

# 第 4 部

## 外部有識者 事業評価委員会による2次評価



# 外部有識者事業評価委員会による 2次評価について

JICAでは、外部有識者の助言を得て評価体制や手法の改善をはかるとともに、評価結果について外部有識者の検証を受けることで、その客観性を高めることを目的に、2002年度に外部有識者評価委員会（2004年度に外部有識者事業評価委員会に改称）を設置しました。委員は、開発援助や評価についての知識や経験をもつ外部有識者（学識経験者、NGO、国際機関関係者、ジャーナリストなど）によって構成されており、評価体制の強化や、評価結果のフィードバック推進に向けた課題など、幅広い事項を議論していただくとともに、JICAが行った評価に対する評価（2次評価）を行い、評価手法の改善や評価を通じた事業の改善について助言をいただいています。

第1回の2次評価は、2002年度から2003年度にかけて、2001年度のプロジェクト終了時評価40件を対象に実施していただきました。その結果は2003年度版の「JICA事業評価年次報告」に掲載しました。JICAでは、プロジェクト終了時評価は外部専門家や相手国関係者の参加を得たうえで、JICA自身が事業管理の観点から責任をもって成否

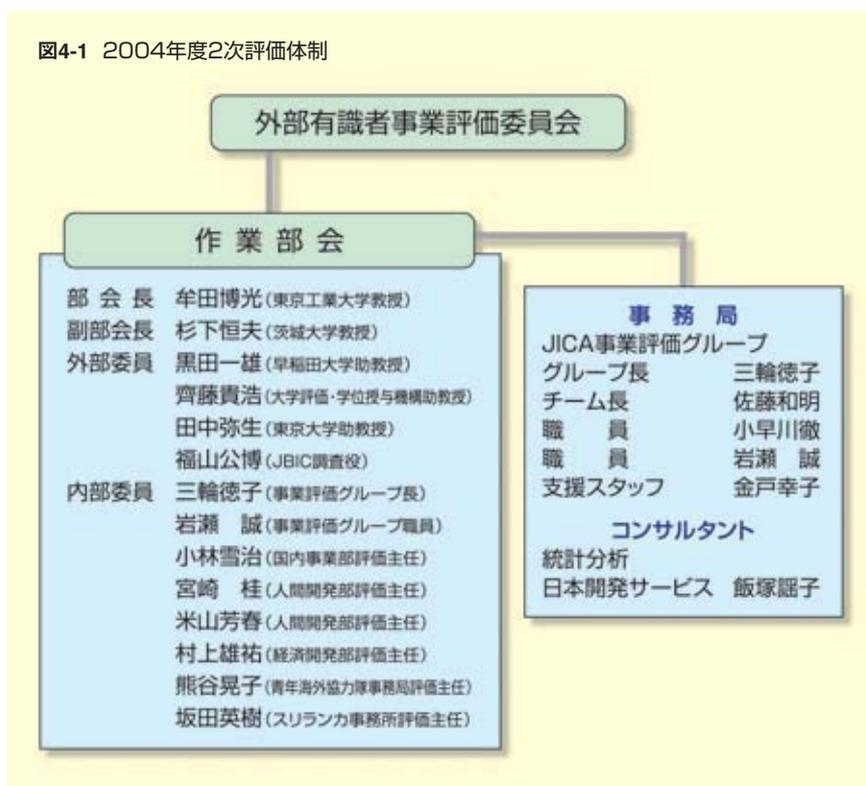
を確認し、必要な判断を行うための内部評価として実施しています。しかし、当事者であるJICAや相手国関係機関が主体となって行う評価は、中立性や客観性を欠く面があるという指摘もあります。2次評価は、こうした問題をふまえて、JICAの内部評価の透明性と客観性を高めることを目的としています。

また、JICAが行う評価の適切性を第三者の目で審査していただき、その現状や課題について指摘していただくことで、評価実施方法や質の改善をはかっていくことも重要な目的です。第1回の2次評価では、評価調査の計画方法と実施方法、評価結果の活用のあり方をはじめ、プロジェクトの運営管理などについて、数々の貴重なご指摘とご提言をいただきました。JICAでは同2次評価結果を反映して、2004年度に事業評価ガイドラインを改訂し、現在、その徹底に努めているところです。

本年度の事業評価年次報告書では、2004年度に行っていた第2回の2次評価の結果を掲載しています。2次評価による検証の対象は、JICAの内部評価の質と評価報告書から読み取れるJICAのプロジェクトの成果です。特に、2004年度の2次評価では、第1回目の経験で明らかになった2次評価の有用性の認識に基づき、JICAが行う評価の現状や課題についてさらに分析を深めるとともに、評価の質の経年的な比較や、外部評価者と内部評価者の評価の違いについても検証を試みました。また、そのために、外部有識者事業評価委員会のもとに、外部評価者と内部評価者からなる2次評価作業部会を設置して作業を行いました。実施体制は図4-1に示すとおりです。

JICAは、こうした評価結果を活用して、JICAが行う評価の透明性を高めるとともに、その質の改善に努めていきたいと考えています。

図4-1 2004年度2次評価体制



# 第1章 2004年度2次評価結果

## —評価の客観性と質の向上のために

外部有識者事業評価委員会／2次評価作業部会

### 1-1 目的・対象・評価方法

#### (1) 目的

外部有識者による2次評価の意義としては、まず、内部評価の透明性を確保することと、評価結果に対する信頼性を確保することがあげられる。

JICAでは、プロジェクト評価は内部評価として実施されている。この評価には外部専門家やコンサルタント、相手国側関係者も参加しているが、彼らは国内協力機関や相手国側実施機関の関係者など、プロジェクトの関係者であることが多い。対象プロジェクトに関するさまざまな事情を熟知した関係者による評価には、きめ細やかな評価ができるという利点があるが、評価の中立性や客観性などの面から問題がないとはいえない。外部有識者による2次評価は、内部評価の透明性を高めて、その適切性を検証するうえで有益である。

また、評価者にはそれぞれ個人的な評価傾向があるため、評価結果はその影響を受ける。外部評価も、プロジェクトとの利害関係がないという意味では客観的であっても、個人的な評価傾向があることから、中立性と公平性の観点において必ずしも優れているとはいえない。これは外部有識者事業評価委員会が2003年度に行った2次評価でも明らかになっている。したがって、より普遍的で信頼性の高い評価結果を得るためには、複数の評価者の目をとおして、個人の評価傾向に偏らない結論を導く努力が必要である。しかし、1つのプロジェクトを何人も外部評価者が評価したり、別の調査団を派遣したりするのは、費用などの面で実用的ではない。2次評価は、1次評価の結果をもとに評価を行うという制約はあるものの、1次評価者とは異なる多くの2次評価者が評価にかかわることで、より偏りが少なく、信頼性の高い評価結果を担保することが期待できる。

以上に加えて、2次評価のもう1つの意義としてあげられるのが、1次評価の質の向上への貢献である。第三者による評価の評価は1次評価における評価の枠組みや調査の実施方法、分析や価値判断の仕方、評価結果の提示方法などの適切性を客観的に検証し、評価の質の改善

に向けた課題を明らかにするのに役立つ。第1回の2次評価では、評価の透明性や信頼性の向上に加え、2次評価のこうした面での有用性が確認された。評価の質の向上は、事業を改善したり、説明責任を遂行したりするうえで、評価がその役割を十分に果たしていくために必須である。1次評価の質の向上に貢献することは、2次評価の重要な意義といえる。

上記のような2次評価の意義をふまえて、2004年度2次評価では、次の評価設問に答えることを具体的な目的として評価を行った。

#### 1次評価(内部評価)の質の検証

- ①評価は一定の質を満たしているか 質の向上に向けての課題は何か
- ②評価の質は経年的に改善されているか
- ③評価の質の向上に向けて留意すべき、内部評価と外部評価の視点の違いは何か

#### 評価対象プロジェクトの成果の検証

- ①2次評価者の視点から見て、対象プロジェクトは効果的・効率的だったといえるか
- ②対象プロジェクトの効果・効率性に影響を与える要因は何か
- ③内部評価と外部評価で、対象プロジェクトの成果に対する2次評価結果に違いはあるか

なお、JICAの評価は、すでに述べたように内部評価として行うものが多い。中立性や客観性の確保の観点からは、2次評価も含めて外部の視点を入れていくことが必要だが、その前に、内部評価者による評価であっても、できる限り客観的で信頼性の高い評価を行うことがまずは重要である。評価の質および評価対象プロジェクトの成果に対する、内部評価と外部評価の違いの検証は、そのための留意点の抽出を目的とするものである。

#### (2) 評価者

2004年度2次評価では上記の目的にもとづき、外部有識者事業評価委員会のもとに、外部評価者と内部評価者から

なる2次評価作業部会を設置して、評価作業を行った。外部評価者は、外部有識者事業評価委員会委員2名、日本評価学会から推薦を得た外部委員3名、JICAと同じくODA事業の実施機関である国際協力銀行職員1名である。内部評価者は、JICA内で評価の質の監理にあたる企画・調整部事業評価グループ職員2名と、事業部門の評価主任6名である。

## 2次評価者

### 2次評価作業部会部会長

牟田博光 東京工業大学教育工学開発センター長  
(外部有識者事業評価委員会委員長)

### 同副部会長

杉下恒夫 茨城大学人文学部教授、元読売新聞社  
(外部有識者事業評価委員会委員)

### 外部委員

黒田一雄 早稲田大学大学院アジア太平洋研究科助教授  
(日本評価学会推薦)

齊藤貴浩 大学評価・学位授与機構評価研究部助教授(同上)

田中弥生 東京大学大学院工学系社会基盤専攻  
国際プロジェクト助教授(同上)

福山公博 国際協力銀行プロジェクト開発部  
開発事業評価室調査役

### 内部委員

三輪徳子 JICA企画・調整部事業評価グループ長

岩瀬 誠 JICA企画・調整部事業評価グループ

小林雪治 JICA国内事業部市民参加協力室チーム長(評価主任)

宮崎 桂 JICA人間開発部第三グループ(保健1)チーム長(評価主任)

米山芳春 JICA人間開発部第四グループ(保健2)チーム長(評価主任)

村上雄祐 JICA経済開発部第二グループ(資源・エネルギー)  
チーム長(評価主任)

熊谷晃子 JICA青年海外協力隊事務局事業管理グループ  
チーム長(評価主任)

坂田英樹 JICAスリランカ事務所次長(評価主任)

※外部評価者：2次評価作業部会部会長、同副部会長、外部委員  
内部評価者：内部委員  
(肩書きは2004年10月時点)

## (3) 評価対象

今回の2次評価では、2002年度に実施されたプロジェクト終了時評価38件と、2004年はじめに2次評価の作業を開始した時点で報告書を公開済みであった2003年度プロジェクト終了時評価10件を主な対象とした。また、すでに第1回の2次評価の対象となった2001年度のプロジェクト終了時評価についても、経年比較を行うために追加的に分析の対象に加えた。分析対象は、第1回2次評価で行われた評価の質についての評点づけを参考に、評点の高いものから低いものまでが均等に盛り込まれるように10件を

選んだ。評価対象とした合計58件の1次評価のリストは表4-1のとおりである。

## (4) 評価方法

対象とする1次評価の報告書を評価者が読み、表4-2の2次評価チェックシートにもとづいて評点づけを行った。

### 1) 評価視点

チェックシートは、1次評価の質を検証する項目と1次評価対象プロジェクトの成果を検証する評価項目からなる。

#### ① 1次評価の質の検証

1次評価の質の検証に関する項目は、項目別評価、総合評価、適切な評価を行うための前提条件(対象プロジェクトの評価可能性)に関する評価項目からなる。それぞれの項目の評価視点は次のとおりである。

■項目別評価：評価方法が適切であったかを、評価調査の枠組み、情報収集方法、分析・評価方法(現状把握、分析、DAC5項目評価)、提言・教訓の抽出・提示方法、報告書のまとめ方の各項目別に評価。

■総合評価：事業評価ガイドラインの「良い評価」の基準(評価情報の有用性、公平性・中立性、信頼性、被援助国の参加度)に照らし、評価の質を総合的に評価。

■評価可能性：目標や成果の達成度の検証可能性、プロジェクトと効果の因果関係についての論理の適切性、評価に必要なデータの蓄積などの観点から評価。

なお、評価可能性を盛り込んだのは、第1回の2次評価で、対象プロジェクトの目標設定やデータ収集、計画の組み立てなどが、評価の質にかかわる問題として指摘されたことによる。当然ながら、報告書だけから評価可能性を評価することには制約があるが、今回の2次評価では、評価可能性の現状や、この問題と評価の関係について分析を深めることを目指して、試行的に対象に加えることとした。

#### ② 1次評価対象プロジェクトの成果の検証

1次評価対象プロジェクトの成果の検証では、DAC評価5項目(妥当性、有効性、効率性、インパクト、自立発展性)および総合評価を項目として用いた。

### 2) 評点

#### ① 1次評価の質の検証

項目別評価の各項目(7項目)、総合評価、評価可能性

表4-1 2次評価対象案件一覧

2001年度(経年的な質の変化の検証のみ分析)			2002年度		
1	インドネシア	酪農技術改善計画	20	ケニア	医療技術教育強化プロジェクト
2	インドネシア	農業普及・研修システム改善計画	21	インドネシア	優良種馬鈴しょ増殖システム整備計画
3	フィリピン	高生産性稲作技術研究計画	22	ベトナム	ハノイ農業大学強化計画
4	マレーシア	化学物質リスク管理プロジェクト	23	バングラデシュ	家禽管理技術改良計画
5	中国	北京消防訓練センタープロジェクト	24	ラオス	ビエンチャン県農業農村開発計画フェーズ2
6	インド	二化性養蚕技術実用化促進計画	25	パナマ	牛生産性向上計画
7	ブラジル	南ブラジル小規模園芸研究計画	26	インドネシア	材木育種計画
8	ブラジル	カンピーナス大学臨床研究プロジェクト	27	インドネシア	生物多様性保全計画フェーズ2
9	アルゼンチン	鉱山公害防止対策研究センター	28	マレーシア	水産資源・環境研究計画
10	モロッコ	鉱物資源探査技術向上プロジェクト	29	ラオス	森林保全・復旧計画フェーズ2
1	インドネシア	国立障害者職業リハビリテーションセンター	30	ボリビア	タリハ渓谷住民造林・侵食防止計画
2	カンボジア	理数科教育改善計画	31	ウルグアイ	林産品試験計画
3	タイ	KMITL情報通信技術研究センター	32	ミクロネシア	漁業訓練計画
4	タイ	都市開発技術向上計画	33	マダガスカル	北西部養殖振興計画
5	フィリピン	マニラ航空保安大学航空管制技術官育成計画	34	フィリピン	電気・電子製品試験技術協力事業
6	マレーシア	日本・マレーシア技術学院	35	フィリピン	工業所有権近代化
7	フィリピン	治水・砂防技術強化ステージ1	36	ラオス	電力技術基準整備
8	インドネシア	初中等理数科教育拡充計画	37	ヨルダン	情報処理技術向上
9	ブラジル	都市交通人材開発	38	パラグアイ	質量分野検定・検査技術向上
10	パラグアイ	日本パラグアイ職業能力促進センター	2003年度		
11	エチオピア	地下水開発・水供給訓練計画	1	インドネシア	地方開発政策支援プロジェクト
12	ケニア	中等理科教育強化計画	2	フィリピン	セブ州地方部活性化プロジェクト
13	インド	新興下痢症対策プロジェクト	3	ガーナ	野口記念医学研究所感染症対策
14	インドネシア	母と子の健康手帳プロジェクト	4	ミャンマー	灌漑技術センター計画フェーズ2
15	スリランカ	ペラデニア大学歯学部プロジェクト	5	メキシコ	農業機械検査・評価事業計画
16	タイ	エイズ予防・地域ケアネットワーク	6	エルサルバドル	農業技術開発普及強化計画
17	ジャマイカ	南部地域保健強化	7	タイ	水管理システム近代化計画
18	ヨルダン	家族計画・WIDフェーズ2	8	ブラジル	アマゾン森林研究計画フェーズ2
19	トルコ	感染症対策	9	ラオス	養殖改善・普及計画
			10	ベトナム	工業所有権業務近代化

の合計9項目について、それぞれ10点満点で評点づけを行った。また、項目ごとにコメント欄を設けて、定量的な評点結果を補完するための、定性的な情報を収集した。

なお、評価項目のそれぞれに、評価する際の着眼点となる小項目を複数記している。小項目は、それぞれの評価項目について、どのような角度から評価するのかを明確にして、評価者間の評価視点の標準化をはかることを目的に新しく設けたものである。小項目は、JICA事業評価ガイドラインのなかで、評価を実施するうえで留意すべき点としてあげられた事項をベースとしている。2次評価では、こうした小項目についても3段階で評価した。

ただし、今回の2次評価では小項目間の重み付けは行っておらず、それぞれの評価者が小項目の評価をふまえて、各項目を評点付ける形をとった。これは、事前に任意の重み付けを行うのではなく、実際に2次評価作業を行ったうえで、今後の2次評価に適用できる、より適

切な重み付けを、作業部会として検討し、提案することとしたからである。

また、すべての評価項目の得点を合算した「総合得点」は設けていない。それは特に低い点があるような場合、総合得点がよくても質の良い評価とはいえず、総合得点が高いことすなわち質の良い評価とはいえないからである。

## ② 1次評価対象プロジェクトの成果の検証

DAC評価5項目の各項目と総合評価の合計9項目について、それぞれ10点満点で評点づけを行った。また、1次評価の質を検証する作業と同じく、項目ごとにコメント欄を設けて、定量的な評点結果を補完するための定性的な情報を収集した。

## ③ 評点づけに際しての工夫

評点づけは、全員がすべての報告書を読んで評価する

のであれば、その平均を求めることで、委員全体の意見が反映され、かつ個人の評価傾向に偏らない得点が得られる。しかし、これは評価者の作業量を考えると現実的にはむずかしい。ただし、報告書1件当たりの評価者が限られれば、その平均値には評価を担当した人の個人的な評価傾向が強く反映するので、評価結果に偏りを生じる。この問題を解決して、普遍性の高い評点を得るために、2004年度2次評価では、第1回の2次評価と同様に統計分析の仕方を工夫し、全員がすべての報告書を読んで評価したとすれば得られるであろう平均値（推定値）を算出した。

すなわち、それぞれの評価者がつけた評点は、理論的に、評価者の個人的な評価傾向によらない部分（評価対象の真の得点）と評価者の評価傾向による部分（評価傾向係数）に分けることができる。したがって、統計分析を活用して両者を分離し、ある報告書の「真の得点」に、その報告書を読んでいない評価者の「評価傾向係数」を加えれば、その評価者がこの報告書を読んだとすれば得

られるであろう推定値を算出できる。また、全員がすべての報告書を読んで評価したとすれば得られると考えられる平均値を求めることもできる。それぞれの評価者の評価傾向の合計は0になるように調整しているため、全評価者の平均値は「真の得点」に一致する。

2002年度と2003年度の終了時評価については、外部評価者1名と内部評価者2名が48件すべての報告書を、ほかの評価者がそれぞれ10件程度の報告書を読むことにした。こうして、1冊の報告書を最低でも2名の外部評価者と3名の内部評価者が読むようにしたうえで、以上述べたような統計分析手法\*を活用し、分析を行った。また、2001年度終了時評価についても、対象となった10件の報告書を外部評価者1名と内部評価者1名が読んで、同様の分析手法を用いた。なお、内部評価者については、所属部署あるいは過去に担当したプロジェクトの報告書の2次評価は行わないようにした。

表4-2 2次評価チェックシート

案件名		
※記入方法		
① 薄緑	の部分でドロップダウン・リストから選択する【I～III】	
基準	A：おおむね良い（10点満点で7点以上） B：不十分な点がある（10点満点で4～7点） C：悪い（10点満点で0～3点）	
② 黄	の部分に小項目（上記3段階評価）の視点を含めた総合評価を10点満点で記載する【I～IV】	
基準	9-10点：良い 7-8点：やや良い 5-6点：普通 3-4点：やや悪い 1-2点：悪い	
③コメント欄には、重要点や特筆されるべき点を記入する（良い事例を含む）【I～V】		
I 適切な評価が可能であったかどうかの前提条件（評価可能性）		
視点	【計画段階のPDMの適切性】当初作成されたPDMは、大きな修正を施さずに評価に用いることができたか	
視点	評価で用いたPDMは、プロジェクトの計画段階で作成されたPDMから、大きく乖離していなかったか（＝プロジェクトの目標や指標の変更により、プロジェクト自体が大きく変わっていなかったか）。	
視点	【達成度（実績）の検証可能性】評価に用いたPDMにおいて、プロジェクトのアウトプットと目標は達成度の評価が可能ないように設定されていたか	
視点	アウトプット、プロジェクト目標、上位目標において、指標とその数値目標（目標値）、および裨益対象が明確・適切に設定されており、PDMに基づいた達成度が評価可能な形になっていたか。	
視点	【プロジェクトの組み立て（因果関係）の検証可能性】評価に用いたPDMにおいて「活動→アウトプット（成果）→プロジェクト目標→上位目標」の論理は妥当だったか	
視点	外部条件を含めた因果関係の仮説が明確かつ現実的であり、妥当な論理が組み立てられていたか。	
視点	【実績・実施プロセスの検証可能性】プロジェクトのモニタリングが行われ、必要な記録（ベースライン・データを含む）が残されていたか	
視点	アウトプット、活動、投入などを中心に、プロジェクトのモニタリングが定期的に行われ、評価に必要な情報・データの蓄積がなされていたか。	
評点	10	コメント

\*章末の「テクニカル・ノート」（\*1 評価得点の推定）参照。

II 項目ごとの二次評価

1. 評価フレームワーク	
	【調査スケジュール】評価調査のスケジュールは適切か
視点	評価に必要な現地調査活動(情報収集や先方関係者との協議など)が十分に行える調査スケジュールが組まれているか。なお、調査スケジュールを見る視点には、質問表配布などの事前の準備・指示の巧拙、調査実施のタイミングの適切性、現地調査期間の十分性、現地調査行程の適切性を含める。
	【調査団構成】中立性・公正性に配慮した調査団構成となっているか
視点	調査団の構成にあたっての公平性・中立性確保への配慮が報告書中に言及されているか。言及がない(N/A)場合は、報告書から読み取れる情報をもとに、「評価の公平性・中立性は害されていないか、調査団構成の偏りがその要因となっていないか」などを基準として判断する。
	【調査団構成】調査団として必要な専門性は確保されたか
視点	調査団としてのバランスのとれた専門性確保への配慮が報告書中に言及されているか。言及がない(N/A)場合は、報告書から読み取れる情報をもとに、「専門知識の不足や評価の質への悪影響はないか、調査団構成の偏りがその要因となっていないか」などを基準として判断する。
	【被援助国の参画度】途上国側関係者の評価調査への参画は十分か《先方関係者の評価者としての評価への参加》
視点	途上国側関係者の評価調査への十分な参画(=評価方法に対する理解に基づき、一連の評価活動にJICAと共同して従事)が報告書中に記述されているか。具体的な記述がない場合は、調査団構成やミニッツなどの報告書から読み取れる情報をもとに、参画度を判断する。
評点	コメント 10
2. 情報収集	
	【評価設問】評価設問は適切に設定されているか
視点	評価目的を踏まえた評価設問が評価グリッドにおいて適切に設定され、現実的な情報収集計画となるよう具体的・実際的なレベル(小項目)までブレイクダウンしてあるか。
	【収集方法】情報収集方法に偏りはあるか(*1)
視点	複数の収集方法を組み合わせることにより、情報の客観性・信憑性を高めるよう配慮しているか。
	【情報源】データ・情報の収集先(訪問先・面会者など)は明らかか
視点	情報収集先に関する十分な説明(出典、面談者リストなど)があるか。
	【情報源】データ・情報の収集先(訪問先・面会者など)は適切か《先方関係者の情報源としての評価への参加》
視点	最終受益者を含め、幅広い情報収集先をバランスよく当たっているか。
	【収集実績】評価を行ううえで十分な情報が収集されているか
視点	設定した情報収集計画(評価グリッド)に基づき、評価設問への回答を得るに十分な情報が得られているか。また、調査の実施過程で新たな疑問が生じた場合は、必要な追加情報を収集しているか。
評点	コメント 10
3. 分析/評価	
3.1 現状把握	
	【プロジェクトの現状把握・検証】実績の検証は十分か
視点	アウトプット、プロジェクト目標、上位目標に関する達成度、もしくは達成予測が、指標・目標値を用いて的確に検証されているか。
	【プロジェクトの現状把握・検証】実施プロセスの検証は十分か
視点	プロジェクトの実施プロセスに対する検証(*2)が、成果や目標達成に当たったの貢献・阻害要因の特定につながる形で、十分に行われているか。
	【プロジェクトの現状把握・検証】因果関係の検証は十分か[1]
視点	計画段階におけるプロジェクトの組み立てに対する検証(*3)が、成果や目標達成に当たったの貢献・阻害要因の特定につながる形で、十分に行われているか。
	【プロジェクトの現状把握・検証】因果関係の検証は十分か[2]
視点	「with/without」や「before/after」の比較を通して、効果がプロジェクトの結果もたらされたものであるかの検証(*4)を、納得性の高い形で試みているか。
評点	コメント 10

3.2. 分析		
		<b>【データ分析全般】客観的な分析を行っているか</b>
視点		科学的な議論の積み上げにより、客観的な分析を行っているか。定量化できるものについては、可能な限り定量化する工夫がはかられているか。
		<b>【データ分析全般】多面的な分析・データの解釈を行っているか</b>
視点		複数のデータ分析方法を組み合わせることにより、さまざまな角度からの検証・分析を行ったうえで、結果の解釈を行っているか。
		<b>【貢献・阻害要因の分析】現状把握に基づいて、貢献要因・阻害要因の分析が十分に行われているか</b>
視点		計画段階におけるプロジェクトの組み立て（因果関係のロジック）と実施段階におけるプロセス（運営体制など）の双方について、効果の発現に対する貢献・阻害要因を十分に分析しているか。
評点	コメント	
		10
3.3. 評価		
		<b>【5項目評価（*5）】「妥当性」の評価は十分か</b>
視点		本評価項目の定義（ニーズとの合致、政策との整合性、手段としての適切性などの視点）に的確に従って、網羅すべきポイントを押さえた評価を行っているか。
		<b>【5項目評価】「有効性」の評価は十分か</b>
視点		本評価項目の定義（プロジェクト目標の明確性、プロジェクト目標達成に関わる因果関係の論理性などの視点）に的確に従って、網羅すべきポイントを押さえた評価を行っているか。
		<b>【5項目評価】「効率性」の評価は十分か</b>
視点		本評価項目の定義（費用節約、コスト面での類似プロジェクトとの比較（費用対効果の考え方）などの視点）に的確に従って、網羅すべきポイントを押さえた評価を行っているか。
		<b>【5項目評価】「インパクト」の評価は十分か</b>
視点		本評価項目の定義（上位目標の明確性、上位目標達成に係る因果関係の論理性、波及効果の有無などの視点）に的確に従って、網羅すべきポイントを押さえた評価を行っているか。
		<b>【5項目評価】「自立発展性」の評価は十分か</b>
視点		本評価項目の定義（政策・制度面、組織・財政面、技術面等における自立発展性の視点）に的確に従って、網羅すべきポイントを押さえた評価を行っているか。本評価項目については、上記各視点からの自立発展性の評価だけでなく、全体としての自立発展性を総合的に判断しているかも重要なポイントとなる。
		<b>【評価全般】評価の判断基準・根拠は妥当か</b>
視点		評価の判断基準・根拠が明記されており、納得性が高いものになっているか。
		<b>【評価全般】結論は適切にまとめられているか</b>
視点		評価結果を横断的に取りまとめ、当初設定した評価の目的に対して的確に判断を下しているか。
評点	コメント	
		10



IV 報告書から読み取れるプロジェクトの評価(5項目評価)

妥当性		
評点		コメント
有効性		
評点		コメント
効率性		
評点		コメント
インパクト		
評点		コメント
自立発展性		
評点		コメント
総合評価		
評点		コメント

V その他コメント(自由記述欄)

[注釈]「JICA事業評価ガイドライン 改訂版」における説明内容(抜粋)

- \*1: 主な情報収集方法の種類  
①文献・既存資料調査 ②直接観察 ③質問紙調査(アンケート調査) ④インタビュー調査 ⑤フォーカスグループ・ディスカッション
- \*2: 実施プロセスの検証を通して把握できること  
①活動の検証 ②ターゲット・グループとの関係 ③プロジェクトのマネジメント ④全体的視点
- \*3: 定性的な因果関係の検証方法:プロジェクト実施状況と変化の関係性やプロジェクトに関わる人々の認識などを丁寧にモニターすることによって、因果関係を推測していく方法。  
①投入から、活動、アウトプット、目標に至るまでの実施プロセスの経緯を積み上げる。  
②プロジェクトの実施と効果のロジックの論理的な説明を試みる。  
③技術の移転、普及過程を分析する。  
④プロジェクトから受益する地域や対象を限定し、より深くデータ分析を行うことにより、プロジェクト実施との関係性を明らかにする。
- \*4: 定量的な因果関係の検証方法:比較調査もしくはベースライン・データを用いて、「with/without」や「before/after」の比較を行う方法。  
①プロジェクト実施前と実施後の受益者や対象社会の変化を比較する(before/after)。  
②プロジェクトの受益者・対象社会と、プロジェクトの影響を受けない人々・社会の状況を比較する(with/without)。  
③①と②の組み合わせ(実験的計画手法)。  
④①と②の組み合わせ(準実験計画手法)。
- \*5: 評価5項目の各項目においてチェックすべき視点については、「JICA事業評価ガイドライン改訂版」の第3部第2章を参照。
- \*6: 提言と教訓の定義は以下のとおり。  
「提言」:評価対象プロジェクトに関して、JICAや相手国の実施機関関係者に対し、具体的な措置、提案や助言を行うもの。  
「教訓」:プロジェクトの経験から特定できるもので、実施中の類似プロジェクトや、将来開始されるプロジェクトの発掘・形成に参考になる事柄。
- \*7: 「良い評価の基準」の定義については、「JICA事業評価ガイドライン改訂版」の第1部第1章(p29)を参照。

(注) 独立行政法人国際協力機構企画・評価部評価監理室編(2004)「プロジェクト評価の実践的手法:JICA事業評価ガイドライン改訂版」国際協力出版会。  
このガイドラインはJICAホームページ(<http://www.jica.go.jp/index-j.html>)の「評価」サイトでもご覧になれます。

評価結果

2-1 1次評価の質の検証

(1) 評価の質の現状と課題

1) 評価結果概況

2002年度と2003年度の終了時評価48件について、各評価項目の平均点は図4-2のとおりである。いずれの評価項目も平均値は6.0点以上で、採点基準の「普通」（5～6点台）のレベルにある。ただし、そのなかでも、DAC評価5項目による「分析/評価（評価）」や「良い評価」の基準による「総合評価」は比較的高い点数となっているが、「評価可能性」や「提言・教訓」の点数はやや低い。

次に、評価項目別の得点分布をみると、図4-3に示すとおり、全体として5.0点から8.0点の間に多く分布している。このうち、たとえば「評価フレームワーク」は6～7点

図4-2 評価項目別の平均点

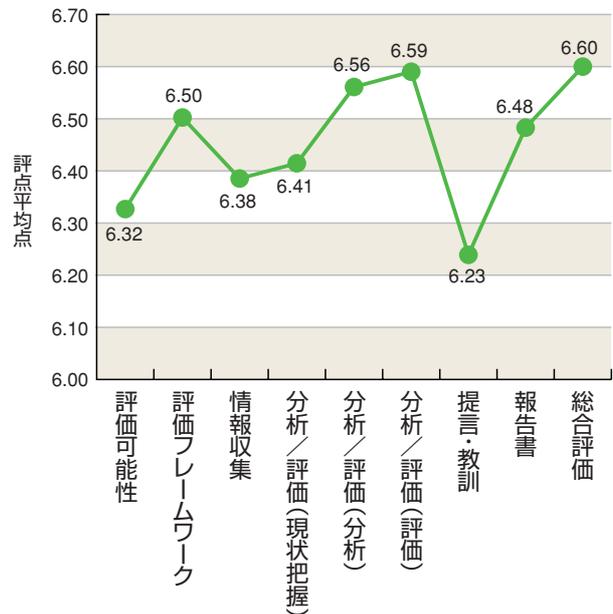
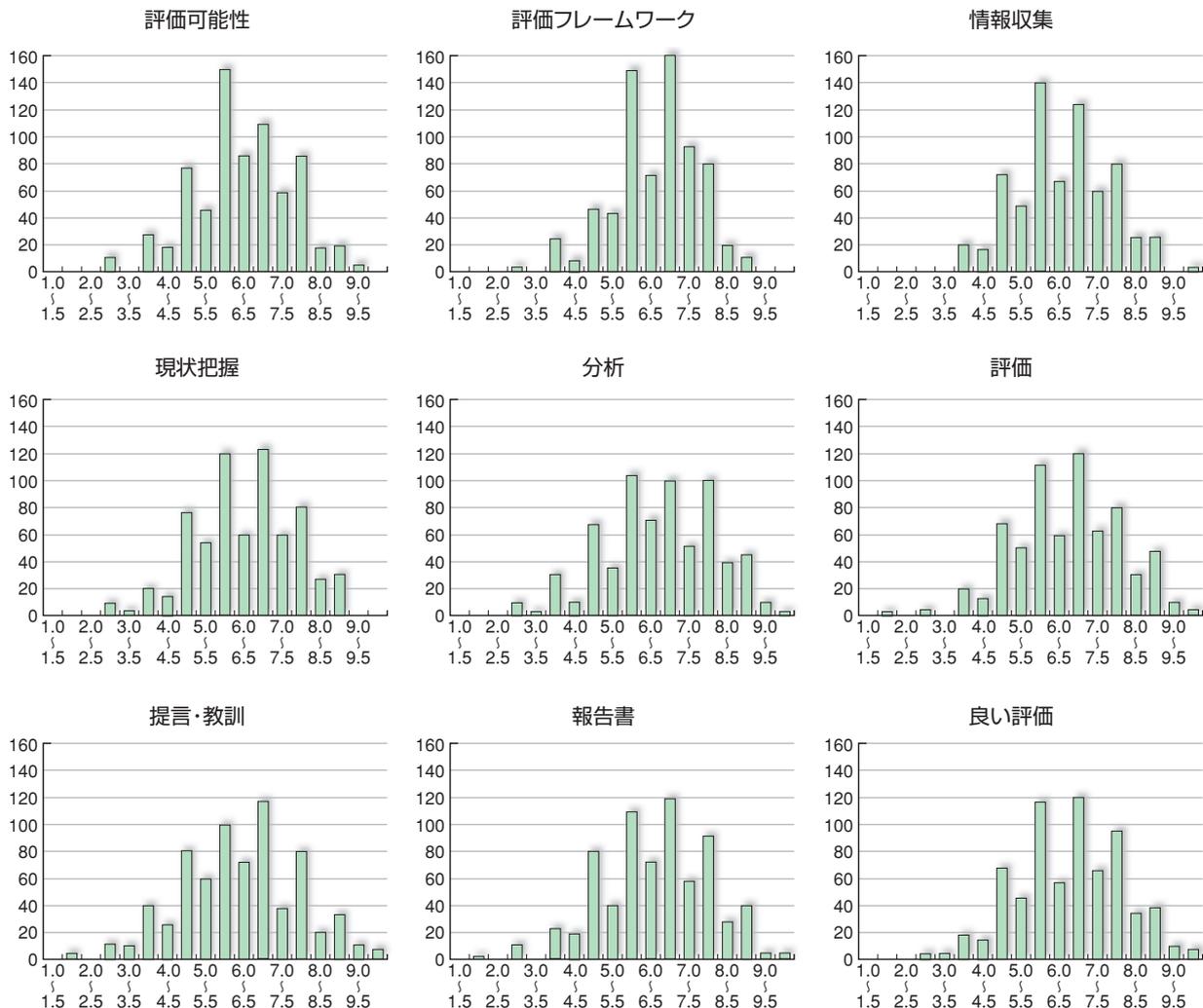


図4-3 評価項目別の得点分布



(注) 縦軸は2次評価の件数。ただし、14名の評価者が48冊の報告書をすべて読んだ場合に得られる、672件を母数とする。

台が大半を占めており、ばらつきも少ない。しかし、「分析／評価」の各項目と「提言・教訓」に関しては、8.0点以上のもも少なくない一方で5.0点以下のももあり、1次評価によって質にばらつきがある。そのなかでも「提言・教訓」については、5.0点以下の件数が相対的に多い。また、「評価可能性」については、平均点前後のものが多く、偏りは少ない。

以上から、対象案件の1次評価の質については、件数としては「普通」（5～6点台）から「やや良い」（7～8点台）が多く、平均して「普通」レベルと一定の質を満たしているが、「分析／評価」や「提言・教訓」を中心に、項目によっては1次評価の質にばらつきがあり、特に「提言・教訓」の質は、相対的に低い状況にあるといえる。

## 2) 項目別評価結果と質の向上に向けた課題

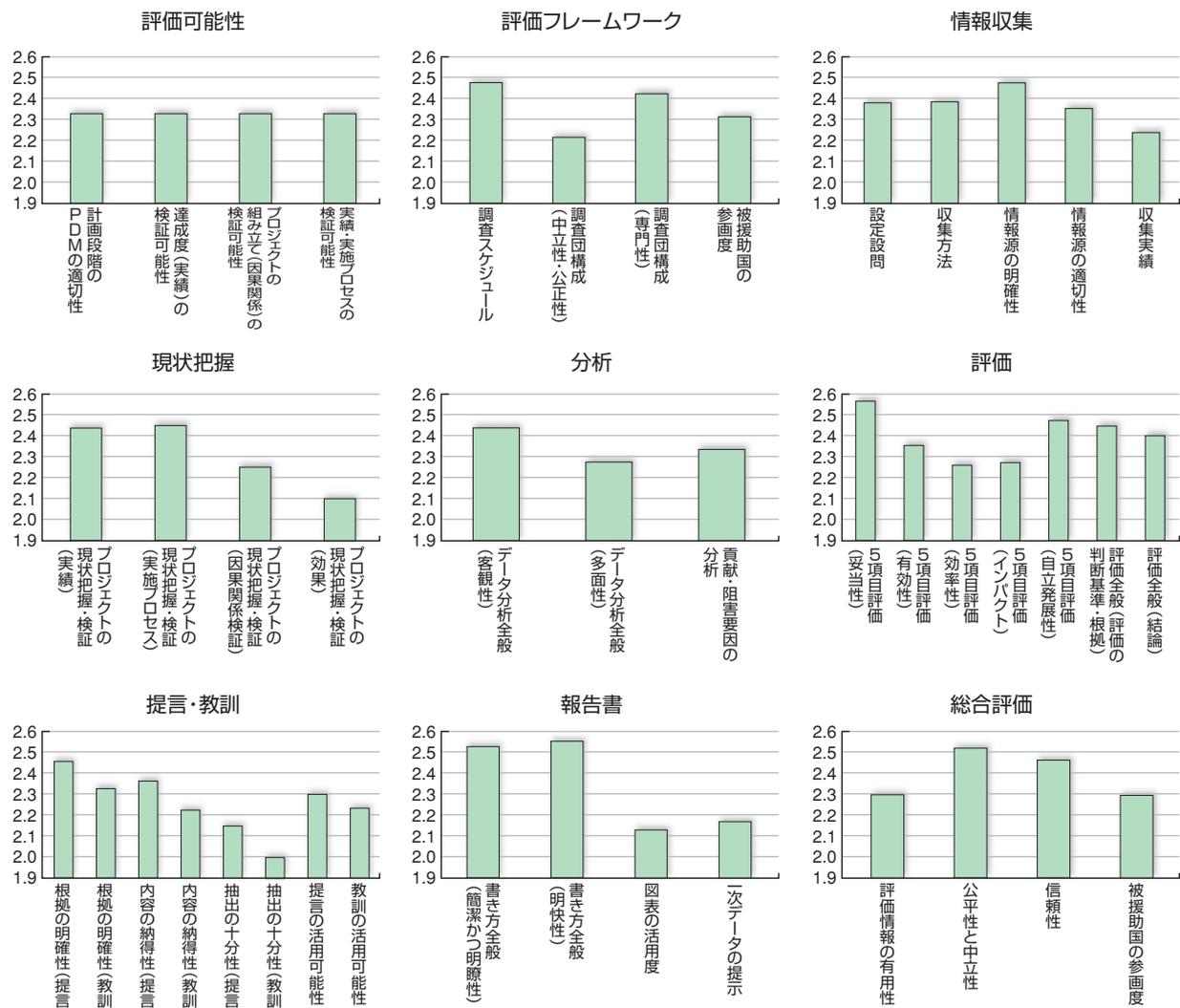
2次評価では、それぞれの評価項目について、小項目

の評点付けを行うとともに、記述式のコメント欄を設けて、評価者コメントの形で定性的な情報も収集した。以下では、各評価項目における小項目の評価結果と、委員のコメントにもとづいて、項目別に評価の質の現状と課題についてより具体的にみる。図4-4は、各評価項目における小項目の評価結果（平均点）である。小項目はA、B、Cの3段階で評価しているが、分析のためにそれぞれを3点、2点、1点に換算して点数化している。

### ① 評価可能性

評価可能性については、小項目の評点は2.2～2.3点台で、ほかの評価項目の小項目得点と比べて、特に点数が高い項目や低い項目はない。ただし、この評価項目のなかでは、「計画段階のプロジェクト・デザイン・マトリックス（PDM）の適切性」と「実績・実施プロセスの検証可能性」の得点が相対的に低い。「PDMの適切性」の項目は、計画

図4-4 各評価項目における小項目の評価結果



(注) 縦軸は評点。

段階で適切な目標や指標が設定されず、その結果生じたPDMと実際のプロジェクトの乖離を埋めるために、評価段階になってPDMに大幅な修正を加えているようなケースがあることが第1回の2次評価で指摘されたことをふまえて設定したものである。評価は計画と比較して達成度をみるものだが、その計画が評価段階になって大きく変更されたりしていないかを問う項目になっている。また、「実績・実施プロセスの検証可能性」については、定期的にモニタリングを行って、評価に必要な情報やデータを蓄積していたかを問う項目であり、これも第1回の2次評価で指摘された事項となっている。

今回の2次評価では、実施段階で必要に応じてPDMを見直していたケースが多く、終了時評価の段階でPDMを大幅に修正していた例は少ない。しかしながら、上位目標とプロジェクト目標が乖離していることを理由に、上位目標を下げる形で修正している例もみられる。これについては、そもそもの課題設定にあたる上位目標を変えるという発想に対する問題が指摘されている。ちなみに上位目標については、計画段階における設定の不十分さ(高すぎる上位目標や矮小化された上位目標の設定)を指摘するコメントも少なくない。なお、実施過程でPDMを変更したとの記述があるのに、その変更内容や変更過程についての情報の記載や資料の添付がないケースも散見された。また、委員からのコメントが最も多かったのが、評価に必要なデータについての問題である。これについては、達成度を検証するための指標データの不足、これに関連してそもそも適切な指標を設定したかという問題、さらにベースライン・データの不十分さが指摘されている。

## ②評価フレームワーク

評価フレームワークについては、「調査スケジュール」をはじめ、総じて小項目の得点は高い。しかしながら、そのなかでも、「調査団構成(中立性・公平性)」の点数はやや低い。これは調査団員のほとんどが、国内協力機関の関係者を含む、プロジェクト関係者であることによる。プロジェクトを技術的に支援してきた国内協力機関の関係者が団員として参加することは、専門性では優れている面もあるが、評価コンサルタント以外はずべてプロジェクト関係者であるような場合、やはり中立性は低く、第三者の参加も得るような構成を検討していくべきとの指摘がなされている。

一方、調査団員については、報告書に氏名と所属しか記されておらず、所属機関から関係者と考えられるものの、プロジェクトと具体的にどうかかわっているかは判

断しにくいという意見、さらに、専門分野や団員としての業務内容がその人の所属機関からだけでは判断できず、団員の人数や構成が適切かどうかわかりにくいという意見が多かった。また、相手国側関係者の参画度についても、いずれの1次評価も合同評価の形態をとり、合同評価委員会を開催して、その評価結果に合意署名を交わしているが、関与の度合い(一緒に調査を行ったのか、日本からの調査団の調査対象としてインタビューを受けたか、委員会に参加したのみなのかなど)は判断しにくいとの意見があった。評価の枠組みについては、それが適切なものかどうか分かるように、情報をもう少し記載すべきだという指摘が多い。

なお、調査スケジュールに関しては、調査日数がやや少ないのではないかという意見が多かった。また、効率よく調査するために、団員の一部が事前に現地に入り、後から残りの団員が合流するなどの工夫をしたケースがあげられている。

## ③情報収集

「収集実績」以外の項目は23~24点台となっている。点数が最も高かった「情報源の明確性」は、情報収集先(訪問先、面談者、データの出典など)が十分に説明されているかを問う項目であり、一部には出典などの記述が不十分な1次評価もみられたが、全体としては適切に示されているとの評価になっている。

一方、評価がやや低かった「収集実績」は、評価を行ううえで必要な情報が十分に収集されていたかどうかを問う項目である。情報収集についてのコメントでは、達成度を検証するためのデータの不足や、活動・成果を中心とした実績情報の収集に偏り、課題(上位目標)に対する効果についての情報収集が不足していることを指摘するものが多い。また、「情報源の適切性」の小項目にも関係するが、プロジェクトの記録のみからデータを収集し、カウンターパートなどの関係者のみにインタビューを行っているようなケースをはじめとし、情報収集先がプロジェクト関係者や機関に偏る傾向があり、幅広い情報源、特に評価において重要な受益者からの情報収集が不十分との指摘が多かった。

## ④現状把握

現状把握については、「実績」と「実施プロセス」の検証はそれぞれ2.4点台となっている。しかし、「因果関係」の検証については、計画段階でのプロジェクトの組み立ての検証を問う項目、プロジェクトの有無や前後の比較

を通じた効果とプロジェクトの関係の検証を問う項目が、ともに低い評価になっている。

コメントでも、プロジェクトの開始前と後や、プロジェクトの対象地域と非対象地域のデータを比較して、プロジェクトの効果を定量的に検証した評価が少ないという指摘が多い。また、効果については、上位目標に対する効果の検証が不十分という指摘も多い。これに関連して、上位目標の設定および同指標データの収集の問題や、プロジェクトの実施と上位目標の達成との間の因果関係の論理の整理の不十分さが言及されている。さらに、成果や目標の達成を指標の達成度だけから評価して、計画の組み立てや実施プロセスを十分に検証しておらず、その結果として貢献要因と阻害要因を分析できていない1次評価も散見された。しかし、中には、ベースラインとの指標の比較や経年的な指標の変化により、定量的に効果を検証している評価や、定量・定性双方のデータを十分に組み合わせて効果の検証や貢献・阻害要因の分析を行っている評価など、ほかの参考になるような1次評価もあった。

#### ⑤分析

分析については、「客観性」では2.4点台となっているが、「多面性」はやや低い。コメントでは、分析が浅いと意見が多く、具体的には、情報・データ不足による分析の弱さ、聞き取りなどの限られた情報のみで頼った分析の不十分さ、定量的な分析の少なさ、分析にあたっての専門的な知見の不足などがあがっており、なかには現象を叙述したり、事実を総括しているのみで分析になっていないとの指摘がなされた1次評価もあった。「多面性」は、複数のデータ分析方法を組み合わせることによって、さまざまな角度からの検証と分析を行っているかどうかを問うものであるが、評点の低さは以上のようなコメントを反映している。

ただし「分析」については、2次評価で8.0点以上の評価を得た1次評価も数件あった。高い評価を得た1次評価では、多面的な情報にもとづいて、定性・定量双方の分析手法を用いたうえで、バランスのとれた分析を行っていた。

#### ⑥評価

「妥当性」「自立発展性」をはじめ、7つの小項目中、5項目が2.3点以上であり、ほかの評価項目に比べて全体的に小項目の評点は高い。しかしながら、「効率性」「インパクト」については、やや低い評価となっている。

「効率性」については、第1回の2次評価でも、DAC評価5項目による評価のなかで最も低い評点付けがなされているが、コスト面、特に費用対効果の視点からの評価が不十分なことが低い評価につながっている。1次評価のなかには、費用面の分析を試みているものも若干みられたが、費用対効果の面で納得性の高い評価に成功している例はなかった。

また、「インパクト」については、課題（上位目標）に対して計画したインパクトの評価が不十分であるという指摘が多い。そしてこの関連では、上位目標の設定、指標データの不足、因果関係の論理性の問題が言及されている。また、計画したインパクトである上位目標に対する効果を十分に評価せず、予期しなかったインパクトについてもっばら記した1次評価もみられるなど、インパクトの定義の混乱を指摘するコメントもあった。

#### ⑦提言・教訓

提言・教訓が調査結果から導かれたものかどうかを問う「根拠の明確性」や提言の「内容の納得性」は2.3点以上であったが、ほかの項目に比べて小項目の評点も総じて低い。特に、評価のプロセスを経て得られた貢献要因や阻害要因などの情報を、提言・教訓に十分反映しているかどうかを問う「十分性」は、小項目すべてに関する評価結果のなかで、最も低い評点がついている。

コメントでも、提言・教訓が表面的であるという指摘が多い。特に教訓については、調査結果のなかで今後の教訓になるような計画面、実施面の問題を多く述べているにもかかわらず、それが教訓に十分に盛り込まれていない、教訓が体系的に抽出されていないなどの意見があげられている。また、「活用可能性」については、短期的に対応すべき事項と中長期的に対応すべき事項、日本側が対応すべき事項と相手国側が対応すべき事項などに分けて提言している1次評価もある一方で、誰がいつまで行うのかがあいまいな提言や、一般論になっていて具体性に乏しい提言・教訓もみられ、提言・教訓を改善することが必要だと指摘するコメントが多い。

#### ⑧報告書

報告書の構成、用語使い、論旨の明確さなど、「書き方」に関する小項目は2.5点台となっている。ただし「図表の活用度」や「1次データの提示」については、ともに2.1点台で評価が低い。基本的な統計データをわかりやすい図表を使って本文中に盛り込んだり、本文中で引用したアンケートやインタビューの結果を簡潔にまとめて、巻末

に資料として添付したりしている報告書もある。一方で、巻末資料を見ないと統計的な数値がわからなかったり、添付している資料は多いが、アンケートの集計結果などの1次資料が盛り込まれていないという報告書も少なくなく、中には資料が大半で本文のページ数が非常に少ないような報告書もみられた。

### ⑨ 良い評価

ほかの評価項目に比べて小項目の評点は総じて高く、特に「公平性・中立性」「信頼性」が高い。やや低いのが「有用性」と「被援助国の参画度」であり、ほかの評価小項目でも指摘があった、相手国側関係者の参加や受益者からの情報収集、分析や提言・教訓の問題が指摘されている。

なお、今回の2次評価では、評価項目間の関係についても、統計分析を用いて検証した\*。この分析結果からは、すべての評価項目の間には正の相関関係があり、特に「情報収集」と「分析／評価」の各項目（「現状把握」「分析」「評価」）との間、「分析／評価」の各項目間、さらに「提言・教訓」と「現状把握」および「分析」との間に、強い相関関係があることが示されている。これは、質の高い「分析／評価」のためには適切な「情報収集」が、また、質の高い「提言・教訓」を引き出すためには適切な「現状把握」や「分析」が重要であると、2次評価者が判断したことを示している。

同様に「総合評価」は、「評価フレームワーク」「情報収集」「分析／評価」「提言・教訓」「報告書」のすべての項目と強い相関関係をもっており、なかでも「分析／評価」と「提言・教訓」とは高い相関を示している。すなわち、「良い評価」たるには、「分析／評価」および「提言・教訓」の質が重視されるといえる。

この分析はまた、「評価可能性」と評価の質の間にもすべて正の相関関係があることを示している。「評価フレームワーク」や「報告書」のように相関関係がそれほど強くないものもあるが、「情報収集」や「評価」では強い相関関係が示されている。「評価可能性」の低さは、評価に際しての適切な「情報収集」や「評価」に影響する、と評価されたことになる。ただし、「評価可能性」と「総合評価」の相関関係は、ほかの評価項目と「総合評価」の間ほどには強くない。「評価可能性」が低ければ、「情報収集」などにむずかしさは生じるが、それによって評価の

質が必ずしも低くなるとは限らない。対象となった1次評価のなかにも実際に見られたが、「評価可能性」の低さを含めて、問題を適切に分析した質の高い評価もあり得るからだと考えられる。

### 3) 質のよい評価報告書の事例

JICAの事業評価ガイドライン改訂版では、「評価フレームワーク」「情報収集」「現状把握」「分析」「評価」「提言・教訓」「報告書」などの各項目について、より適切な評価を実施するために留意すべき事柄が詳細に説明されている。しかし、一般論だけでは、なかなか具体的には理解できない。2次評価結果などを活用して、質のよい評価報告書を客観的に示すことができれば、それらの報告書を一種の手本として利用できる。また、こうした作業を続ければ分野や課題ごとに手本がそろい、その手法や内容を参考としながら、評価調査や報告書作成を行うことで、質の確保が期待できる。

こうした考えのもと、2004年度の2次評価対象48件の終了時評価の中から、ほかの手本となるような評価として抽出されたのは、次の4つの終了時評価報告書である。いずれも、評価そのものの質にかかわる項目（「項目別評価」および「総合評価」）の平均点が7.5点以上であり、かつすべての項目において6.0点を下回るものがなく、全体としてバランスのとれた質のよい評価報告書となっている。また、「現状把握」「分析／評価」「提言・教訓」に8点台のものが多く、いずれも定量・定性双方のデータから十分に現状を把握し、分析することで、納得性の高い評価が行われ、具体的に有用性の高い提言・教訓を抽出しているところが共通している。

これは、評価調査にあたった1次評価者が評価や専門分野に対して知識・理解が深く、調査手法にも習熟していたことが、要因の1つと判断される。それとともに、いずれの評価も現状把握と分析に必要なデータが十分に入手されていることから、プロジェクト実施段階から適切なモニタリングが行われ、必要なデータが蓄積されていたことがうかがわれ、これも質の高い評価を可能とする要因であったと考えられる。

なお、多くの項目で最高得点を得たエチオピア「地下水開発・水供給計画」については、評価対象のプロジェクト自体には若干の問題もみられたが、プロジェクトの現状を十分に把握し、貢献要因と阻害要因を掘り下げて分析したうえ、納得性の高い根拠にもとづいて問題を的

\* 章末の「テクニカル・ノート」(\*2 評価項目間の相関関係) 参照。

確・率直に指摘していることが、多くの2次評価者が質の高い評価として認める点となっている。

## 2次評価において、質のよい評価と判断された終了時評価報告書

- エチオピア「地下水開発・水供給訓練計画」
- ヨルダン「情報処理技術向上プロジェクト」
- ヨルダン「家族計画・WIDプロジェクト」
- ケニア「中等理数科教育強化計画」

## (2) 評価の質の経年的な変化

以上では、2002年度と2003年度の終了時評価48件を対象に、評価項目別に評価の質をみてきた。この48件に2001年度の終了時評価を加えて、年度別・項目別に平均点を示したのが表4-3と図4-5である。いずれの年度も、評価方法のところで説明したとおり、対象の1次評価報告書を全評価者が読んで評価した場合の評点を推定して、平均点を算出している。2001年度については、2002年度と2003年度の結果と比較するための追加的な分析であり、評価者数も異なることから、区別するために破線で示してある。

この図からは、年度間で動きに差はあるものの、2002年度と2003年度の間では「評価フレームワーク」、「情報収集」、

「分析／評価」の「現状把握」、「分析」、「評価」のいずれもが0.1から0.2点台で評点が上昇している。このうち「分析」は、3年間継続して0.2点台の上昇をみせている。しかし、「提言・教訓」、「報告書」および「総合評価」については、ほとんど変わっていない。また、「評価可能性」は、下がっている。

以上は対象となった1次評価だけについての結果であり、その年度の終了時評価全体にこうした変化があるかどうかの結果ではない。すなわち、2003年度と2001年度では、対象となったのはそれぞれ10件の1次評価のみであり、すべての終了時評価を2次評価した場合に同様の結果となるかどうかはわからない。このため2次評価では、その年度のすべての終了時評価を対象とした場合に、経年的な質の変化がみられるかどうかについても、統計分析を用いて検証してみた\*。表4-4はこの分析結果をまとめたものである。その年度に行われたすべての終了時評価の2次評価結果の平均点に差があったと判断された項目と、そうでない項目を区分してある。また、差があった場合に、評点が高かった年度を括弧内に記している。

この分析結果では、「評価可能性」については、すべての終了時評価の平均点で見れば、2001／2002年度、2002／2003年度、2001／2003年度のいずれの間でも、評点に差は

表4-3 年度別・項目別の評点の推移

年度	評価可能性	評価フレームワーク	情報収集	現状把握	分析	評価	提言・教訓	報告書	総合評価
2001	6.21	6.58	6.13	6.27	6.33	6.50	6.45	6.11	6.26
2002	6.36	6.46	6.36	6.37	6.50	6.56	6.23	6.48	6.59
2003	6.16	6.65	6.45	6.57	6.77	6.71	6.23	6.50	6.61

図4-5 経年的な評価の質の変化

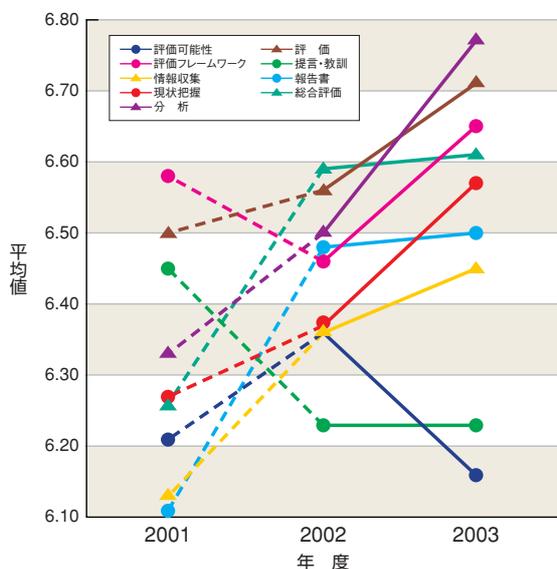


表4-4 経年的な質の変化の検証結果

設 問	2001/2002	2002/2003	2001/2003
<b>I 適切な評価が可能であったかの前提条件 (評価可能性)</b>			
評価可能性	-	-	-
<b>II 項目ごとの二次評価</b>			
評価フレームワーク	-	○(03)	-
情報収集	○(02)	-	○(03)
分析／評価 (現状把握)	-	-	○(03)
分析／評価 (分析)	-	○(03)	○(03)
分析／評価 (評価)	-	-	-
提言・教訓	-	-	-
報告書	○(02)	-	○(03)
<b>III 「良い評価の基準」に基づく総合評価</b>			
総合評価	○(02)	-	○(03)

○は母集団の評点平均点に差があると判断された値、-は差はないと判断された値。  
(02)は2002年度のほうが評点平均点が高いグループ、(03)については2003年度のほうが評点平均点が高いグループ。

\*章末の「テクニカル・ノート」(\*3 年度間の質の差の検証) 参照。

みられないという結果が示されている。

また、「情報収集」、「報告書」および「総合評価」については2001/2002年度間で2002年度のほうが、「評価フレームワーク」と「分析」については2001/2003年度間で2003年度のほうが、「情報収集」、「現状把握」、「分析」、「報告書」および「総合評価」については、2001/2003年度間で2003年度のほうが平均点が高いことが明らかになった。ただし「評価」と「提言・教訓」については、年度間の平均点の差は特にみられないという結果になっている。

以上からは、評価の質については、「提言・教訓」、「評価可能性」などを除き、傾向としては経年的に質が向上しているといえる。

### (3) 外部評価と内部評価の視点の違い

#### 1) 全体傾向

今回の2次評価では、外部評価者と内部評価者の違いによる、評価結果の違いについても検証を試みた。表45と図46は、2002年度と2003年度の終了時評価48件について、評価項目別の2次評価結果の平均点を、外部評価者と内部評価者に分けて示したものである。

外部評価者と内部評価者ともに、項目別の評価傾向はおおむね相似しており、評点は全体的に外部評価者のほ

うが高いことがうかがえる。ただし、「評価フレームワーク」と「総合評価」では、両者の評点差はほかの評価項目に比べて小さい。

外部評価と内部評価の間の差について、質の経年変化と同様に、対象となった年度のすべての終了時評価の2次評価を行った場合にも、外部評価者と内部評価者で差があるかどうかを統計的に分析したのが表46である\*。差があったと判断された項目とそうでない項目を区分したうえで、外部評価者と内部評価者でつけた評点が高かったほうを括弧内に記している。この結果からも、「評価フレームワーク」と「総合評価」を除き、外部評価者のほうが内部評価者よりもつけた評点が高いことが明らかになっている。

#### 2) 評価視点の違い

外部評価者と内部評価者の視点の違いをさらに分析するために、それぞれの評価項目の小項目の平均点を比較したのが図47である。

この図が示すとおり、小項目についても、全体的には外部評価者のほうが評点が高いが、外部評価者のほうが低い点をつけている小項目もいくつかみられる。そのうち、「評価フレームワーク」は、外部評価者のほうがすべ

表4-5 外部評価・内部評価の評点の違い

	評価可能性	評価フレームワーク	情報収集	現状把握	分析	評価	提言・教訓	報告書	総合評価
全体	6.32	6.50	6.38	6.41	6.56	6.59	6.23	6.48	6.60
外部評価者	6.62	6.50	6.73	6.58	6.89	6.84	6.46	6.89	6.68
内部評価者	6.09	6.50	6.12	6.29	6.31	6.41	6.06	6.18	6.53

図4-6 外部評価・内部評価の評点の違い

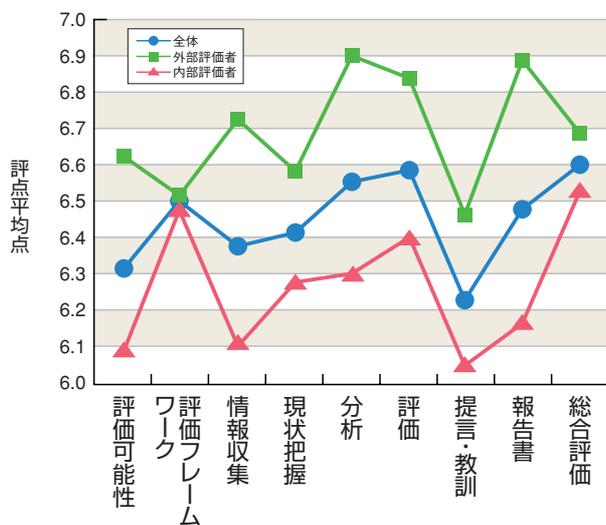


表4-6 外部評価・内部評価の評点の違いの検証結果

設 問	母平均の差の有無
<b>I 適切な評価が可能であったかの前提条件 (評価可能性)</b>	
評価可能性	○ (外部)
<b>II 項目ごとの二次評価</b>	
評価フレームワーク	-
情報収集	○ (外部)
分析/評価 (現状把握)	○ (外部)
分析/評価 (分析)	○ (外部)
分析/評価 (評価)	○ (外部)
提言・教訓	○ (外部)
報告書	○ (外部)
<b>III 「良い評価の基準」にもとづく総合評価</b>	
総合評価	-

○は母集団の評点平均点に差があると判断された値、-は差はないと判断された値。  
(外部)は外部委員のほうが母集団の評点平均点が高いと判定された設問。

\* 章末の「テクニカル・ノート」(\*4 外部評価者と内部評価者の評価結果の差の検証)参照。

での小項目に対して評点が低く、特に「被援助国の参画度」は、外部評価者と内部評価者の評点差が比較的大きい。このほか、外部評価者のほうが評点が低かった小項目には、「情報収集」の「収集方法」および「情報源の適切性」、「現状把握」の「実績の検証」および「プロジェクト実施前後や有無の比較を通じた、効果とプロジェクトの因果関係の検証」、「評価」の「インパクト」評価、「報告書」の「図表の活用」、「総合評価」の「有用性」および「被援助国の参画度」があげられる。

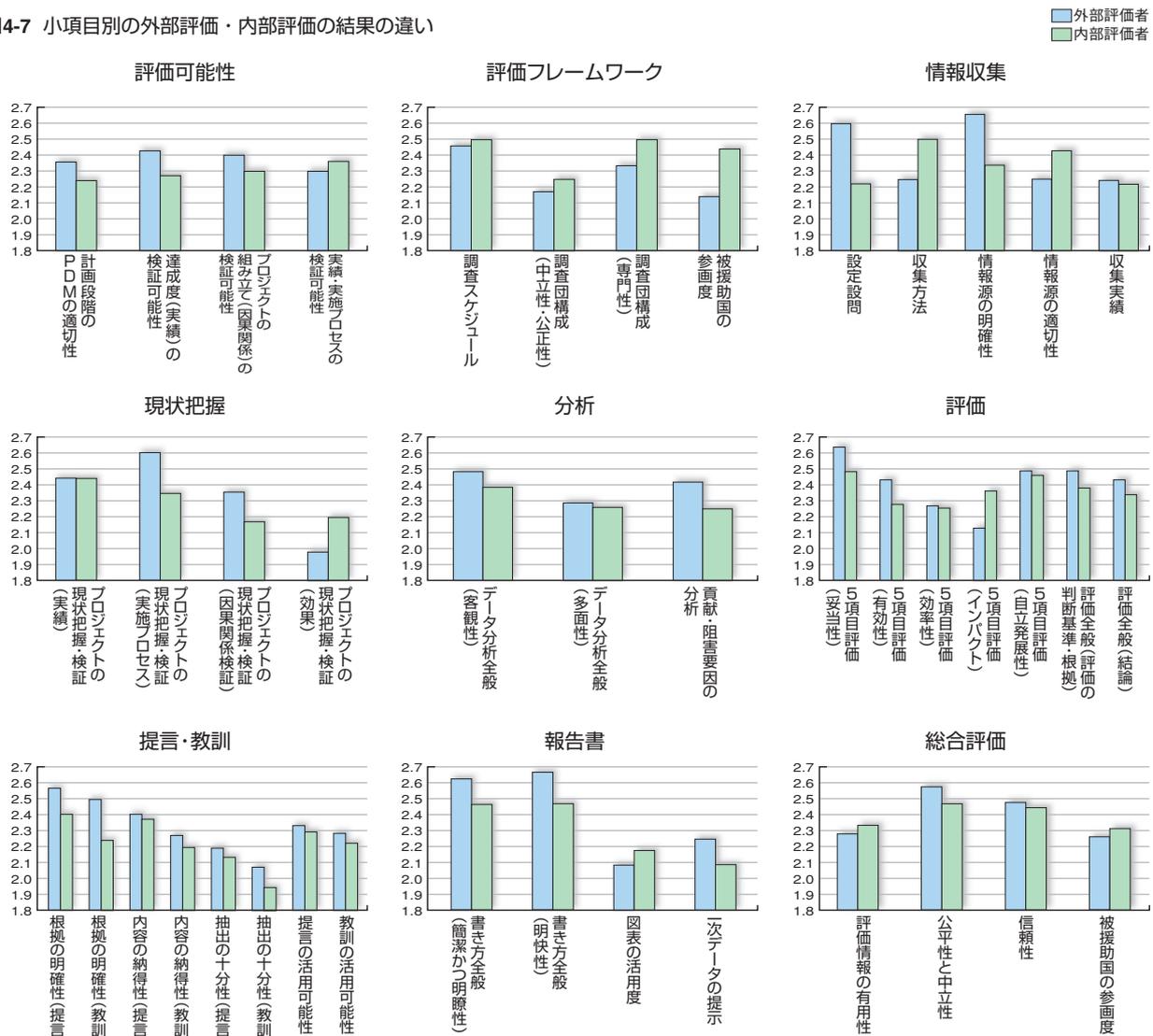
「情報収集方法」の「情報源の適切性」は、すでに述べたとおり受益者からの情報収集の不十分さが課題として指摘されており、「評価フレームワーク」や「総合評価」の「被援助国の参画度」と視点が共通する。また、「現状把握」についての上記の2つの小項目と「インパクト」は、いずれも、効果の適切な検証に関する項目となっている。こうしたことから、外部評価者が内部評価者より

も辛い点をつけた項目は、相手国側関係者の参画や受益者の視点の評価への反映や、適切なデータにもとづく協力効果、特に課題（上位目標）に対する効果の検証に関わるものが多いといえる。また、情報収集方法や図表の活用に関しては、外部評価者の多くが調査研究方法や論文のまとめ方に精通した学識経験者であったことも、内部評価者よりこれらの項目の評点が低かった理由の1つと考えられる。なお、「評価フレームワーク」は、内部評価者がすべての小項目で外部評価者よりも甘い点をつけている。これは調査スケジュールや調査団構成はJICA内ではある程度形が決まったものになっており、それを所与に評価したことによるのではないと思われる。

### 3) 評価傾向の違いについての考察

2次評価では、外部評価者と内部評価者の評価傾向の違いについて、評価結果を用いていくつかの分析も試みた。

図4-7 小項目別の外部評価・内部評価の結果の違い



(注) 縦軸は評点。

まず、評価項目別の評価結果では、外部評価者のほうが総じて評点が高いという結果が示された。しかしながら、これは、全員の平均点での比較である。そこで、個人個人の傾向をみるために、項目ごとに評価者別の評点を高い順にランク付けしてみたのが表4-7である。外部評価者、内部評価者ともに、個人によって傾向は異なるが、全体としてはやはり内部評価者のほうが辛い評点付けを行っていることが示されている。

内部評価と外部評価では、一般的には内部評価のほうが評価が甘いと考えられがちであるが、今回の2次評価で内部評価者のほうが総じて評価が辛いという結果が出た。その要因を分析するために、それぞれの評価者が対象となった地域や分野についてどの程度の知識をもっているかを調査した。対象地域や分野に対する知識が、評価傾向に影響するのではないかと考えられたからである。

このため1次評価ごとに、実際に2次評価を行った評価者に対し、当該地域あるいは分野にどの程度馴染みがあるかを4段階（4点＝非常に馴染みが深い、3点＝ある程度馴染みがある、2点＝一般的な情報として知っている、1点＝まったく知らない）で報告してもらい、外部評価者と内部評価者の平均値を求めた。外部評価者については地域の馴染み度が2.53点、分野の馴染み度が2.27点であり、内部評価者はそれぞれ1.97点および1.79点であ

った。また、地域と分野のいずれの馴染み度とも、外部評価者のほうの標準偏差が大きく、ばらつきがあった。

次に、統計分析を用いて、馴染み度が評価項目の評点にどのように影響しているかを検証した\*。分析結果は、表48に示すとおりである。

評価者の区別に、馴染み度が評点に影響を与える場合とそうでない場合を区別したうえで、評点が高くなる傾向にあるのか、低くなる傾向にあるのかを括弧内に記している。この表からは、外部評価者と内部評価者では、地域や分野に対する馴染み度が評点に与える影響が異なることが示されている。すなわち、外部評価者では地域の馴染み度と評点との間に因果関係がみられないのに対し、内部評価者の場合には対象地域に馴染みがある場合に総じて評点が低くなる、すなわち評価が厳しくなる傾向がある。また、分野の馴染み度については、外部評価者の場合は馴染みがあるほど「分析」や「報告書」の評点付けが、内部評価者の場合は「評価フレームワーク」の評価が厳しくなる傾向がある。以上の結果からは、今回の2次評価で内部評価者の評価が比較的辛かった1つの要因としては、内部評価者の多くが勤務歴が長く、対象地域に対して馴染みがあったことが考えられる。

さらに2次評価では、外部評価者と内部評価者の評価傾向の違いをみるために、統計分析を用いて、個々の評

表4-7 外部評価者・内部評価者の評価傾向

設 問	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位	11位	12位	13位	14位
<b>I 適切な評価が可能であったかの前提条件(評価可能性)</b>														
評価可能性	内部-E	外部-C	外部-D	外部-B	内部-H	外部-E	内部-F	外部-A	内部-C	内部-A	外部-F	内部-D	内部-G	内部-B
<b>II 項目ごとの二次評価</b>														
評価フレームワーク	内部-H	内部-G	内部-E	外部-B	外部-E	外部-F	外部-C	内部-F	内部-C	外部-D	内部-A	内部-B	外部-A	内部-D
情報収集	外部-D	内部-E	外部-B	外部-C	内部-H	外部-E	内部-C	内部-G	外部-F	内部-A	内部-F	外部-A	内部-B	内部-D
分析/評価(現状把握)	外部-B	外部-E	内部-H	内部-E	外部-C	内部-G	外部-D	内部-C	内部-F	外部-A	内部-A	内部-D	内部-B	外部-F
分析/評価(分析)	内部-H	外部-B	外部-D	外部-E	外部-C	内部-E	内部-G	内部-C	外部-A	内部-D	内部-A	外部-F	内部-F	内部-B
分析/評価(評価)	内部-H	内部-E	外部-B	外部-C	外部-D	外部-E	外部-A	内部-C	内部-G	内部-F	内部-A	外部-F	内部-D	内部-B
提言・教訓	外部-E	外部-C	内部-G	内部-E	外部-D	内部-H	外部-B	内部-C	内部-F	内部-A	内部-D	外部-F	外部-A	内部-B
報告書	外部-B	外部-A	外部-D	内部-H	内部-E	外部-E	外部-C	内部-G	内部-C	内部-A	内部-F	内部-B	外部-F	内部-D
<b>III 「良い評価の基準」にもとづく総合評価</b>														
総合評価	内部-H	外部-E	外部-B	内部-G	外部-D	内部-E	内部-C	外部-C	外部-F	内部-F	内部-A	内部-D	外部-A	内部-B
<b>IV 報告書から読み取れるプロジェクトの評価(5項目評価)</b>														
妥当性	内部-H	内部-E	外部-B	外部-C	外部-E	外部-D	外部-A	内部-G	内部-C	外部-F	内部-A	内部-F	内部-B	内部-D
有効性	内部-H	外部-B	外部-D	外部-E	内部-E	外部-C	外部-A	内部-C	外部-F	内部-G	内部-A	内部-F	内部-D	内部-B
効率性	内部-H	外部-E	外部-A	内部-E	外部-C	外部-B	外部-D	内部-C	内部-G	内部-A	内部-F	外部-F	内部-D	内部-B
インパクト	外部-D	外部-B	内部-H	外部-E	内部-E	外部-A	内部-G	外部-C	内部-C	内部-A	内部-D	内部-F	内部-B	外部-F
自立発展性	外部-D	内部-H	外部-C	外部-B	外部-A	内部-E	外部-E	内部-C	内部-D	外部-F	内部-G	内部-A	内部-F	内部-B
総合評価	内部-H	外部-D	外部-B	外部-E	内部-E	外部-C	外部-A	内部-C	内部-G	内部-F	内部-A	外部-F	内部-D	内部-B

\* 章末の「テクニカル・ノート」(\*5 馴染み度の評点への影響の分析) 参照。

表4-8 地域・課題への馴染み度と各設問への影響

分析の種類	地域の馴染み度		課題の馴染み度	
	外部評価者	内部評価者	外部評価者	内部評価者
<b>I 適切な評価が可能であったかの前提条件（評価可能性）</b>				
評価可能性	-	-	-	-
<b>II 項目ごとの二次評価</b>				
評価フレームワーク	-	○(辛)	-	○(辛)
情報収集	-	-	-	-
分析／評価（現状把握）	-	○(辛)	-	-
分析／評価（分析）	-	-	○(辛)	-
分析／評価（評価）	-	○(辛)	-	-
提言・教訓	-	○(辛)	-	-
報告書	-	-	○(辛)	-
<b>III 「良い評価の基準」にもとづく総合評価</b>				
総合評価	-	○(辛)	-	-

○は評点に影響を与える場合、-は与えない場合。(辛)は設問の評点が辛くなる傾向。

評価者が評価項目と小項目の何を重視して評点づけを行ったかを分析して、同じような評価者をグループ分けした\*。この結果が図4-8であり、左側で線が結合しているほど、評価傾向が近いことを示している。また、この分析結果にもとづいて、仮に評価者を2つのグループに分けてみたのが表4-9である。評価者の評価傾向を2グループに分けた場合、内部評価者のほうは75%が同じグループに分類されるのに対して、外部評価者は67%であり、内部評価者のほうが評価傾向が類似していることが示された。

内部評価者のほうが評価者間の評価傾向が似ており、外部評価者のほうが評価者間の評価傾向にばらつきがあるのは、内部評価者は同じ組織で同じような業務に従事している一方、外部評価者は専門分野やバックグラウンドも違うことを考えれば、ある種当然といえる。また、評価傾向が似ていたり、ばらつきがあったりすること自体に問題があるわけではない。しかしながら、以上は、評価傾向の類似性が高い内部評価者だけが評価する場合、評価の視点が偏る可能性があるこ

とを、また外部評価者であっても個人的な評価傾向があり、中立性・公平性の面で必ずしも優れているわけではないことを示唆している。このことから、内部評価に外部評価者の参加をはかっていくことや、外部評価でも複数の視点を盛り込むことが、より信頼性の高い評価結果を得るためには重要といえる。

図4-8 評価傾向による評価者のグループ分け

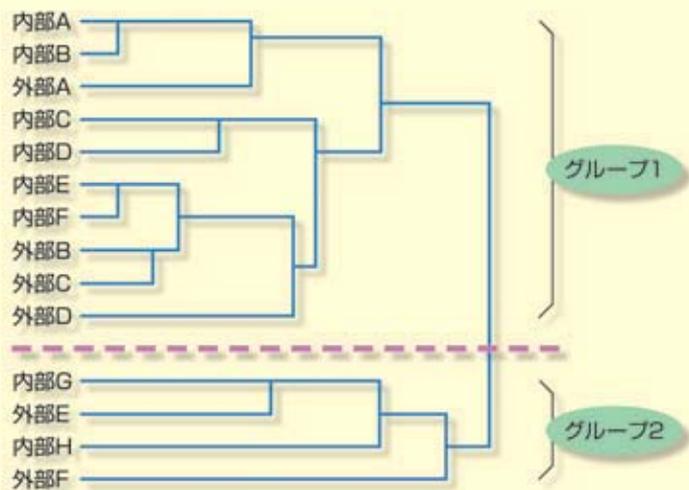


表4-9 評価傾向による評価者の分類

グループ1		グループ2	
外部評価者	内部評価者	外部評価者	内部評価者
外部A、外部B、外部C、外部D (外部の67%)	内部A、内部B、内部C、内部D、内部E、内部F (内部の75%)	外部E、外部F (外部の23%)	内部G、内部H (内部の25%)

\*章末の「テクニカル・ノート」(\*6 評価傾向の違いの分類) 参照。

## 2-2 1次評価対象プロジェクトの成果の検証

### (1) 2次評価結果

報告書をもとに、2002年度と2003年度の終了時評価48件の対象プロジェクトを、DAC評価5項目と総合評価で評価した。2次評価なので、情報面での制約と、報告書の質の差という制約はあるが、報告書から読み取れるプロジェクトに対する評価結果（各項目の平均点）は図4-9のとおりである。

「総合評価」は6.34点で、採点基準の「普通」（5～6点

図4-9 評価項目別平均点

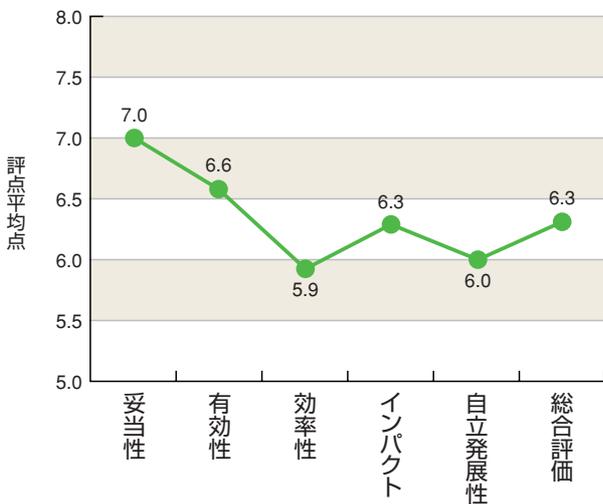
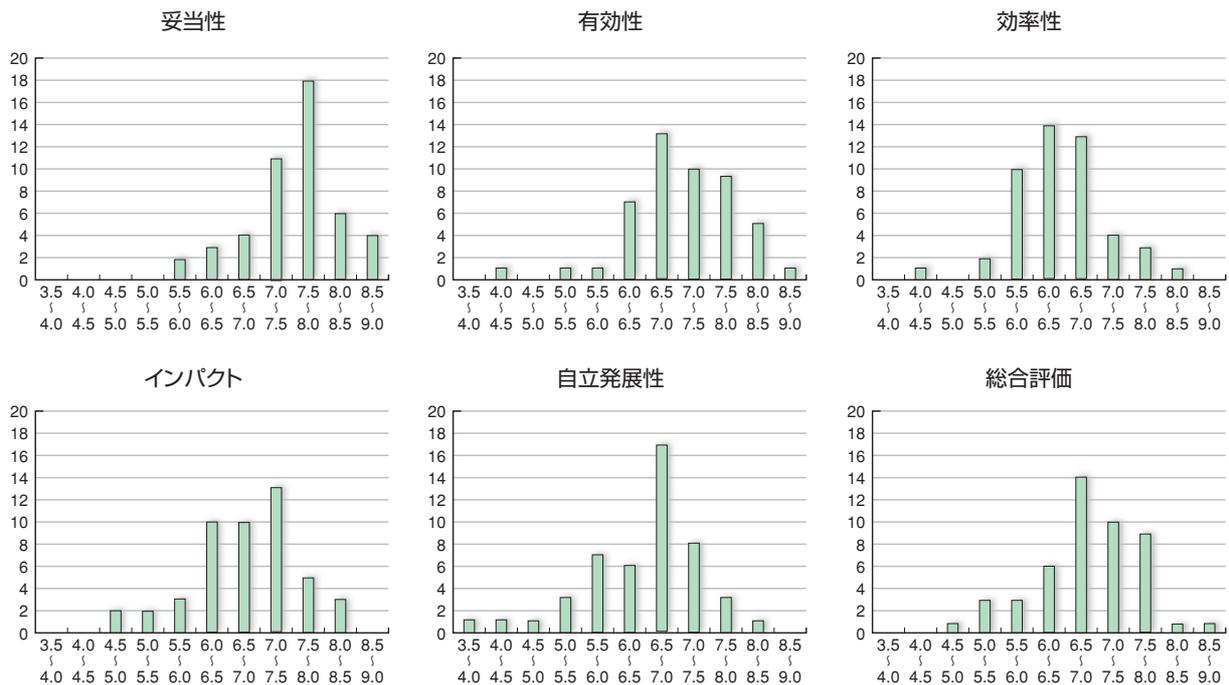


図4-10 対象プロジェクトの評価項目別得点分布



(注) 縦軸はプロジェクト数

台) のレベルにある。評価5項目については、「妥当性」の評点が最も高く「やや良い」（7～8点台）のレベルにある。残りの4項目も5～6点台で「普通」のレベルにあるが、「効率性」と「自立発展性」はやや低い評価となっている。

次に得点分布をみる。図4-10は、対象プロジェクトごとにそれぞれの評価項目の平均点を算出して、得点分布を示したものである。全体として5.5点から8.0点の分布が多い。そのうち、「妥当性」については大半のプロジェクトが7点台以上で、5点を下回るものはない。また、「有効性」と「インパクト」については多くが6.0～7.5点の間にあり、8点台以上のものもある一方、5点を下回るものも見られる。さらに「効率性」は5.5～7.0点の間のもが多く、全体としてやや評価が低い。「自立発展性」は、最も多いのは6.5～7.0点であるが、ばらつきが大きく、3点台のものもあった。

### (2) 評価結果の分析

2次評価では、プロジェクトによって成果が異なることが示されたが、対象地域や分野によって、プロジェクトの成果に影響を与える要因に違いがあるのかどうかを分析するために、対象プロジェクト48件を地域別、分野別に分類し、項目別の評価結果の平均点を示したのが図4-11と図4-12である。

図4-11 地域別・項目別評価結果

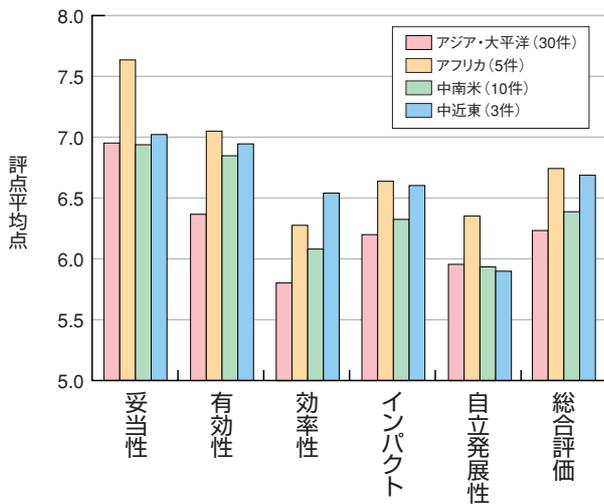
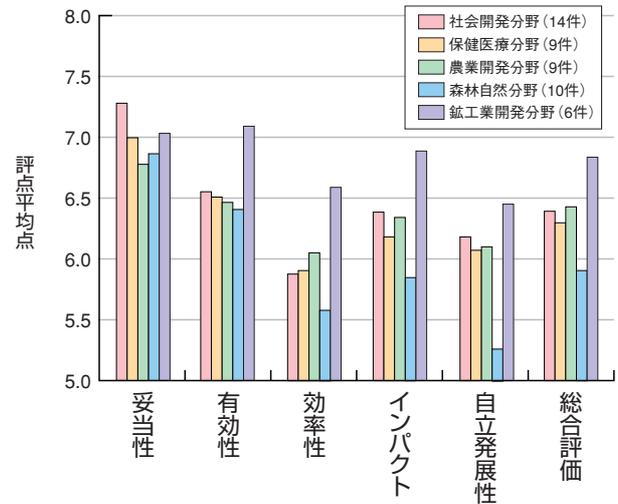


図4-12 分野別・項目別評価結果



### 1) 地域別分析

地域別では、アジア・大洋州地域以外は対象プロジェクト数が限られており、これらのみから地域別の傾向を分析するのは適当ではないが、今回の2次評価結果に限って言えば、アフリカ地域の平均点が総じて高く、特に妥当性に対する評価が高い。また、アジア・大洋州地域の平均点がほかにくらべて低い。

アフリカ地域の平均点が高かったのは、今回の2次評価の対象となった案件のなかに、非常に成功したプロジェクトがあったこと、また、残りのプロジェクトも5.0点を下回るような項目がなく、総じて良好な評価を得たことが理由にあげられる。前者はケニア「中等理数科教育強化」であり、DAC5項目と総合評価のすべてにおいて、48件中で最も高い評価を得ており、「有効性」と「総合評価」では、唯一8.0点以上の評価結果を得ている。また、各項目について「やや良い」（7～8点台）という評価を得たプロジェクトをみると、エチオピア「地下水開発・水供給訓練計画」が「妥当性」の8点台をはじめ、「有効性」、「効率性」、「インパクト」および「総合評価」で7点台の評価を得ている。

この2つのプロジェクトは、1次評価の質でも高い評価を得ているが、プロジェクトの計画とモニタリングを含む運営が適切だったことが質の高い1次評価につながり、質の高い1次評価が、プロジェクトに対する以上のような2次評価につながったといえる。

一方、アジア・大洋州地域については、プロジェクトの数が多いことからばらつきがある。DAC5項目の「妥当性」、「有効性」、「インパクト」、「自立発展性」および「総合評価」で7点台の評価を得たインドネシア「母

と子の健康手帳」のようなプロジェクトもあるが、カンボジア「理数科教育改善計画」やフィリピン「治水・砂防技術強化」のように、「総合評価」を含む複数の項目で「やや悪い」（3～4点台）という評価となったプロジェクトもある。

このうちカンボジア「理数科教育改善計画」は、カンボジアの基礎教育の質を改善するために、中等理数科教員の養成・訓練のための中・長期計画を策定して、教員養成校の機能・能力の強化を支援するプロジェクトである。中・長期計画の策定については目標をほぼ達成しているが、教員養成校の機能・能力強化については、当初の予想を上回るカンボジア側教員の基礎能力の低さや訓練設備の悪さから、計画の見直しが必要となり、当初のプロジェクト期間内に十分に目標が達成できなかった。こうしたことから、2次評価でも「妥当性」はあるが、「有効性」、「効率性」、「自立発展性」および「総合評価」は低いとの評価になっている。

また、フィリピン「治水・砂防技術協力強化」は、台風による洪水や土砂崩れなどで毎年多くの人的・資金的損害が生じているフィリピンで、新しく設置された防災事業実施機関の体制強化を、特に地方の現場事務所関係者の能力強化を中心に支援するプロジェクトである。このために、技術基準の改善、技術者訓練、防災事業に必要な基礎的情報システムの確立などの協力を行うこととなっていた。しかし、フィリピン側の予算が慢性的に不足して投入が遅れたり、防災関連情報の不足や誤りなどから追加調査が必要になったりと、活動が遅れが生じたため、計画した成果が達成できなかった。そのため、2次評価では、「有効性」、「効率性」、「自立発展性」で低い

評価となった。また、「妥当性」についても、必要性は認められるが、フィリピン側の優先度に疑問を呈するコメントもあり、結果として低い「総合評価」になっている。

なお、カンボジア「理数科教育改善計画」とフィリピン「治水・砂防技術強化」は、終了時評価結果を受けて、協力期間を延長し、現在、目標達成に向けた協力を実施している。

中近東には、それぞれの項目で5.0点を下回る得点となったプロジェクトはない。「やや良い」（7～8点台）という評価が多かったプロジェクトとしては、「妥当性」、「有効性」、「効率性」、「インパクト」および「総合評価」で7点台を得たヨルダン「情報処理技術向上」があげられる。これも、先に紹介したケニアやエチオピアのプロジェクトと同様に、1次評価の質もよかったプロジェクトである。

また、中南米では、エルサルバドル「農業技術開発普及計画」が「妥当性」、「有効性」、「インパクト」、「総合評価」で7～8点台の評価を得ており、特に「総合評価」はケニアに次いで高い。一方、中南米では、ブラジル「アマゾン研究計画」が「妥当性」、「有効性」、「効率性」では「普通」（5～6点台）と評価されているが、「インパクト」と「自立発展性」がやや低く、「総合評価」も4点台となっている。このプロジェクトは、アマゾンの森林資源を持続可能な形で活用していくために、森林保全や荒地回復に関する技術の改良を支援するプロジェクトである。このために、研究機関を対象として、荒地の立地特性、森林保全と荒地回復に重要な樹種の特性の解明などに対する技術移転を実施している。計画した成果は達成されて、研究所の能力も高まっていることから、技術面では自立発展性を有している。しかし、予算不足による財政面の問題が指摘されており、2次評価でも



エチオピア「地下水開発・水供給計画」では、「妥当性」をはじめとする各項目で良好な評価結果が得られた

「自立発展性」の評価が低くなっている。また、2次評価者からは、基礎研究は重要ではあるが、技術改良には至っておらず、技術の活用につながるルートもみえないとして、課題に対する貢献の観点から問題を提起する意見が少なからずみられる。これが「インパクト」や「総合評価」の評価結果につながっている。

## 2) 分野別分析

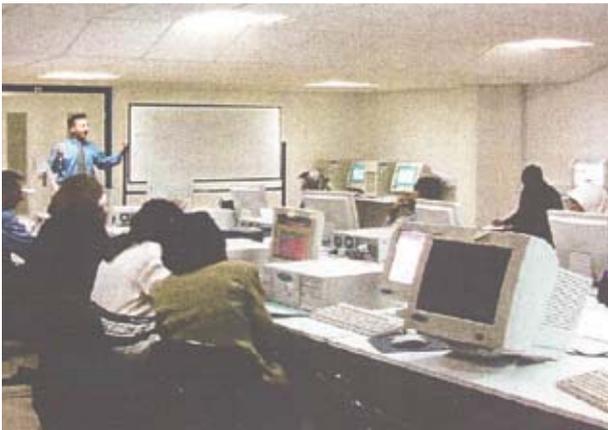
分野別についても、分野によって対象プロジェクト数が異なるため、今回の2次評価結果のみから傾向を分析することはできない。しかし、対象とした48プロジェクトに限れば、図4-12に示されるように、全体的に鉱工業開発分野のプロジェクトの平均点が高い。また、森林自然分野の平均点が総じて低く、特に自立発展性が低い。

鉱工業分野の平均点が高いのは、既述のヨルダン「情報処理技術向上」をはじめ、フィリピン「電気・電子製品試験技術」、ラオス「電力技術基準整備」、パラグアイ「質量分野検定・検査技術向上」など、「総合評価」において7点台、かつ他の項目についても6点台の、総じて評価がよかったプロジェクトが多かったことによる。

一方、森林自然分野では、ブラジル「アマゾン森林研究計画」の「総合評価」が4点台であるほか、残りのプロジェクトも「普通」（5～6点台）であることが全体として平均点がやや低かった要因となっている。また、「自立発展性」については、ラオスの「森林保全・復旧計画」と「養殖改善・普及計画」の2つのプロジェクトでいずれも4点台と評価されたことが平均点を低めている。このうち「森林保全・復旧計画」については、技術面の自立発展性はあるが、組織面と財政面の自立発展性の問題が、「養殖改善・普及計画」についても、財政面の自立発展性の問題が指摘されている。このことから、この2件については、分野というより対象国に起因する要素も少なくないと考えられるが、ラオス「電力技術基準整備」については、政府がコミットメントを明言しているとして特に問題にされていない。1次評価の情報にもとづくため制約もあり、一概には比較できないが、政府の優先度は財政面での自立発展性に関係するため、これが分野による違いに反映していることも考えられる。ただし、政府の優先度が、開発ニーズに照らした必要性といつも合致しているわけではない。

## 3) 外部評価と内部評価の視点の違い

報告書から読み取れるプロジェクトの評価についても、1次評価の質と同様に、内部評価と外部評価の評価結果を

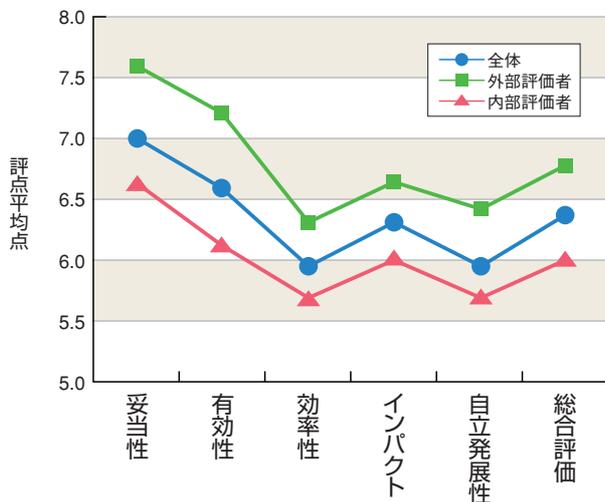


ヨルダン「情報処理技術向上」では、質の高い評価が行われると同時に、案件自体にも良い評価結果が得られた

比較した。図4-13は、評価項目別の平均点を、外部評価者と内部評価者に分けて示したものである。また、表4-10は、評価の質について行ったのと同じ手法を用いて、外部評価者と内部評価者の差について統計的に分析した結果である。

これらの図表が示すように、全体として内部評価者のほうが、報告書から読み取れるプロジェクトの評価についても厳しく評価している。そのうち特に「妥当性」と「有効性」については平均点の差が大きく、内部評価者のほうがかなり辛めの評価をしている。ただし、「効率性」と「インパクト」では、内部評価者と外部評価者の差は比較的少ない。これは1次評価の質に対する2次評価において、DAC評価5項目のうち、「インパクト」は外部評価者のほうが内部評価者よりも辛い評価をしていたこと、「効率性」も残りの項目に比べれば両者の差が少なかったことと、整合する結果となっている。

図4-13 評価項目別平均点



### 3 評価結果の総括と提言

#### (1) 1次評価の質

2次評価結果からは、JICAのプロジェクト終了時評価の質は、全体として一定のレベルは満たしているといえる。また、経年的な質の向上もみられ、特に「情報収集」、「現状把握」、「分析」などは着実に質が高まっていることがうかがえる。しかし、その一方で、「提言・教訓」はほかの項目に比べてやや質が低く、質の向上の面でもまだあまり進展がみられない。「提言・教訓」は、評価結果を活用して事業を改善していくためにも重要であり、質の向上に向けたさらなる取り組みが望まれる。

評価方法については、評価に際して被援助国側の参画度をさらに高めるとともに、プロジェクト関係者や機関のみでなく、受益者からも情報を収集することが必要である。また、定性・定量双方の情報・データを十分に収集したうえで、複数の分析手法を組み合わせ、より掘り下げた分析を行うことが求められる。この関連では、定量的なデータの収集や、定量的な分析手法の活用などといった、定量的な面がやや弱い。さらに、上位目標の現状把握や、上位目標に対するプロジェクトの効果などの視点が、総じておざなりになっている傾向がみられる。プロジェクトの終了時点では、上位目標に対する効果の発現に限られるものが少なくないことも一因と考えられるが、課題解決のためにプロジェクトを実施していることから、こうした効果が上がっているのかいないのかは、まさにプロジェクトの価値を決める点であり、評価においても大切な点である。

なお、適切な評価を実施するには、プロジェクト自体において目標や指標を適切に設定して、効果につながる論理を適切に構築していることが重要な役割を果たす。計画やデータが十分でない、評価可能性の低いプロジェクトを評価することはむずかしく、評価の質も低くなる

表4-10 外部評価・内部評価の違いの検証結果

設問	母平均の差の有無
報告書から読み取れるプロジェクトの評価 (5項目評価)	
妥当性	○ (外部)
有効性	○ (外部)
効率性	○ (外部)
インパクト	○ (外部)
自立発展性	○ (外部)
総合評価	○ (外部)

○は母集団の評点平均点に差があると判断された値、一は差はないと判断された値。  
(外部)は外部委員のほうが母集団の評点平均点が高いと判定された設問。

傾向がみられる。当然ながら、評価可能性の低いプロジェクトであっても、質の高い評価もあり、逆にプロジェクト自体は評価可能性が高くて、質の低い評価もある。しかし、2次評価の結果、質のよい評価として選出された4つの終了時評価とその対象プロジェクトに対する2次評価結果が示すように、プロジェクトの適切な計画、モニタリング、運営は、質の高い評価を実施するために重要である。また、質の高い評価は、良い成果を上げているプロジェクトが、報告書を通じて読者に正当に評価されるうえで重要である。

2次評価の対象となった1次評価のなかには、「評価フレームワーク」、「情報収集」、「分析／評価」、「提言・教訓」および「報告書」の各項目、または複数の項目で、すでに述べた4件の評価と同じように、ほかの参考となるような質の高い評価もあった。こうした質の高い評価がどのように評価を実施したり、評価結果をまとめたりしているかを、広く共有することは、評価の質を改善していくうえで有益と考えられる。たとえば、報告書のなかには、レポート・ライティングの技術が不十分なものも少なくなかったが、こうした問題は、レポート・ライティングの面で優れた報告書を参考にすることで、比較的早期に改善がはかれるものと考えられる。また、報告書のまとめ方について、関係者のための執務資料としての用途を主として想定しており、対外的な説明の用途はあまり意識されていないようなまとめ方の報告書が少なくないが、認識の統一が必要と思われる。

## (2) 報告書から読み取れるプロジェクトの評価

対象プロジェクトは全体として「妥当性」が高く、結果としても一定レベルの成果を達成しているといえる。しかし、高い成果を上げたプロジェクトがある一方で、少数ではあるが評価の低いプロジェクトも存在し、「自立発展性」はプロジェクトの間の差が大きい。ちなみに「自立発展性」については、2001年度のプロジェクト終了時評価40件を対象とした第1回の2次評価でも、プロジェクト間の差が大きいという結果が出ている。JICAプロジェクトの「自立発展性」を全体として高めていくためには、2次評価結果なども活用して、「自立発展性」の高いプロジェクトを特定したうえで、それらのプロジェクトの経験を分析するとともに、結果を組織全体の資産として広く共有し、グッド・プラクティスを広めていくことが有益と考えられる。

また、「効率性」については、ほかの項目に比べて全体的に評価が低い。「効率性」についての評価者のコメントをみると、プロジェクトによっては、人や機材を投入するタイミングなどが問題として指摘されている。しかし、投入の経済性や費用対効果についての記述が十分でないことから、「効率性」が高いと評価することがむずかしいという意見が多い。「効率性」については、第1回の2次評価での「報告書から読み取れるプロジェクトの評価」でも、同様の理由で低い評価になっている。コスト面の視点からの評価が不十分であることは、1次評価の質の検証でも「効率性」の評価にかかわる問題として指摘されている。プロジェクトにコスト面の視点を盛り込んでいくことは、プロジェクトの計画、実施、評価における課題の1つである。

## (3) 2次評価チェックシートの改善について

今回の2次評価作業部会では、2次評価作業結果をふまえて、2次評価チェックシートの改善について提言することも作業の1つとした。具体的には、評価項目ごとの小項目の重みづけを提案すること、チェック項目について改善すべき点を提案することである。

このため2次評価作業部会では、統計分析を用いて、それぞれの評価者が評価項目と小項目の何を重視して評価を行ったのかを分析して、統計分析案を策定した。その策定とともに、2次評価チェックシートの小項目のなかで、互いに相関関係が強く、項目を統合しても類似の情報が得られるものを統計的に分析して、小項目の削減・合体案を策定した\*。

そして、これらを参考に、2次評価結果にもとづいて、今後、評価の質をさらに高めていくうえで重視すべき項目やその重みについて議論のうえ、チェックシートの項目と重み付けについての作業部会案を策定した。同案は表4-11に示すとおりである。重み付け案を策定する際には、それぞれの小項目の持ち点を5点とし、大項目の得点(小項目の合計点)を算出したうえで、大項目ごとに小項目間の重みを検討し、大項目の得点を重みに応じて割り振る形で小項目の得点を決めている。大項目ごとに小項目の数が異なることから、大項目の得点は同じではない。また、「1-1(4)評価方法」に記したとおり、大項目を合計した「総合得点」が評価の質をただちに表すわけではないことから、「総合得点」は設置していない。よって大項目間にも重み付けは設定していない。

\* 章末の「テクニカル・ノート」(\*7 小項目の重み付けに関する統計分析案 \*8 評価項目の精選) 参照。

表4-11 2次評価チェックシート改善案

大項目	小項目	作業部会案
評価可能性	計画段階のPDMの適切性	4
	達成度(実績)の検証可能性	4
	プロジェクトの組立て(因果関係)の検証可能性	6
	実績・実施プロセスの検証可能性	6
評価フレームワーク	調査スケジュール	3
	調査団構成(中立性・公正性)	6
	調査団構成(専門性)	
	被援助国の参画度	6
情報収集	評価設問	4
	収集方法	8
	情報源の適切性	
	情報源の明確性	2
	収集実績	6
分析/評価(現状把握)	プロジェクトの現状把握・検証(実績)	4
	プロジェクトの現状把握・検証(実施プロセス)	4
	プロジェクトの現状把握・検証(因果関係検証)	6
	プロジェクトの現状把握・検証(効果)	6
分析/評価(分析)	データ分析全般(客観性)	5
	データ分析全般(多面性)	6
	貢献・阻害要因の分析	4
分析/評価(評価)	5項目評価(妥当性)	5
	5項目評価(有効性)	5
	5項目評価(効率性)	5
	5項目評価(インパクト)	3
	5項目評価(自立発展性)	5
	評価全般(評価の判断基準・根拠)	0
	評価全般(結論)	7
提言・教訓	根拠の明確性(提言)	6
	内容の納得性(提言)	
	抽出の十分性(提言)	4
	提言の活用可能性	5
	根拠の明確性(教訓)	6
	内容の納得性(教訓)	
	抽出の十分性(教訓)	4
	教訓の活用可能性	5
報告書	書き方全般(簡潔かつ明瞭性)	7
	書き方全般(明快性)	
	図表の活用度	4
	一次データの提示	4
総合評価	評価情報の有用性	5
	公平性と中立性	5
	信頼性	5
	被援助国の参画度	5

最後に、今後の2次評価の実施に向けて、今回の作業プロセスで気づいた点を記しておく。まず、2次評価は、1次評価の報告書にもとづいて評価するという制約がある。このため、報告書に十分な情報が盛り込まれていないと判断がむずかしい。2次評価で必要な情報は、2次評価のためのみでなく、対象となった1次評価が適切に行われたかどうかを第三者が判断するうえでも必要な情報である。その意味で、すでに述べた「調査団構成」についての情報や、「評価方法」についての情報などは、報告書にきちんと記載することが望まれる。また、「評価可能性」など、プロジェクトそのものについての項目には、ほかの項目以上に判断にむずかしさがともなう。

さらに、今回の2次評価では、実際の2次評価で用いたチェックシートの考案も含めて、同一の報告書を用いて全委員で数回にわたるトライアル評価を実施した。この過程で、評価視点や小項目の着眼点をもつ意味についての共通認識が深まり、評価者間の評価結果のばらつきも標準化された。2次評価を異なる評価者が行う場合には、評価視点について評価者間で認識の共有をはかることが重要であり、トライアル評価はそのための有益な方法といえることができる。