必要不可欠であり、例えば、家畜を高台へ移動するトレーニングや、食糧を高潮被害から守るために容器にいれて土に埋めるといった、その土地に見合った防災知識の普及をおこなう必要がある。

第三報告の『コミュニティ防災協力の効果と課題～ソロモン諸島での事例～』（三村悟）は2010年度から3年間にわたってJICAが防災担当機関の能力向上とコミュニティ住民の災害対応能力の向上を図る技術協力を実施してきたソロモン諸島のガダルカナル島で、2014年5月に発生した洪水において政府機関および地域コミュニティに対する支援の効果が実際の災害時にどのように発現されたかを論じた。この中で、これまでJICAが実施してきたほぼすべてのコミュニティ防災協力で、数か所のコミュニティでパイロットプロジェクトを実施してそれを周辺の地域に面的に拡大していくというアプローチがとられているが、そのような面的展開がプロジェクト期間中だけでなく終了後も受け入れ国側の自助努力によって図られてきたケースはどれだけあるのであろうかという疑問を呈しつつ、地域社会や国家の概念がごく最近まで希薄であり、数百人程度の小規模のコミュニティが近隣のコミュニティと緊張した関係性の中で日常生活を営んでいたような近隣との協力が難しい社会などがあることを踏まえ、国ごと地域ごとの社会的、文化的背景も踏まえたアプローチへの意識を強くすることが指摘された。

三件の報告とも、防災政策、プログラム、プロジェクトがどのように効果を発現しているのか、そこから今後の協力にどのようにフィードバックしていくかを共通のテーマとして、さらなる評価の必要性を強調する内容となった。質疑応答では防災に関する社会的なキャパシティーの向上などを中心に議論が行われた。

共通論題セッション５「行政評価のアカウンタビリティー指定管理者評価の功罪」

座長　山谷 清志（同志社大学）
討論者　高橋 由紀（国立女性会館）

男女共同参画社会基本法が1999年に成立し、地方公共団体の責務が9条に明記されている。「第九条　地方公共団体は、基本理念にのっとり、男女共同参画社会の形成の促進に関し、国の施策に準じた施策及びその他のその地方公共団体の区域の特性に応じた施策を策定し、及び実施する責務を有する」。ほぼ同時期に日本各地の地方公共団体、とくに三重県や静岡県、岩手県、秋田県、北海道などから評価ブームが起こり、はからずも男女共同参画政策と評価は流行のトピックとして普及した。ただし、ごく一部、たとえば内閣府の男女共同参画会議の専門調査会、一部の地方公共団体、そして日本評価学会をのぞき、自治体評価（行政評価）と男女共同参画政策を同じアジェンダで論じることはなかった。

このような中で地方公共団体の男女共同参画政策をめぐる評価には、多くの課題が出ている。問題事例や疑問視される活動が多いのである。すなわち、男女共同参画「政策」を評価する方法の無知、評価手法の選択ミス、行政がまちがった評価方法を指定管理者のNPOに押しつけ、その結果役に立たない評価が増加、評価をさせられる人びとの徒労感の蔓延などである。そこに事業仕分けやそれに類する公開レビュー・ヒアリングが行われ、指定管理者制度も対象にされた。

こうした状況の認識を共有した上で、第一報告者の内藤和美氏（芝浦工業大学）は、科学研究費補助金基盤研究C、課題番号235103520001の研究成果をふまえ以下の諸点について報告された。まず、各地方公共団体の男女共同参画所管課についてはアンケートや実態調査に基づいて、①男女共同参画担当課の体制と業務、②拠点施設との関係、③実施または受けている評価の全体像、④男女共同参画計画の進行管理の体制と流れ、⑤男女共同参画計画の進行管理における評価の主体･対象･内容･方法の概要、⑥評価における指標による量的評価と質的方法の組み合せ方、⑦評価における評価基準とその設定理由・根拠、⑧男女共同参画計画の進行管理における評価と拠点施設の評価の関係を明らかにした。これは実務担当者と一部の専門家が経験的に知り得ていたが、アカデミズム一般ではあまり知られていなかった事実である。

また、男女共同参画拠点施設に関しては、3つのタイプが観察された。拠点施設の事業が計画に深く包摂され、男女共同参画担当課が日常的に拠点施設と協働し、拠点施設事業を計画事業として密に管理するタイプ1〔松戸市、川崎市施設〕、男女共同参画計画の成果指標として拠点施設に係る指標を用いるタイプ2〔越谷市、静岡県、名古屋市施設、広島県、三重県〕、拠点施設が個別の計画掲載事業の担当部門となるだけでなく、事業に分解されない拠点施設機能全体を基本計画の施策以上の階層に位置付け、施策として拠点機能の発揮に取り組むタイプ3〔秋田県施設、さいたま市、静岡県、名古屋市施設、広島県、松戸市、三重県〕の3つである。

内藤氏によると、ここから男女共同参画計画の進行管理と拠点施設の機能管理における評価の性格が明らかになる。すなわち調査した24の評価活動のうち22活動は業績測定型評価で、対象は事業の進捗状況である。そこで使われる指標は、事業等実施数・開催数等のインプット指標、利用者数・参加者数のような「アウトプット指標」が中心で、その事業が施策・政策・計画の目標達成や施設使命の実現にどのように貢献できたのかを確かめるには甚だ限界がある。ただ、「受講料を助成した技能習得講座等受講者のうち受講後就労に結びついた女性人数」、「女性センター運営協議会からの改善につながった意見数」〔松戸市〕というように、例外的に事業次元の「政策指標」と解される指標が散見された。指標の成熟が期待されるが、それは一般の自治体評価と同じように難しい。

第二報告では渋谷典子氏（NPO参画プラネット／名古屋大学）が、NPOのトップとして男女共同参画関連施設の運営と評価に携わった経験から、重要な報告をされた。すなわち、「公務との均等待遇と評価～NPO『活動者』を手がかりに～」と題して、民営化や民間委託された業務について責任をもって引き継ぐシステムをどのように構築するか、である。「公務」がさまざまな形で民間部門の業務に転換されると、公と民の境界が曖昧となる。ここでは一方で公務員削減論へと結びつき、公務員の賃金や人事管理等に民間類似の手法が取り入れられ、人事管理に能力と成果主義を取り入れる動きが繰り返される。他方、単なる経費削減が動機で指定されたNPO・指定管理者をはじめとする男女共同参画施設関係者は「安使い」を繰り返され、意欲を失っていく。公務員とNPO関係者の男女共同参画政策に関わる「モラール」をどうやって維持できるのか、これが課題として提示されたのである。

ただ、NPOが指定管理者になるメリットもある。①会社、組織本位ではない運営ができる、②当事者である市民が、自身の権利基準を維持するための“サービスの担い手”となり、利用者（市民）として自らがサービスを選択し責任をもつような事業が実施できる、③評価士や大学院博士課程などのさまざまな形で専門家（エキスパート）としての能力を磨く機会を得られる。これらの点に関して渋谷氏は、希望を持たれていた。

以上の2報告に対して討論者の高橋由紀氏から、これもまた長年の研究と実務の経験をふまえた論点整理、視点を変えた問題提起がなされた。

ただ、男女共同参画「政策」と指定管理者制度、それを担うNPO活動においては、男女共同参画に関する政策のサブスタンスの実現についても、管理と経営についても、「アカウンタビリティ（結果について納得できるように説明できる責務）は開発途上ではないか」、この懸念が明らかになったのが本セッションの成果である。

自由論題セッション報告

自由論題セッション１「自治体評価の諸相」

座長　小野 達也（鳥取大学）

本セッションでは、「自治体評価の諸相」というタイトルのもと、地方議員の活動の評価、自治体の行政評価手法及び自治体の評価担当者の意識という、文字通り自治体に関する評価の諸相を論ずる3本の報告が行われた。

まず、最初の報告となる本田正美氏（東京大学大学院情報学環交流研究員）の「地方議員の議会での質問に対する年間評価」では、ある市議会議員の1年間の質問を検証評価した結果が紹介された。これは同じ議員の一回の議会における質問の評価であった第13回全国大会での報告に続くものであり、地方議員の活動全般を評価するための1ステップと位置づけられるものである。質疑では、質問を評価する尺度や、今回の個別特定議員に関する調査を一般化するための方策などについて質問がなされた。

2番目の報告は、谷口敏彦氏（行政経営アナリスト）の「地方自治体における行政評価に関する提案」であった。本報告は、東京都の中野区・練馬区・豊島区の行政評価制度を取り上げ、米国農務省の戦略計画と年次報告、フランス政府の農業教育に関する評価を参照した上で、自治体の行政評価において期待されるアカウンタビリティとマネジメントの2つの機能の改善に向けて、①計画策定との一体化、②長期目標の明確化、③達成目標の再検討可能性、④マネジメントに活用できる指標整備、⑤評価項目の選択と集中、⑥事業設計段階での必要性等の見極め、⑦外部評価の力点の移行という7点の提案がなされた。会場からは、自治体の内部評価・外部評価の現状を踏まえ、多くのコメントや質問が提出されている。

最後の報告である、高千穂安長氏（アジアサイエンスカフェ）の「自治体評価の向上～評価担当者の視点から～」では、自治体の評価能力向上のためには評価担当者が評価業務を通じて満足を得てやる気を増すことが重要であるという観点から、川崎市の評価担当職員を対象に2008年に実施されたアンケートの分析に基づき、評価担当者が理解される評価内容を作成し、結果として満足をえるための要因について報告された。質疑では、サービス業における従業員満足と顧客満足の関係に基づくモデルの公共部門への適用に関する質問などが提出された。

本セッションでの3本の報告の多面性は、地方行政・自治体経営における評価の役割、及びそれらを論ずる評価研究の多様性を反映しているともいえよう。各研究の今後の進展が期待されるところである。

自由論題セッション２「評価手法」

座長　牟田 博光（東京工業大学）

「自由論題セッション2　評価手法」の分科会では、様々な評価手法に関する4研究の報告が行われた。

大島章嘉氏の研究「道路の問題点は何か調査結果」は、道路の問題点は何かを235名に予備調査し、その結果に基づきアンケート調査票を作成した。次に、全国20才以上の977名にアンケート調査を実施し、得られた結果を集計分析した。さらに、問題ない度、満足度、信頼度を被説明変数に、様々な不満の理由を説明変数にして回帰分析を行い、どのような不満が、問題ない度、満足度、信頼度に影響を与えるかを示した。

山﨑 その、伊多波良雄、宮嶋恒二氏の研究「大学のガバナンスと経営効率性」は、大学経営効率化に

関し、国公私立大学に対して行ったアンケート調査で得られた194大学のデータを用いて、包絡分析によって大学経営の効率性評価を行った。さらに、外部要因の影響も考慮にいれて、ガバナンスやマネジメントと効率性の関係について、構造方程式モデリングにより分析を行った結果を示した。

青柳恵太郎、神谷祐介氏の研究「コミュニティ保健介入が妊産婦の健康希求行動に与える影響：バングラデシュ母性保護サービス強化プロジェクトの事例」は、JICAがバングラデシュで行った「母性保護サービス強化プロジェクト」の効果を検証するために、プロジェクトが独自に実施したサーベイ調査、および、既存の大規模調査であるDHSを結合したデータを用いて、インパクト評価を実施した結果を示した。

菊田怜子、岸本昌子氏の研究「研修の波及効果についての費用便益分析～JICA課題別研修を事例として～」は、JICAが実施した2010年度課題別研修受講者に質問紙調査を行い、191名から回答を得、研修で得た知識・技術を帰国後第三者に伝播する行動を通じて生じるインパクトを便益で測り、研修のために必要であった費用と比較することによって、研修の効率性を評価した。さらに感度分析によって、伝達行動の活性化の重要性を示した。

各発表の後、熱心な質疑応答が行われた。評価対象、時期などに応じて、評価手法も様々に工夫されており、今後もさらに、様々な分野で、その分野の課題解決に資する評価手法の研究、発表が重ねられることが期待される。

自由論題セッション３「参加型評価」

座長　高千穂 安長 （アジアサイエンスカフェ）

本セッションの目的は、実務家や市民が評価に参加し、より適切、的確な評価を実施するための方策について研究することであった。参加の当事者や参加方法、利用する手段等に対して内外で行われた新たな知見について以下の3件の発表があった。

2時間の枠の中で3本という発表数とコメンテータを置かなかったことから時間に余裕があり、フロアとも十分な質疑応答の時間が確保できた。

1.「効果的プログラムモデル形成のための実践家参画型評価アプローチ法の開発（その3）:「実践家参画型ワークショップの活用」に注目して」（新藤健太、大島巌、大山早紀子、方真雅、高野悟史、鈴木真智子、宇野耕司）は、実践家が積極的に評価活動に参画することを促進するためのアプローチであるCD-TEP法のうち、実践家との協働作業の中で効果モデルの形成や改善などの評価の場としての実践家参画型ワークショップに着目し、事例を通してその効果を検証した。

ワークショップの効果として、グループダナミクスの発生、オーナーシップの向上、評価を意識した目標の設定行動など良い効果が見られたとしている。なお課題として、ファシリテーターの質的・量的確保を挙げている。

2.「モスト・シグニフィカント・チェンジ（MSC）の実践　指標を用いない参加型評価手法～バングラデシュでの実践からの教訓～」（田中博）は、モスト・シグニフィカント・チェンジ（MSC）について概要を説明した後、その特徴を3点明示し、バングラデシュでの実践結果について報告した。

物語として参加者に話し理解を進めさせるため、読み書きができない参加者も理解がされるメリットがあり、現場で発生する複雑な変化を把握・分析して教訓を学ぶ手法として優れていることを示した。今後は、プロジェクト開始段階から定期・継続的に実施し、担当スタッフの研修が望まれるとした。

3.「ロジック・モデルを活用した市民参加型意思決定システム～東海市を事例として～」（海川能理子）は、行政を取巻く諸般の状況から、市民のニーズに応答的な成果志向の行政運営の実現が重要であり、そのために、ロジック・モデルを活用し、行政運営の意思決定に市民参加を取り入れる事が重要とし、それにより市民のニーズに応答的で分野横断的な政策が実現可能になる事を、予算編成への影響を主題として

東海市を事例として明らかにした。

今後はロジック・モデルの活用について、ロジック・モデルの習得は時間がかかること、市民の入れ替わりなどあることに対する対応が望まれるとした。

以上の3報告は、いずれも今後の評価に対する新しい貢献を予見させる、示唆に富むものであり、今後のさらなる研究が望まれる。

企画委員会

企画委員会は、日本評価学会が行う各種の活動を企画、実行することをとおして、わが国の評価学の普及と発展に寄与することを目的としています。最近の活動としては以下の活動があります。

１．評価士養成・認定

公共事業、学校教育、福祉、ODA等をはじめとする様々な分野の政策や事業に関するアカウンタビリティや効率化が求められており、こうした社会一般からの需要に応えるために、評価システムの導入や事業評価の実施が進められていますが、評価の専門性を持つ人材は未だ十分に育っていないのが現状です。そこで、評価に関する専門的能力を身につけた人材を養成してその能力を認定し、各分野における評価の向上に資することが目的です。

評価士はそのレベルによって初級、中級、上級に区分しています。それぞれのレベルの考え方は次のとおりです。

初級レベル（評価士）：評価に関する幅広い知識と技法を身につけた専門家。

中級レベル（専門評価士）：各専門分野（セクター）の評価に関する深い知識と技法を身につけた専門家。

上級レベル（上級評価士）：評価に関する体系的な研究を行うことにより、評価活動の発展に寄与することができる専門家。

平成25年度後半～26年度前半には、初級レベルの第15、16期評価士養成講座を開講し、計58人が受講されました。上級レベルとしては第3期上級評価士認定審査を行い、2人が申請されました。

２．顕彰事業

日本評価学会では年一回、学会賞の選定という形で会員の顕彰を行うことにしています。学会賞の受賞者は企画委員会の下に設けられている学会賞審査委員会によって推薦し、理事会で決定されます。学会賞としては、本学会誌に掲載された原著論文等が評価学研究に大きく貢献したと認められる者に授与される「日本評価学会論文賞」、評価学研究の進歩に寄与する優れた研究をなし、将来の発展を期待される者に授与する「日本評価学会奨励賞」、評価学の発展に顕著な功績があった者に授与される「日本評価学会功績賞」の3つの賞があります。

平成26年度は10月15日まで受賞者の推薦を受け付けています。詳細は学会ホームページに記載してあります。多くの自薦、他薦を期待しています。受賞者は所定の手続きを経て決定され、11月15、16日に大阪大学で開催予定の第15回全国大会の際に表彰されます。

企画委員会委員長　牟田　博光

国際交流委員会

国際交流委員会では海外の評価専門家や評価学会との交流や意見交換を通じて相互に学習を行っている。2013年10月から2014年9月の1年間に以下の活動を行った。

1．ネパール国サンクー市における固形廃棄物処理対策事業を通じて、ネパール評価協会（Nepal Evaluation Society）の評価能力育成を支援してきた。

2．UNDP評価部長講演会
2013年10月29日、国際協力機構（JICA）本部にて国連開発計画（UNDP）評価部長、Indran Naidoo氏を招聘し、講演会を実施した。

3．Asia-Pacific Evaluation AssociationのBoard Meeting開催
2014年5月30日スカイプによりAPEAのBoard Meetingが開催された。参加者は以下の通り。
会長　廣野良吉
副会長　Romeo Santos
事務局長 Aru Rasappan
理事 Laura Luo
理事 Champak Pokharel
理事 Subrato Mondal
議題は以下の通り。
A　開会の挨拶（会長）
B　APEA進展の報告（事務局長）
C　APEA財務報告
D　2014年3月26日クアラルンプールでのAPEA非公式会合のレビュー
E　マレイシア評価学会のコンファレンスとAPEA
F　理事会の選挙のレビューと決定
G　APEA代表のIOCE　Boardへの立候補。
H　2014年APEAの活動計画のレビューと決定
I　その他

4．春季全国大会で国際セッションを開催
2014年5月31日JICA地球ひろばにて開催された春季第11回全国大会において国際機関と開発途上国政府からそれぞれ評価を担当しているパネリストを招待し、国際セッション「ネパールおよびその他の開発途上国におけるモニタリング～評価体制と評価人材の育成の現状と課題～」を開催した。
議長：廣野良吉　成蹊大学名誉教授
発表（1）：大内文香　国連開発計画評価室評価顧問
発表（2）：Dilip Kumar Chapagain　ネパール政府国家計画委員会事務局M$E課長
コメント（1）：湊　直信　国際大学客員教授
コメント（2）：佐々木亮　国際開発センター主任研究員、立教大学講師
議論：全ての参加者、パネリスト間で質疑応答、議論が行われた。

国際交流委員会委員長　湊　直信

広報委員会

１．活動方針

広報委員会は、その目的である、「評価の意義を広く普及し、学会活動の強化を支援するために、日本評価学会および学会の行う種々の事業を広報し、学会の情報活動に寄与する」ことを目的とします。評価学会の活動のみならず広く評価に関わる情報も会員の皆さんと共有すべくあらゆる機会を捉えてその実現を目指します。

２．活動状況

(1)「学会活動の強化と学会の健全な運営を目指して会員獲得に努めます」
会員獲得のための学会説明などのツールについては、ホームページを第一に考えています。
その効果を高めるため、文字情報とともに、写真など視覚情報も充実していきます。
また、自治体、企業などの説明ツールとしては、マス・メディアを使ったパブリシティも利用する方向で検討を進めていきます。

(2)「広報の手段は主として学会ホームページを活用します。従って、学会からのおしらせはホームページに随時掲載します」。
引き続き、より読みやすく、タイムリーに情報をお伝えするよう検討していきます。

３．今後の方針

広報委員会の役割について見直し、進捗状況管理、新企画の検討などを行い広報面での有効性、効率性を高めて行く所存です。

広報委員会委員長　高千穂　安長

日本評価学会第15回全国大会のご案内

第15回全国大会
実行委員長　齊藤　貴浩

来たる2014年11月15日（土）－16日（日）に日本評価学会第15回全国大会が大阪大学吹田キャンパス（大阪府吹田市）にて開催されます。

プログラム等詳細につきましては、学会メーリングリストおよびホームページにて随時ご案内いたします。

会員各位におかれましては奮ってご参加くださいますようよろしくお願い申し上げます。

記

1．　日程：2014年11月15日（土）－16日（日）

2．　場所：大阪大学吹田キャンパス
〒565-0871 大阪府吹田市山田丘1-1

以上

日本評価学会誌刊行規定

2005.2.15改訂
2002.9.18改訂
2001.9.9改訂

（目的および名称）

1. 日本評価学会（以下、「学会」という）は、評価に関する研究および実践的活動の成果を国内外の学界をはじめ評価に関心をもつ個人および機関に広く公表し、評価慣行の向上と普及に資することを目的として、「日本評価研究（仮名）」（英文仮名：“The Japanese Journal of Evaluation Studies”、以下、「評価研究」という）を刊行する。

（編集委員会）

2. 「評価研究」の編集は、後で定める「編集方針」にもとづいて編集委員会が行う。
3. 編集委員会は、学会会員20名以内をもって構成し、委員は学会理事会が選任する。編集委員の任期は2年とし、再任を妨げないものとする。
4. 編集委員会は、互選により委員長1名、副委員長2名および常任編集委員若干名を選出する。
5. 編集委員会は、最低年1回編集委員会を開き、編集方針、編集委員会企画、その他について協議するものとする。
6. 編集委員会は、その活動等について、随時理事会へ報告し、承認を受けるとともに、毎年1回学会年次大会の場で、過去1年の活動成果と翌年の活動計画に関する報告を行う。
7. 委員長、副委員長および常任編集委員は、常任編集委員会を構成し、常時、編集実務に当たる。

（編集方針）

8. 「評価研究」は、原則として、年2回刊行する。
9. 「評価研究」の体裁は、B5版とし、和文又は英文とする。
10. 「評価研究」に掲載する原稿（以下「論文等」という）の分類は、以下の5カテゴリーからなるものとする。
    （1）総説
    （2）研究論文
    （3）研究ノート
    （4）実践・調査報告
    （5）その他
11. 「評価研究」への投稿有資格者は、学会会員および常任編集委員会が投稿を依頼した者とする。学会会員による連名での投稿および学会会員を主筆者とする非会員との連名での投稿は、これを認める。編集委員による投稿はこれを認める。
12. 投稿原稿を上記分類のどのカテゴリーとして扱うかは、投稿者の申請等をもとに常任編集委員会が、下記の「作業指針」に従って決定する。
    （1）「総説」は、評価の理論あるいは慣行について概観する論文とし、その掲載については編集委員会が企画・決定する。
    （2）「研究論文」は、評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文とし、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
    （3）「研究ノート」は、「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
    （4）「実践・調査報告」は、評価事業の実践あるいは評価にかかわる調査の報告で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。

(5)「その他」には、編集委員会が独自に企画する特集に掲載する依頼原稿や学会誌の刊行に関する編集委員会からの学会会員への連絡等が含まれる。

13. 論文等は2名の査読者により査読することとし、その人選は編集委員会が行う。「研究論文」については、査読結果と編集委員会が査読者とは別に指名する担当編集委員1名の参考意見をもとに、編集委員会が掲載に関する決定を行う。「総説」、「研究ノート」、「実践・調査報告」および「その他」の論文については、査読結果にもとづき編集委員会が掲載に関する決定を行う。

14. 編集委員が「評価研究」に投稿した場合には、当該委員はその投稿に係わる常任編集委員会あるいは編集委員会の議事に一切参加しないものとする。

15. 上記いずれのカテゴリーの投稿についても、常任編集委員会による掲載の判断は可・不可の二者択一で行うこととする。但し、場合によっては編集委員会の判断で、小規模の修正による掲載も認める。「研究論文」としての掲載が適当でないと判断された場合でも、投稿者が希望すれば、常任編集委員会は「研究ノート」あるいは「実践・調査報告」としての掲載を決定できる。

（投稿要領の作成公表）

16. 編集委員会は、上記の編集方針にもとづき投稿要領を作成し、理事会の承認を得て、広く公表する。

（配布先）

17. 「評価研究」は、学会会員に無償で配布するほか、非会員に有償で提供する。

（抜刷の配付）

18. 「評価研究」掲載論文等の抜刷り30部を、投稿者（原著者）に無料で配布する。それ以上の部数を希望する場合は投稿者（原著者）の自己負担とする。

（インターネット上の公開）

19. 「評価研究」掲載論文等は、投稿者（原著者）の了承を得て全文をインターネット上で公開する。

（著作権）

20. 「評価研究」に掲載された論文等の著作権は各投稿者（原著者）に帰属するものとし、編集権は本学会に帰属するものとする。

（事務局）

21. 「評価研究」編集及び配布の事務は、それに関連する会計も含めて学会事務局が担当する。

（以上）

『日本評価研究』投稿規定

2008.9.29改訂
2003.4.18改訂
2002.3.25改訂
2001.9.9改訂

1. 『日本評価研究』（The Japanese Journal of Evaluation Studies）は、評価に関する論文、論考、調査報告等を掲載する。

2. 『日本評価研究』は、会員間の研究成果交流の場を提供し、内外における評価研究の一層の発展に資することを主目的として発行されており、原則として会員による寄稿を掲載する。なお、依頼原稿を除き、ファーストオーサーは学会員でなければならない。また、投稿は、一時に一原稿に限るとともに、他学会誌などへ二重に投稿などのない未発表のものとする。

3. 投稿された原稿は、編集委員会の責任において審査を行ない、採否を決定する。審査にあたっては、1原稿毎に2名の査読者を選定し、査読結果を参考にする。（査読者には、投稿者名を伏せて査読を依頼する。）

4. 原稿料は支払わない。

5. 『日本評価研究』に掲載された論文等は、その全文をインターネット上の本学会のホームページに掲載する。

6. 投稿にあたっては、投稿原稿が、①研究論文、②総説、③研究ノート、④実践・調査報告、⑤その他のうち、どのカテゴリーに入るかを明記する。ただし、カテゴリーについての最終判断は、編集委員会で行なう。「研究論文」は評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文、「総説」は、評価の理論あるいは慣行について概観する論文、「研究ノート」は「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考、「実践・調査報告」は評価事業の実践あるいは評価にかかわる調査の報告、「その他」は編集委員会が独自に企画する特集に掲載する依頼原稿等である。

7. 投稿方法
   （1）使用言語は日本語又は英語とする。
   （2）著者校正は原則として第一校までとする。
   （3）英文原稿については、ネイティブスピーカーによる英文チェックを済ませ、完全な英文にして投稿すること。
   （4）ハードコピー4部（A4版）を提出する。その際、連絡先（住所、Tel、Fax、Email）と原稿の種類を明記すること。掲載可と判断された原稿については、必要なリライトを経た後に、最終原稿のハードコピー2部とDOS/Vフォーマットのフロッピーを用いたTEXTファイルを提出する。その際、オリジナル図表を添付すること。
   （5）刷り上がりは最大14ページとする。これを超える場合は、その経費は著者負担とする。

（6）日本語原稿の最大文字数は以下のとおり。①研究論文20,000字、②総説15,000字、③研究ノート15,000字、④実践・調査報告20,000字、⑤その他適宜。それぞれ和文要旨を400字程度、英文要旨を150words程度、及び和文・英文でキーワード（5つ以内）を別に添付する。印刷は1ページ、20字×43行×2段（1,720字）とする。20,000字の原稿の場合、単純計算では英文要旨1ページを加えて合計13ページとなるが、図表の量によっては、それ以上のページ数となり得るので、注意すること。

（7）英文ではA4版用紙に左右マージン30mmをとり、10ポイントフォントを使用し、1ページ43行のレイアウトとする（1ページ約500words）。論文冒頭に150words程度のAbstractをつける。14ページでは、7,000words相当になるが、タイトルヘッド等を考慮して、最大語数を約6,000words（図表、注、文献込み）とする。図表の量によっては、ページ数が予想以上に増える場合もあり得るので、注意すること。

8.　送付先

〒140-0002　東京都品川区東品川4－12－6　日立ソリューションズタワーB 22F
一般財団法人 国際開発センター内 日本評価学会事務局
TEL: 03-6718-5931
FAX: 03-6718-1651
E-mail: jes.info@evaluationjp.org

『日本評価研究』執筆要領

2002.9.18改訂
2002.3.25改訂

1．本文、図表、注記、参考文献等

（1）論文等の記載は次の順序とする。

日本語原稿の場合

第1ページ：表題、著者名、所属先、E-mail、和文要約（400字程度）、和文キーワード（5つ以内）
第2ページ以下：本文、謝辞あるいは付記、注記、参考文献
最終ページ：英文表題、英文著者名、英文所属先、E-mail、英文要約（150words程度）、英文キーワード（5つ以内）

英文原稿の場合

第1ページ：Title; the author's name; Affiliation; E-mail address; Abstract (150 words) ; Keywords (5 words)
第2ページ以下：The main text; acknowledgement; notes; references

（2）本文の区分は以下のようにする。

例１（日本語）

１．
（1）
①
（2）
（3）

例２（英文）

1.
1.1
1.1.1
1.1.2

（3）図表については、出所を明確にする。図表は原則として、筆者提出のものをそのまま写真製版するので、原図を明確に作成すること。写真は図として扱う。

例１：日本語原稿の場合

図１　○○州における生徒数の推移

（注）
（出所）

表１　○○州における事故件数

（注）
（出所）

例２：英文原稿の場合

Figure １　Number of Students in the State of ○○

Note:
Source:

Table １　Number of Accidents in the State of ○○

Note:
Source:

（4）本文における文献引用は、「‥‥である（阿部 1995、p.36）。」あるいは「‥‥である（阿部 1995）。」のようにする。英文では、(Abe 1995, p.36) あるいは(Abe 1995)とする。

（5）本文における注記の付け方は、（‥‥である[1]。）とする。英文の場合は、（‥.[1]）とする。
（6）注記、参考文献は論文末に一括掲載する。

注記

1 ‥‥‥‥。

2 ‥‥‥‥。

（7）参考文献は、日本語文献は著者の五十音順、外国語文献は著者のアルファベット順に記し、年代順に記載。参考文献の書き方については以下のようにする。

日本語単行本：著者（発行年）『書名』、発行所

（例）日本太郎（1999）『これからの評価手法』、日本出版社

日本語雑誌論文：著者（発行年）「題名」、『雑誌名』、巻（号）：頁－頁

（例）日本太郎（1999）「評価手法の改善に向けて」、『日本評価研究』、1（2）：3-4

日本語単行本中の論文：著者（発行年）「題名」、編者『書名』、発行所、頁－頁

（例）日本太郎（2002）「行政評価」、日本花子『評価入門』、日本出版社、16-28

複数の著者による日本語文献：著者・著者（発行年）『書名』、発行所

（例）日本太郎・日本花子（2002）『政策評価』、日本出版社

英文単行本：著者 （発行年）. *書名*. 発行地：発行所.

（例）Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6th edition.* Beverly Hills, Calif: Sage Publications.

英語雑誌論文：著者（発行年）. 題名. *雑誌名*, 巻（号）, 頁－頁.

（例）Rossi, P. H. (1999). Measuring social judgements. *American Journal of Evaluation,* 15(2), 35-57.

英語単行本中の論文：著者（発行年）. 題名. In 編者 (Eds.), *書名*. 発行地：発行所, 頁－頁.

（例）DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco, Calif: Jossey-Bass, 177-196.

2名の著者による英語文献：姓, 名, and 姓, 名（発行年）. *書名*. 発行地：発行所.

（例）Peters, T., and Waterman, R.(1982). *In Search of Excellence: Lessons from America's Best Run Companies.* New York: Harper & Row.

3名以上の著者による英語文献：姓, 名, 姓, 名, and 姓, 名（発行年）. *書名*. 発行地：発行所.

（例）Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement.* Washington: Urban Institute.

（注１） 同一著者名、同一発行年が複数ある場合は、 (1999a)、(1999b) のようにa,b,cを付加して区別する。

（注２） 2行にわたる場合は2行目移以降を全角1文字（英数3文字）おとしで記述する。

『日本評価研究』査読要領

日本評価学会　『日本評価研究』編集委員会
2005年9月10日決定

1. 本査読要領の趣旨

本査読要領は、『日本評価研究』における掲載論文等の審査の要である査読手続きについて、投稿する会員及び査読を依頼される会員に対して解説を行い、審査手続きを効率的かつ効果的に行うことを目的として、定めるものです。

2. 査読の目的と投稿者の責任

査読は、投稿原稿が『日本評価研究』に掲載される論文等としてふさわしいものであるか否かについての判定を当編集委員会が行う上で必要とされるものです。

査読に伴って見いだされた疑問や不明な事項について、必要な場合は修正意見をつけて、修正を求めることがあります。査読は、その意味で、投稿原稿の改善に資するものでもあります。ただし、修正が求められた場合においても、論文等の内容に関する責任は著者が負うべきものであり、査読者の責に帰するものではありません。

査読者は2名で、編集委員会において学会会員の中から当該分野の専門家を選び依頼されますが、学会会員以外に依頼することもあります。

3. 査読の視点

査読は、以下の5つの視点によりますが、投稿原稿の種類によって、重点が違います。

（1）テーマの重要性・有用度
（2）研究の独自性
（3）論理の構成
（4）実証法・方法論の妥当性
（5）評価理論・実践への貢献

・研究論文の査読については、上記の5項目全てに配慮する。
・研究ノートの査読については、上記5項目のうち、特に（1），（2），（3），（4）の諸項目に配慮する。
・実践・調査報告の査読については、上記5項目のうち、特に（1），（3），（5）の諸項目に配慮する。
・総説の査読については、上記5項目のうち、特に（3）と（5）の諸項目に配慮する。

4. 投稿に当たっての留意点

2. に掲げた査読の視点以外に、基本的な論文の完成度の問題があります。例えば、

- 論文等として体裁が整っているか、
- 執筆要領にしたがっているか、
- 簡潔明瞭に記述されているか、
- 実証的なデータは適切に位置づけられているか、
- 注や参考文献は本文と対応しているか、
- 専門用語の使用は適切か、
- 語句や文法的な誤りがないか、
- 誤字脱字はないか、
- 句読点に誤りはないか、
- 英文要約などの英文表現は適切か、（必ずしも和文要約の直訳である必要はなく、英文としてまとまっていること）
- 字数は規程に従っているか、

など、内容及び形式に関する留意点があります。

大学院生及び実務家の投稿において、論文としての体裁が整わないまま送付されている例があり、査読そのものに至らないものもあるので、しかるべき指導を受けた後に投稿されるよう強く勧めます。

5. 査読にあたっての判断事例

(1) 完成度において不十分であるが掲載を考慮できる場合

萌芽的な研究、発展が期待できる論文等は評価論の発展のためにできるだけ評価してください。

- 検証は十分とはいえないが、理論や定式化が学問の発展に有用である。
- 考察は十分とはいえないが、新たな理論の形成・促進に有用である。
- 文献調査は十分とはいえないが、研究の位置づけは明確である。
- 比較研究は十分とはいえないが、適用例としては意義がある。
- 考察は十分とはいえないが、社会的、または、歴史的に重要な事例の評価として意義がある。
- 考察は十分とはいえないが、特定の社会活動の評価として意義がある。
- 論文の構成や表現は適切とはいえないが、内容は評価できるものがある。
- 論理性は十分とはいえないが、実務上の有用性がある。
- 有意義な実践・調査報告である。

(2) 掲載を考慮するのが困難と判断される事例

- 問題意識や問題の設定が不明確。
- 基本的な用語の概念の理解や分析枠組が不明確または不適切。
- 論拠とするデータ等の信頼性が乏しい。
- 論旨の明確さや論証の適切さがない。
- 論文の構成、表現（用語、引用、図表等）が適切でない（または整合性がとれていない）。

6. 判定

掲載についての判定は以下の4つの類型に分かれ、最終的に常任編集委員会において決定します。ただし、これらの判定は、評価できる項目や問題のある項目の多少によるものではありません。(3) 及び (4) にあるように、投稿論文の種類以外であれば、掲載を考慮できるとする場合があります。別の種類となる場合、字数の関係で、大幅に修正を要することがあります。

(1) 掲載可とする。

(2) 小規模の修正による掲載可とする。

(3) 大幅な修正による掲載可とする。
但し、（総説／研究論文／研究ノート／実践・調査報告）として掲載を考慮できる。

(4) 掲載不可とする。
但し、（総説／研究論文／研究ノート／実践・調査報告）として掲載を考慮できる。

Publication Policy of the Japanese Journal of Evaluation Studies

Last revised on 15th February 2005

The Purpose and the Name

1. The Japan Evaluation Society (hereinafter referred to as "evaluation society") publishes "The Japan Journal of Evaluation Studies (hereinafter referred to as "evaluation study") in order to widely release evaluation studies and outputs of practical activities to domestic and international academic societies, interested individual and institutions, and contribute to the advancement and prevalence of evaluation practice.

Editorial Board

2. The editorial board administrates editing of evaluation study based on the editorial policy stated below.
3. The editorial board is formed with less than 20 members of the evaluation society who are assigned by the board of directors. Terms of editors are two years but can be extended.
4. The editorial board assigns one editor-in-chief, two vice-editors-in-chief, and a certain number of standing editors among the members.
5. The editorial board may hold at least one meeting to discuss the editing policy, plans of editorial board, and others.
6. The editorial board reports activities to the board of directors as needed and receives approval. Also it is required to report the progress of the past year and an activity plan for the following year at the annual conference.
7. The editor-in-chief, the vice-editors-in-chief and the standing editors organize the standing committee and administrate editing on a regular basis.

Editorial Policy

8. The evaluation study, as a principle, is published twice a year.
9. The evaluation study is printed on B5 paper, and either in Japanese or English.
10. Papers published in the evaluation study are categorized as five types;

10.1. Review
10.2. Article
10.3. Research note
10.4. Report
10.5. Others

11. The qualified contributors are members of the evaluation society (hereinafter referred to as "members") and persons whose contribution is requested by the standing editors. Joint submission of members and joint submission of non-members with a member as the first author are accepted. Submission by the editors is accepted.
12. Submitted manuscripts are treated as the above categories, however, the standing editors will decide based on the application of the contributors and the following guidelines;

12.1. "Review" is a paper, which provides an overview of evaluation theory or practice. The editorial board will make the decision regarding publication.
12.2. "Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.
12.3. "Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical enquiry. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.4. “Report” is the study report related to a practical evaluation project or evaluation. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.5. “Others” includes requested papers for special editions organized by the editorial board and announcements from editorial board to members regarding publication.

13. The editorial board selects two referee readers. For the “article”, the editorial board makes adoption judgments referring to the results from referee readings and comments provided by one editor assigned by the editorial board. For “review”, “research note”, “report” and “others”, the editorial board makes adoption judgments referring to the results from referee readings.

14. When editors submit a manuscript, the editors are not allowed to attend any of the standing editors committee meetings or editorial board meetings regarding the manuscript.

15. The standing editors have alternative of approval or not-approval for adoption judgment of manuscripts submitted to any categories. However exception is permitted if the editorial board approves the publication after minor rewrite. Even if the manuscripts are considered insufficient as an “article”, standing editors can decide whether the manuscripts are published as a “research note” or “report” if the authors wish to publish.

Formulation and Release of Submission Procedure

16. The editorial board formulates the submission procedure based on the editorial policy described above and release after approval from the board of directors.

Distribution

17. The evaluation study is distributed to all members for free and distributed to non-members for a charge.

Distribution of the Printed Manuscript

18. 30 copies of the respective paper are reprinted and distributed to the authors. The authors must cover any costs incurred by author’s requests for printing more than 30 copies.

Release on the Internet

19. The papers published in the evaluation study are released on the internet with approval from the authors.

Copyright

20. Copyright of papers which appear in the evaluation study is attributed to the respective authors. Editorial right is attributed to the evaluation society.

Office

21. The office is in charge of administrative works for editing, distribution, and accounting.

Information for Contributors
(For English Papers)

Last revised on 29th September 2008

1. "The Japanese Journal of Evaluation Studies" is the publication for reviews, articles, research notes, and reports relating to evaluation.

2. "The Japanese Journal of Evaluation Studies" is primary published to provide opportunities for members of the Japan Evaluation Society (hereinafter referred to as "members") to exchange findings, and to contribute to further development of the study of evaluation both domestically and internationally. As a principle, this journal publishes the contributions submitted by the members. With the exception of requested papers, the first author must be a member. A submission (as the first author) is limited to one manuscript that has not been published or submitted in any form for another journal of academic association etcetera.

3. Adoption judgments of the manuscript are made at the discretion of the editorial board. Comments from two referee readers who are appointed for every manuscript are referred to in the screening process (the editorial board requests referee readers without notifying the author of manuscript).

4. Payment for the manuscript is not provided.

5. Papers published in "The Japanese Journal of Evaluation Studies" are released on the Internet at homepage of this academic society.

6. Regarding submission, manuscripts must be identified as one of the following categories: 1) article, 2) review, 3) research note, 4) report, and 5) others. However, the final decision of the category is made by the editorial board.
   "Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice.
   "Review" is a paper which provides an overview of evaluation theory or practice.
   "Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical study in the process of producing an "article".
   "Report" is the study report related to a practical evaluation project or evaluation.
   "Others" are manuscripts for special editions requested by the editing committee.

7. Manuscript Submission
   (1) Manuscripts may be written in either Japanese or English.
   (2) Correction by the author is only for the first correction.
   (3) English manuscripts should be submitted only after the English has been checked by a native speaker.
   (4) Submit four hard copies (A4 size) of the manuscript. Contact information including mailing address, telephone number, fax number, and e-mail address, and the category of the manuscript should be clearly stated.
   For approved manuscripts, after necessary rewriting, the author needs to submit two hard copies of the final paper as well as a text file saved on a DOS/V formatted floppy disk. Original figures, charts, and maps should be provided.
   (5) Total printed pages should not exceed 14 pages. Any cost incurred by printing more than 14 pages must be covered by the author.
   (6) The layout for English papers should be 30 mm of margin at left and right side, 10pt for font size, 43

lines on A4 paper (about 500 words per page). An abstract of 150 words should be attached to the front. 14 pages are equivalent to 7,000 words but the body should not exceed 6,000 words to allow for the title, header, figure, chart, footnotes, and references. Please note that the number of pages may be more than expected depending on the number of figures included.

8. Mailing address
Office of Japan Evaluation Society at International Development Center of Japan
Hitachi Soft Tower B 22nd Floor, 4-12-6, Higashi-Shinagawa, Shinagawa, Tokyo,
140-0002, Japan
Phone: +81-3-6718-5931, Facsimile: +81-3-6718-1651
E-mail: jes.info@idcj.or.jp

Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers)

Revised on 18th September 2002

1. Text, Charts, Figures, Graphs, Diagrams, Notes, and References

(1) The paper should be written in the follow order:

First page: Title; the author's name; Affiliation; E-mail address; Abstract (150 words); Keywords (5 words)

Second page: The main text; acknowledgement; notes; references

(2) Section of the text should be as follow:
1.
1.1
1.1.1
1.1.2

(3) Source of the charts, figures, graphs, and diagrams should be clarified. Submitted charts and others will be photoengraved, therefore it is important that the original chart is clear. Pictures shall be treated as figures.

Figure 1 Number of Students in the State of ○○

Note:
Source:

Table 1 Number of Accidents in the State of ○○

Note:
Source:

(4) Citation of literature in the text should be, (Abe 1995, p.36) or (Abe 1995).
(5) Note in the text should be, (------.[1])

(6) Note and references should be written all together in the end.
Note
1 --------.
2 --------.

(7) Reference should list the literature in alphabet order, and arranged in chronological order. Follow the examples:

Book: author (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6th edition.* Beverly Hills, Calif: Sage Publication.

Article from magazine: author (year of publication). Title. *Title of the magazine*, volume (number), page-page.

(e.g.) Rossi, P. H. (1999). Measuring social judgments. *American Journal of Evaluation*, 15(2), 35-37.

Article in Book: author (year of publication). Title. In editor (Eds.), *Title of the book.* Published location: publishing house, page-page.

(e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

Book by two authors: surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Peters, T., and Waterman, R. (1982). *In Search of Excellence: Lessons from America's Best Run Companies.* New York: Harper & Row.

Book by more than three authors: surname, first name, surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement.* Washignton: Urban Institute.

(note 1) If some references are from the same author with the same publication year, differentiate by adding a,b,c as (1999a), (1999b).

(note 2) If the reference is more than a single line, each line from the second should be indented by three spaces.

(e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

Referee-Reading Guideline

The Japanese Journal of Evaluation Studies Editorial Board,
The Japan Evaluation Society
Approved on 10th September 2005

1. Content of the Referee-Reading Guideline

This Referee-Reading Guideline is to provide explanation of the main publication judgment, procedure of the referee-reading, to the members who submit the manuscript and for the members who are requested to conduct referee-reading in order to carry out the procedure efficiently and effectively.

2. Purpose of Referee-Reading and the Responsibility of the Author

Referee-reading is necessary for the editorial board to make decisions of whether submitted manuscripts are appropriate to publish in the Japanese Journal of Evaluation Studies or not.

If there is doubt or obscurity identified in manuscripts during the referee-reading corrections may be required. Therefore, referee-reading also contributes to the improvement of the submitted manuscripts. However, although the manuscripts are requested corrections, the author is still solely responsible in regards to the contents and it is not attributed to the referee-readers.

Referee-readers are two persons who are requested by the editorial board depending on the specialty or the field of the submitted manuscript. People who are not members of this academic society also may be requested.

3. Items of Consideration in Referee-Reading

Five points are considered in referee-reading, however, the importance of each may be different depending on the type of manuscript.

(1) Importance and utility of the theme
(2) Originality of the study
(3) Structure of the logic
(4) Validity of verification and methodology
(5) Contribution to evaluation theory and practice

- For the article, all of above five are considered.
- For the research note, especially (1), (2), (3), and (4) are considered.
- For the report, especially (1), (3), and (5) are considered.
- For the review, especially (3) and (5) are considered.

4. Attentions in submission of manuscript

Besides above five viewpoints, basic completeness as a paper is also considered, for example;

- appearance of the paper is organized
- written according to the writing manual
- described simply and distinctive
- verification data is appropriately used
- notes and references are corresponding with the text
- terminology is appropriately used
- no wording and grammatical mistakes
- no errors and omission
- no punctuation mistakes
- expression in English abstract is appropriate
- word count is according to the manual

The above mentioned forms and contents are also considered. There have been cases in which graduate students and practitioners posted without organizing the manuscripts as a paper. On those occasions, referee-reading was not conducted. Necessary consultation is strongly recommended prior to submission.

5. Judgment Cases in Referee-Reading

(1) In the case of the manuscript which is considered acceptable for the publication but is not yet complete:
The referee reader should evaluate carefully whether the paper can contribute to the development of evaluation theory or evaluation studies.

- Verification is lacking but the theory and formulation are useful for academic development.
- Analysis lacking but useful for formation and promotion of new theory.
- The literature review is not of a high standard but, the overall study is meaningful.
- Comparative study is not up to standard but is meaningful as an example of application.
- Analysis is lacking but it is meaningful as an evaluation of socially and historically important cases.
- Analysis is lacking but it is meaningful as an evaluation of particular social activities.
- Organization and expression are not up to standard as a paper but the contents are worthy to evaluate.
- Logic is not strong enough but useful in practice.
- The paper has significance as a report.

(2) In case of the manuscript which is considered as difficult for publication:

- Awareness of the issue or setting of the problem is indecisive.
- Understanding or analytical framework of notion of basic terminology is indecisive or inappropriate.
- There is a lack in credibility of data for the grounds of an argument.
- There is no clear point of an argument or appropriateness of proof.
- Organization of the paper and presentation (terminology, citation, chart, etc) are inappropriate (or not consistent).

6. Judgment

The final decision will be made on publication at the standing editors committee following one of four patterns (listed below). However, these judgments are not based on the number of errors but on the strength of the overall report. In the case of (3) and (4), there is a possibility to be published as a different type of paper. If it is published as a different type of paper, major rewrite concerning the number of words may be required.

(1) The paper will be published.
(2) The paper will be published with minor rewrite.
(3) The paper will be published with major rewrite, however as a different type of paper (review, article, research note, or report).
(4) The paper will not be published; however there is the possibility that it will be published as a different type of paper (review, article, research note, or report).

～投稿案内～

日本評価学会では、「日本評価研究」掲載のための投稿原稿を募集しております。随時、投稿を受け付けておりますので、ご興味をお持ちの方は投稿規定・執筆要領をご参照のうえ、奮ってご投稿ください。

投稿先：特定非営利活動法人日本評価学会　事務局
〒140-0002　東京都品川区東品川4-12-6 日立ソリューションズタワーB 22F
一般財団法人国際開発センター内
TEL: 03-6718-5931　　FAX: 03-6718-1651

『日本評価研究』第14巻第1号
2014年11月26日

編集・発行　特定非営利活動法人　日本評価学会
〒140-0002　東京都品川区東品川4-12-6
日立ソリューションズタワーB 22F
一般財団法人国際開発センター内
TEL: 03-6718-5931　　FAX: 03-6718-1651

印　　刷　株式会社　研恒社

©日本評価学会

本誌に掲載されたすべての内容は、日本評価学会の許可なく転載・複写はできません。

ISSN 1346-6151

Japanese Journal of Evaluation Studies

Vol. 14, No. 1, November 2014

CONTENTS

Article

Administrative Project Review～Beyond Change of Governmental Regime～
Yayoi Tanaka

Probative Logic: its evaluation methodological implication
Kunpei Nishimura

Rethinking the Human Development Index
:A Path toward Comprehensive Development Evaluation
Aki Yonehara

Report of the 14th Annual Conference

Report of the 11th Spring Conference

Announcement for the 15th Annual Conference

Report of the Committees

Japan Evaluation Society