

## 第3部／外部評価

### －外部有識者評価委員会による2次評価



# 第3部 外部評価

## —外部有識者評価委員会による2次評価

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図るとともに、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を設置しました。委員は開発援助や評価に関する知識や経験を有する外部の有識者（学識経験者、NGO、ジャーナリスト等）によって構成されています。（「第1部第2章 2-5 外部評価の拡充」）

外部有識者評価委員会は、2002年6月の第1回会合以来、ほぼ隔月で開催されてきており、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議論いただくとともに、評価体制の強化に向けての課題、新たな評価対象事業における評価のあり方、評価結果の公表方法の改善など、幅広い事項について助言をいただいています。

特に、外部有識者による2次評価については、客観性や透明性の確保に加え、第三者の視点でその質の現状と課題を御指摘いただくことでJICAの評価の質の向上に役立てることを目的に、2001年度に実施された旧プロジェクト方式技術協力事業の終了時評価全40件の評価を行っていただきました。同2次評価結果の全文は第1章に掲載のとおりです。

同2次評価では、JICA内部においても日頃から問題として指摘されてきている点、また、第三者の視点を通すことによって初めて明らかになった点を含め、JICAの評価が現在抱える問題があまるところなく抽出されていると考えます。JICAでは、同評価結果と委員からの提言をふまえ、後述（「第2章外部有識者2次評価結果を受けて—JICAの対応」）のように評価の質の改善に向けて具体的な取り組みを図っていくつもりです。また、引き続きこうした形での2次評価を拡充していきたいと考えています。

### 外部有識者による2次評価の評価者

#### 外部有識者評価委員会委員長

牟田博光（東京工業大学教育工学開発センター長）

#### 外部有識者評価委員会委員（アイウエオ順）

青山温子（名古屋大学大学院医学系研究科教授）

池上清子（国連人口開発基金東京事務所長）

熊岡路矢（日本国際ボランティアセンター代表理事）

杉下恒夫（茨城大学人文学部教授、元読売新聞社）

長尾真文（広島大学教育開発国際協力研究センター教授）

古川俊一（筑波大学社会工学系教授）

山越厚志（社団法人日本経済団体連合会国際協力本部国際協力グループ副長）

# 第1章●外部有識者評価委員会による2次評価結果

委員長 牟田博光

## 1-1 2次評価の目的

### (1) 手続きの透明性の確保

終了時評価を外部有識者によって2次評価する目的は、大きく次の2つである。まず、透明性の確保が挙げられる。すべてのプロジェクトについて終了時評価が行われているが、評価団員に外部有識者、コンサルタントを交える場合があっても、それらは国内委員会委員や役務提供者など、何らかの意味でプロジェクトの関係者であることが多く、いわば内部評価に近い。関係者は様々な事情に熟知しているところから、きめ細かな評価ができる利点があると同時に、評価の公平性、中立性等の観点からは問題がないとはいえない。

一方、第三者による外部評価であっても、JICAや当該プロジェクトと利害関係が皆無とはいえないし、何よりも後で証明するように誰しも個人的な評価傾向があり、公平性、中立性の観点から必ずしも優れていると結論づけられるものではない。

そこで、内部評価結果を複数の外部有識者評価委員が再評価する事によって、これらの問題の解決を図ることが期待される。

### (2) 結果の信頼性の確保

次に評価結果の信頼性、不偏性の担保である。終了時評価はある特定の評価調査団によって1回行われるだけである。もし、別の評価調査団が評価を行ったとしたら、評価結果はまったく同じではないであろう。もちろん、何度も違う評価調査団を派遣し、評価結果がどのように違うかを調べるようなことはできない。そこで、特定の評価調査団によって1回しか行われない評価結果から、信頼のおける結論を導き出す方法として、1次評価者とは異なる多くの2次評価者が評価作業に加わることが考えられる。

1次評価の結果を基に評価するという制約はあるものの、より多くの人々が評価に関わることによって、偏りの少ない、信頼のおける結論が導かれることが期待できる。この事は内部評価についてだけ当てはまるのではなく、外部評価についても同様に当てはまる。

さらに、2次評価は1次評価実施後1年間程度の時間差をおいて行われるという事実も重要である。単なる見直しだけではなく、効果発現などの新たな証拠が得られる可能性は大きい。

## 1-2 2次評価の方法

### (1) 2次評価の視点

今回の2次評価の視点は二つある。一つは報告書の内容がこれで十分か、評価方法はこれで良いのかといった評価方法や評価報告書の質に関する視点、もう一つは評価報告書を通したプロジェクトそのものの2次評価である。この2つの視点は密接に関連している。今回は平成13年度に実施された終了時評価結果40件を例に取り上げ分析する。ここで用いる2次評価の手法はメタ評価手法の一種である。

### (2) 評価項目ごとの評価点

まず、外部有識者評価委員が終了時評価報告書を読んで、定められた評価票に基づき、評価報告書の質の評価（評価フレームワークに関し8項目、調査の実施状況に関し6項目、情報分析・評価に関し9項目、教訓・提言の内容に関し4項目、計27項目）、プロジェクトそのものの評価（6項目）の観点から、5段階の評点で2次評価を行った<sup>1</sup>。評価票に記載されている評価事項及び評価基準は、JICA事業評価が

<sup>1</sup> 評価者は青山委員、池上委員、熊岡委員、杉下委員、長尾委員、古川委員、牟田委員長、山越委員。

イドラインの「良い評価の基準」を基礎として作成したものである。

全員がすべての評価報告書を読んで2次評価を行うのであれば、各評価項目（評価事項／基準）について平均値を求めれば、委員全体の意見が反映され、委員個人の評価傾向に偏らない評価得点が求められる。各委員はそれぞれ立場や意見が異なるため、ある程度の数の委員の意見を平均化しないと結果が偏るからである。しかし、これでは全体の作業量が莫大で、現実には不可能である。そこで、委員長は全件について、他の委員は6～9件の評価を行うことによって、各報告書を最低2名が評価することとした。この場合、各報告書はそれぞれ2～3名の委員しか評価点をつけていないため、各評価項目について平均値を求めれば、たまたまそれを評価した少数の委員の評価傾向が結果に強く反映することになり、2次評価結果の信頼性、不偏性が問題となる。

各委員がつけた各評価項目の評価点は理論的に、評価対象の真の得点（評価者によらない評価得点）と評価者の評価傾向係数とに分けることが出来る。そこで、統計分析の仕方を工夫することによって、これらを分離し、評価者が一般的に点を甘く付けるか辛く付けるか等の評価傾向の調整を行って、評価対象ごとに評価者の傾向によらない不偏の評価得点を推定することにした。

### （3）評価項目ごとの評価得点の推定値

こうして、まず40件の報告書について、評価項目ごとに評価点の推定値（真の評価得点）と個人の評価傾向係数を求める。ここで、個人の評価傾向係数を評価項目ごとに合計が0になるように調整することによって、評価項目ごとの評価得点の推定値は全員が評価したら得られると考えられる評価点の平均値となる。

表3-1は、各評価項目についてこうして求めた40件の評価得点の平均値及び標準偏差を示している。小評価項目は33項目あるので、より大きなブロックに分けてまとめた値も示す。

### （4）個人の評価傾向

評価項目ごとに、個人の評価傾向係数がどのように変化するかを求めてみると、全般的に評価が高い者、低い者、特定の評価項目に関してだけ評価が高い者、低い者がいるなど様々なタイプがあることが判った。従って、5段階評価のような評価方法では評価者間の判断の違いは避けることが出来ず、ある程度の人数の評価者を確保して平均という形で意見の集約をはかる必要があることが理解される。また、評価者の評価傾向係数間の相関を求めたところ、概して相関は小さくなく、今回の評価者の評価傾向は相互に独立であることを示している。

さらに、評価項目ごとに評価者個人の評価傾向係数の標準偏差を求めた。結果は表3-1に書き込んでいる。評価報告書の質の評価に関しては、0.13（「調査プロセス」の「信頼性」）から0.67（「教訓」の「有用性」）までの値をとり、評価項目によっては40件の項目得点の標準偏差よりも大きい。この値が大きいことは評価者間に意見の差が大きいことを意味している。すなわち、一人で評価結果の信頼性、安定性には本質的な問題があることを示している。評価項目によっては案件の違いによるばらつきより、評価者個人の違いによるばらつきの方が大きいからである。

報告書から読みとれるプロジェクトの評価については、比較的意見がまとまっており、標準偏差は0.22（妥当性）から0.50（効果）までの範囲である。効果や総合評価点で意見の違いが大きい。社会的インパクトをどのように測定、評価するかについてはまだ十分な手法が確定していないことにもよろう。

意見の差の中には、価値観に基づく本質的な差もあるであろうが、言葉の定義や判断基準が明確ではないために起こる差もあろう。しかし、統一的な判断基準を示すなど、多くの数の評価を行う前に、少数の評価例について評価者全員で十分な議論を行う場を設けるなどの準備によって用語や見解の統一を行えば、ある程度は意見の違いを小さくすることが可能であることが、これまでの研究から知られている。

表3-1 評価項目ごとの平均値と標準偏差

|                               |                    |                    |                    |                    |                    | ( ) 報告書の標準偏差<br>(斜体) 評価者の評価傾向係数の標準偏差 |
|-------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------------------------|
| <b>I. 評価調査／報告書の評価</b>         |                    |                    |                    |                    |                    |                                      |
| 事項                            | 有用性                | 公平性・中立性            | 信頼性                | 被援助国側              | 総計得点               |                                      |
| <b>1. 評価フレームワーク</b>           |                    |                    |                    |                    |                    |                                      |
| ・ 評価のタイミング、期間                 | 3.24 (0.40) (0.57) |                    |                    |                    |                    |                                      |
| ・ 評価者／団員構成                    |                    | 2.69 (0.27) (0.55) | 3.12 (0.30) (0.58) | 3.18 (0.57) (0.47) |                    |                                      |
| ・ 調査プロセス                      | 3.16 (0.51) (0.26) | 3.07 (0.39) (0.37) | 3.12 (0.53) (0.13) |                    |                    |                                      |
| ・ 調査コスト(規模の適性)                | 2.96 (0.40) (0.24) |                    |                    |                    |                    |                                      |
| 合計                            | 3.12 (0.45) (0.40) | 2.88 (0.38) (0.46) | 3.12 (0.43) (0.42) | 3.18 (0.57) (0.47) | 3.07 (0.46) (0.46) |                                      |
| <b>2. 調査の実施状況</b>             |                    |                    |                    |                    |                    |                                      |
| ・ 調査事項 (TOR)                  | 3.25 (0.49) (0.39) |                    |                    |                    |                    |                                      |
| ・ 情報収集方法                      | 3.16 (0.55) (0.43) | 2.86 (0.47) (0.42) | 3.10 (0.51) (0.14) |                    |                    |                                      |
| ・ 訪問先／面会者構成                   |                    | 2.80 (0.38) (0.48) |                    | 3.28 (0.47) (0.29) |                    |                                      |
| 合計                            | 3.20 (0.52) (0.41) | 2.83 (0.43) (0.45) | 3.10 (0.51) (0.14) | 3.28 (0.47) (0.29) | 3.08 (0.51) (0.40) |                                      |
| <b>3. 情報分析／評価</b>             |                    |                    |                    |                    |                    |                                      |
| ・ 情報の整理／分析                    | 3.13 (0.75) (0.39) | 3.00 (0.44) (0.54) | 3.23 (0.65) (0.25) | 3.23 (0.50) (0.32) | 3.15 (0.60) (0.40) |                                      |
| ・ 5項目評価の結果                    |                    |                    |                    |                    |                    |                                      |
| 1) 妥当性                        | 3.45 (0.60) (0.36) |                    |                    |                    |                    |                                      |
| 2) 有効性                        | 3.25 (0.56) (0.41) |                    |                    |                    |                    |                                      |
| 3) 効率性                        | 2.76 (0.63) (0.56) |                    |                    |                    |                    |                                      |
| 4) 効果                         | 3.33 (0.64) (0.40) |                    |                    |                    |                    |                                      |
| 5) 自立発展性                      | 3.45 (0.56) (0.46) |                    |                    |                    |                    |                                      |
| 合計                            | 3.25 (0.65) (0.44) |                    |                    |                    | 3.25 (0.65) (0.44) |                                      |
| ・ 教訓／提言の内容                    |                    |                    |                    |                    |                    |                                      |
| 1) 教訓                         | 3.14 (0.61) (0.67) |                    |                    |                    |                    |                                      |
| 2) 提言                         | 3.21 (0.61) (0.44) | 3.01 (0.38) (0.46) | 3.12 (0.46) (0.50) |                    |                    |                                      |
| 合計                            | 3.18 (0.61) (0.57) | 3.01 (0.38) (0.46) | 3.12 (0.46) (0.50) |                    | 3.12 (0.53) (0.56) |                                      |
| <b>II 報告書から読みとれるプロジェクトの評価</b> |                    |                    |                    |                    |                    |                                      |
| 基準／事項                         | 5段階評価              |                    |                    |                    | 総計得点               |                                      |
| ・ 妥当性                         | 3.49 (0.67) (0.22) |                    |                    |                    |                    |                                      |
| ・ 有効性                         | 3.27 (0.66) (0.37) |                    |                    |                    |                    |                                      |
| ・ 効率性                         | 2.90 (0.65) (0.39) |                    |                    |                    |                    |                                      |
| ・ 効果                          | 3.23 (0.68) (0.50) |                    |                    |                    |                    |                                      |
| ・ 自立発展性                       | 3.12 (0.87) (0.25) |                    |                    |                    |                    |                                      |
| 合計                            | 3.20 (0.74) (0.36) |                    |                    |                    | 3.20 (0.74) (0.36) |                                      |
| 総合評価                          | 3.21 (0.71) (0.47) |                    |                    |                    | 3.21 (0.71) (0.47) |                                      |

注： 5段階（1：低い 2：やや低い 3：普通 4：かなり高い 5：高い）で評点を付した。

特に、報告書の評価に関する評価項目の中では教訓の有用性について判断の個人差が大きいのが目につく。提言の信頼性の判断についても個人差が大きい。教訓は提言を導いたり、当該プロジェクトの次に行われる類似プロジェクトの改善に用いたりするものであるため、この内容について識者の意見が大きく分けられるということであれば問題である。コメントの中にも、教訓や提言が導かれる根拠について疑問の声も聞かれた。評価結果のフィードバック、活用として最も重要な部分について、評価者の意見が分けられるのは分析が客観的でないということでもあり、報告書における教訓や提言の記述の仕方について、根拠を明示するなどの統一的なガイドライン設定が望まれる。

### (5) 2次評価項目の精選

33の評価項目間の相関を計算してみると、強い相関がみられる項目が多くある。これはもう少し評価項目数を絞っても類似の情報が得られることを意味する。そこで、項目数を少なくすることにより評価を効率的に行うことを考える。客観的に項目数を絞り込むために、因子分析の手法を用いて項目の再検討を行った。評価5項目、及び教訓・提言の事項の有用性に関しては項目削減の対象としないこととし、計算から除外して分析した。

分析の結果、評価の事項に関しては、調査プロセス、評価コスト、調査事項（TOR）、訪問先／面会者構成は他の変数で代替できること、基準に関しては、公平性・中立性と信頼性は同類の指標とみなすことができ、公平性・中立性が高ければ、信頼性が増すと考えられ、これらを別々に分けなくても得られる情報に大きな差はないことなどが明らかになった。

一部について、事項のカテゴリ分けを変更したり、事項と基準を入れ替えたり、有用性を適切性と言い換えたりして、報告書の評価項目として、表3-2に示すような2次評価票を考案した。今回使用した評価票同様、評価はすべて5段階評価で行うものとする。また、簡略化するだけでなく、報告書に

対するコメントの中で指摘されたが元の調査票には盛り込まれていなかった「評価報告書の分かりやすさ」と「報告書の総合評価点」を追加した。

評価項目数は当初の33評価項目から表3-2の22項目へ、3分の2に減少したが、得られる情報量はほとんど減少しない事になる。この新2次評価票は有用であり、今後2次評価において本案の活用を提言する。

## 1-3 2次評価の結果

### (1) 評価調査／報告書の評価

図3-1（109～111頁）は評価報告書の質の評価及び、評価報告書から読みとれるプロジェクトの2次評価結果について、40件の頻度分布図を求めたものである。評価報告書の質の評価結果は表3-1や図3-1に示すように、いずれの評価項目についても評価得点はおおむね普通(3)を上回っており、全体的に比較的高い評価が得られている。しかし、改善点が無いわけではない。まず評価報告書作成にあたっての全般的な問題点を述べ、次いで表3-1、図3-1の定量的分析結果、及び委員からの定性的コメントを評価項目群別にまとめて述べる。

#### 1) プロジェクト・デザイン・マトリックス (PDM)

評価のためにPDMを作り替えている例が多くある。原理的には終了時評価は直近のPDMに基づいて行うべきである。確かに直近のPDMが漠然としていたために、指標を明確にするためにPDMを再考するのはわかるが、少なからぬ報告書において実際のプロジェクトがPDMの上位目標まで到達するには遠すぎるからといって、上位目標を現実に近くように評価用に作り変えたものをPDMeとしている。これでは本末転倒である。本来であれば上位目標がまず定まり、それを達成すべくプロジェクトが設計されているはずである。しかし現実には、これまでのやり方ではプロジェクトがまずあり、書類の上で上位目標を考えてきたからである。成果重視の

表3-2 外部有識者評価新2次評価票

| I. 評価報告書の評価             |       |               |     |      |
|-------------------------|-------|---------------|-----|------|
| 基準<br>事項                | 有用性   | 公平中立性、<br>信頼性 | 適切性 | コメント |
| <u>1. 評価フレームワーク</u>     |       |               |     |      |
| ・評価のタイミング               |       |               | ( ) |      |
| ・評価者/団員構成               |       | ( )           |     |      |
| ・被援助国の評価協力度             |       |               | ( ) |      |
| <u>2. 情報収集/分析</u>       |       |               |     |      |
| ・情報収集                   | ( )   | ( )           |     |      |
| ・情報分析                   |       |               | ( ) |      |
| <u>3. 分析/評価</u>         |       |               |     |      |
| ・5項目評価の結果               |       |               |     |      |
| 1) 妥当性                  | ( )   |               |     |      |
| 2) 有効性                  | ( )   |               |     |      |
| 3) 効率性                  | ( )   |               |     |      |
| 4) 効果                   | ( )   |               |     |      |
| 5) 自立発展性                | ( )   |               |     |      |
| ・教訓/提言の内容               |       |               |     |      |
| 1) 教訓                   | ( )   |               |     |      |
| 2) 提言                   | ( )   | ( )           |     |      |
| 4. 報告書のわかりやすさ           |       | ( )           |     |      |
| 5. 報告書の総合評価点            |       | ( )           |     |      |
| II. 報告書から読み取れるプロジェクトの評価 |       |               |     |      |
| 基準/事項                   | 5段階評価 | コメント          |     |      |
| ・妥当性                    | ( )   |               |     |      |
| ・有効性                    | ( )   |               |     |      |
| ・効率性                    | ( )   |               |     |      |
| ・効果                     | ( )   |               |     |      |
| ・自立発展性                  | ( )   |               |     |      |
| プロジェクトの総合評価点            | ( )   |               |     |      |

5段階評価 1：低い 2：やや低い 3：普通 4：やや高い 5：高い

立場からは、上位目標は変えないで、必要に応じてプロジェクトのやり方を変えるべきである。これからはプロジェクトを通して社会的ニーズ（上位目標）を見るのではなく、ニーズに応える手段としてプロジェクトの善し悪しを判断していかねばならない。

## 2) 報告書の書き方

適切な評価が行われ、わかりやすいと高く評価される報告書がある一方で、データを引用しているがその出典が明記されていなかったり、面談者のリストが整えられていなかったり、情報の信頼性が分からないものがある。また、論理的に書かれていないので読みにくい、数量的データの記載がない、インタビューや質問紙内容及び結果が記載されていない、報告書中の使用用語が統一されていないなど、報告書の質が貧弱なものもある。報告書の書き方に工夫が必要であると同時に、指導が必要である。報告書の書き方のマニュアルを作る事や、評価の高い報告書を手本とする事なども考慮すべきである。

根本的な問題として、儀礼的・社会的に書かれた報告書が多く、具体的なところまで踏み込めていない。もっと厳しいところまで、外部の人が読んで知りたいところを明確にさせる必要がある。その理由として、評価文化が根付いていないこと、フィードバック先が考えられていないこと、分析そのものが十分出来ていないことが挙げられる。

## 3) 評価フレームワーク

「評価フレームワーク」についての評価点はおおむね3点（普通）を少し上回るが、評価者／団員構成の公平性・中立性は他の項目に比べて評価点が低い。構成がプロジェクト関係者に偏っているからである。また、調査コスト（規模の適正）の有用性もいくらか低い。この場合のコストはかかった直接の金額による判断ではなく、団員の専門や人数で判断している。調査内容に比して人数が多いのではないか、あるいは団員の構成内容などから見て、不必要な団員が含まれているのではないかといった視点からの判断である。

評価フレームがしっかりしているものもあるが、弱いものもある。調査内容と比較して、調査期間が長いものもあれば、短すぎて十分でないものもあるなど、内容と期間のバランスがとれていないものが見受けられる。全般的に、調査団員が多く、調査内容を考慮すると調査団員の職務を兼任できる場合も多い。調査期間、団員数は調査コストに直結する要素であることから、評価の効率性にも十分注意を払うべきである。

また、評価者が関係者に偏っている傾向があるので、関係者以外の評価者を得て、より透明性に富む評価を心がけるべきである。案件によっては、専門分野に関して基本的な知識のない者が評価している例がある。逆に、分野の専門性は問題ないが、開発分野に関する視点が十分ではなく、その国やセクター全体を見る視点が不足している場合、他のプロジェクトに関する知識がないため、十分な比較がなされていない例もある。また、評価に関する知識が十分ではなく、分析が不十分、不適切な例もある。

調査時期が、終了時6カ月前などと画一的であるが、場合に応じて柔軟に行うことを考える必要がある。例えば、当初予定の機材納入が遅れた時や、政変があったときに、杓子定規なタイミングで実施するのではなく、明らかな状況変化を加味すべきである。

## 4) 評価の実施状況

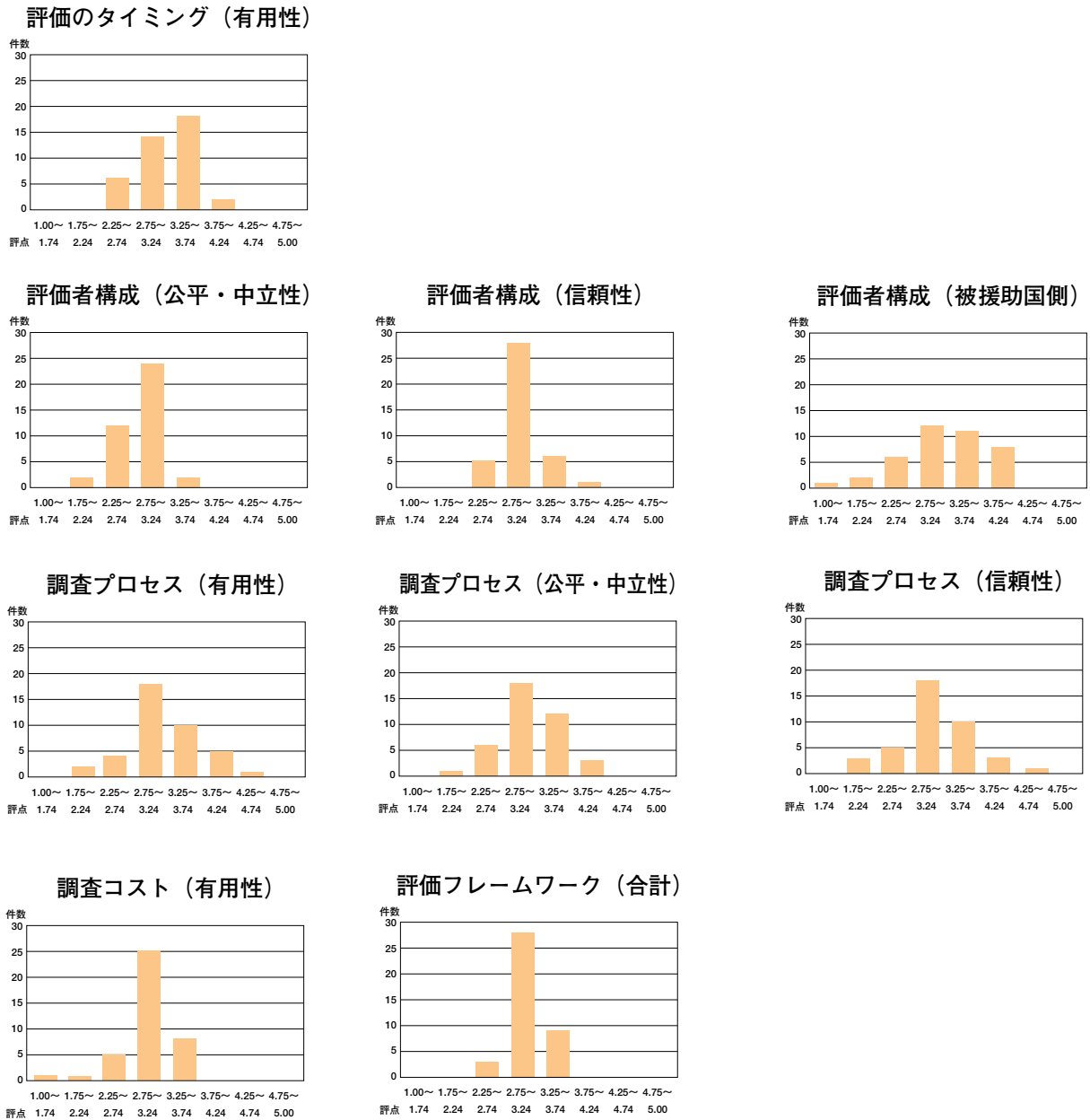
「調査の実施状況」についての評価点もおおむね3点（普通）に近いが、情報収集方法、訪問先／面会者構成の公平性・中立性に対する評価結果は他に比べていくらか低い。もう少し複合的な手法や多様な情報収集先の考慮が望まれる。特に、最終的な裨益者に対する調査の強化が必要である。

調査事項はおおむね適切で、情報収集方法も有益である。また、幅広い情報収集が行われている。しかし、例えば、面会者数が少ない、あるいは面会者リストが不備で、情報の入手先が不明な報告書もある。また、プロジェクトの直接、間接の裨益者である研修生、農民などから意見聴取が十分行われてい

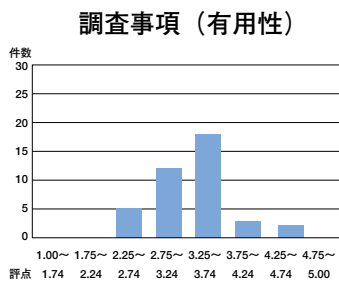


図3-1 報告書の質と報告書から読みとれるプロジェクトの2次評価得点分布

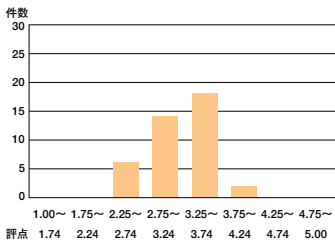
1 評価フレームワーク



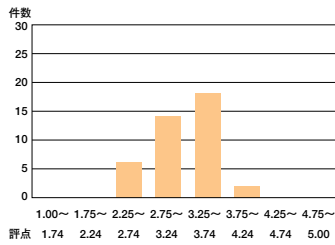
2 調査の実施状況



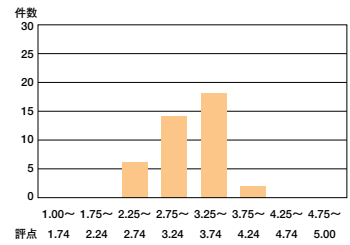
評価のタイミング（有用性）



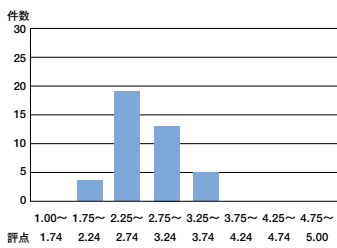
評価のタイミング（有用性）



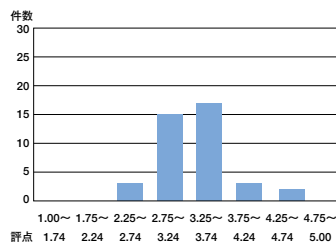
評価のタイミング（有用性）



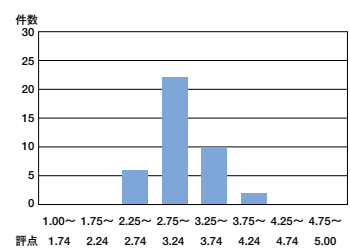
訪問先構成（公平・中立性）



訪問先構成（被援助国側）

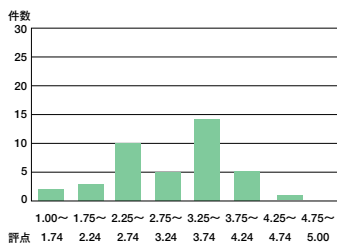


調査の実施状況（合計）

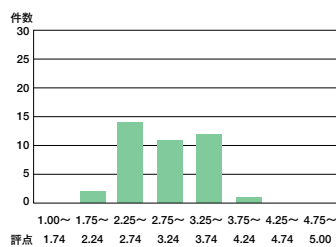


3 情報分析／評価

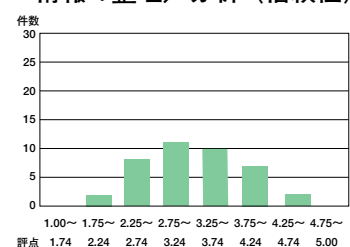
情報の整理／分析（有用性）



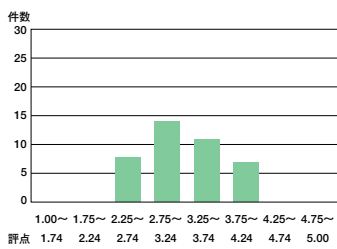
情報の整理／分析（公平・中立性）



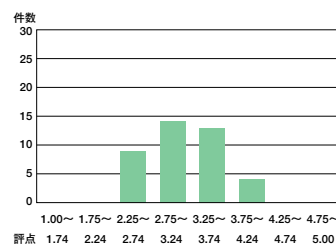
情報の整理／分析（信頼性）



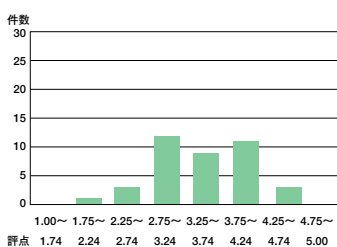
情報の整理／分析（被援助国側）



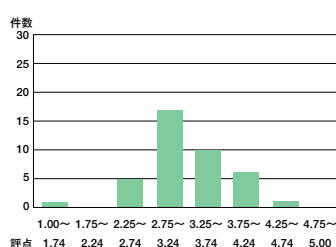
情報の整理／分析（合計）



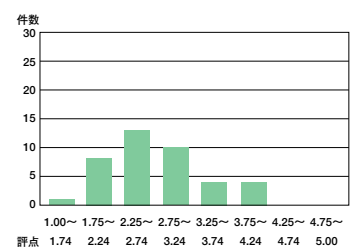
5項目評価・妥当性（有用性）



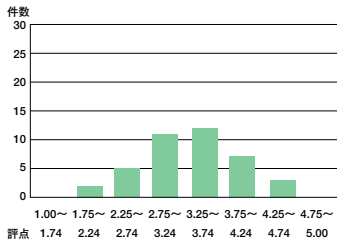
5項目評価・有効性（有用性）



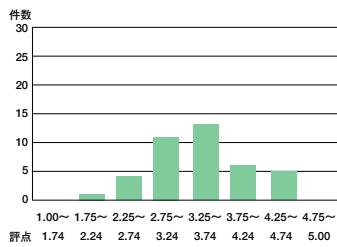
5項目評価・効率性（有用性）



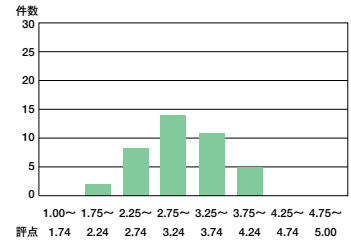
5項目評価・効果（有用性）



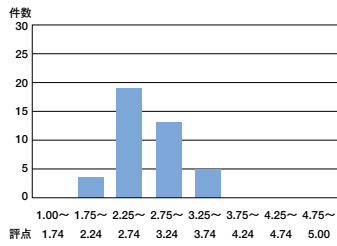
5項目評価・自立発展性（有用性）



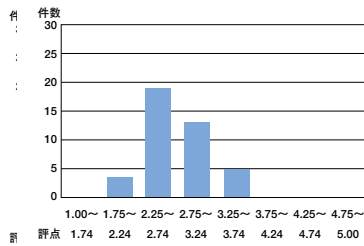
5項目評価の結果（合計）



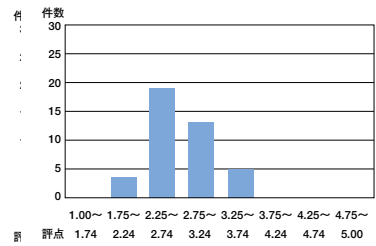
訪問先構成（公平・中立性）



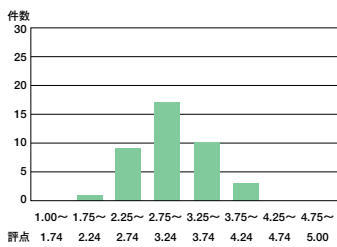
訪問先構成（公平・中立性）



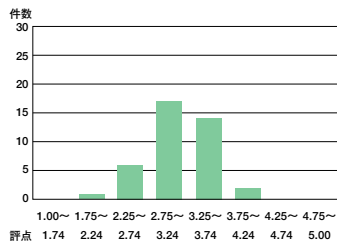
訪問先構成（公平・中立性）



提言（信頼性）

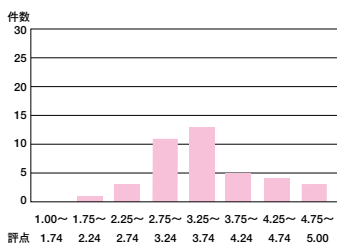


教訓／提言の内容（合計）

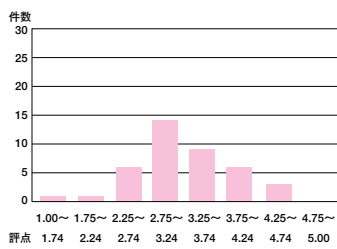


#### 4 報告書から読み取れるプロジェクトの評価

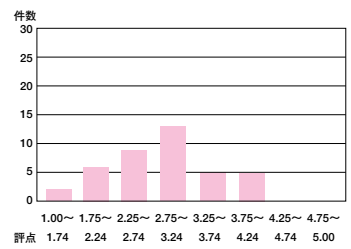
プロジェクトの妥当性



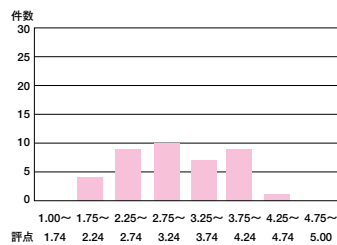
プロジェクトの有効性



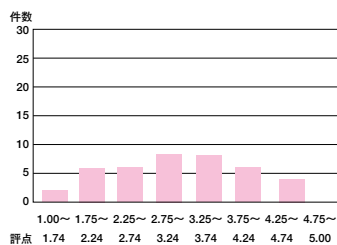
プロジェクトの効率性



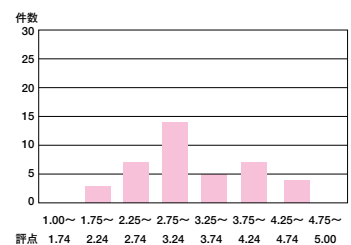
プロジェクトの効果



プロジェクトの自立発展性



プロジェクトの総合評点



ない例や派遣専門家など重要な情報源が抜けている場合もある。さらに、客観性の担保から、プロジェクト関係者以外からの意見収集も必要である。

プロジェクトによって裨益するのはパートナー国であることを考えれば、これらの評価はパートナー国との合同評価として行われるのが望ましいが、現実に多くの評価が合同評価として行われていることは評価できる。

### 5) 情報分析／評価

適切な情報の分析や評価がされているものが多い。「情報分析／評価」については効率性の有用性を除いては3点（普通）を超える値となっている。これは効率性という評価項目に関する理解の混乱によるものである。例えば、予算をきちんと使ったことを効率的と記した報告書もあるが、効率的とは資金を含めた資源を最大限利用することであり、「必要な機材は現地ですできるだけ安く調達する」場合などである。もし、現地調達できるよりも高価な機材を購入したのがあるとするれば、その理由（性能等）の説明がなければ効率性の判断はできない。効率性に関する分析の明確な基準を作り、評価者に周知させる必要がある。

図3-1に示すように、大きな評価項目のくくりとして、評価フレームワーク、実施状況については案件間のばらつきは少ない。他方、情報の整理／分析や評価5項目に関する評価項目に関しては、案件間のばらつきが大きく、質の高い評価結果もあれば低いものもあることが理解できる。

達成度を多段階評価で表すなど、定量的分析を心がけ、客観性を保つような努力がなされているが、一方で、評価の根拠が希薄なものや、示されていないものもある。また、一部ではあるが、統計的資料がない、あるいは古いなど、データの質が低いものもある。

研修や技術移転の効果は受講者数などの人数だけではなく、移転された知識や技術でなければならないが、そのような評価分析はあまりなく、研修や技術移転の実質的な評価をどのように行うかについ

て、検討が十分でない場合が多い。また、情報が十分分析されていない場合もある、さらに、社会的波及効果を分析する上では経済・政治・社会要因を十分考慮することが必要である。

さらに、一部ではあるが、評価が甘く、プロジェクト寄りの評価であると感じられるものもある。

### 6) 教訓、提言の内容

具体的・有益な教訓、提言がある一方、十分な提言・教訓が引き出されていない場合もある。提言の中には、評価結果からは導かれなかったり、記述内容が良すぎたり、さらには主観的、感情的であるものもある。また、具体的な提言ができるところを十分行っていない場合もある。教訓・提言としてあまり一般化しすぎれば、当たり前すぎることを言うことになる。だからといって、個別プロジェクトに特殊な教訓、提言であれば他のプロジェクトで利用できないことになる。

## (2) 報告書から読みとれるプロジェクトの2次評価

報告書から読みとれるプロジェクトの2次評価として、評価5項目による評価結果では、効率性を除いて3点（普通）を超える値となっており、特に妥当性は高い。効率性が低い理由は前項で述べたとおりである。報告書の質の評価と比較すれば評価点の標準偏差が大きく、プロジェクト間の差が大きい。特に、自立発展性に関しては差が大きい。

総合評価に関しては、大半は3点（普通）以上の評価が得られており、4点（かなり高い）以上も7件ある。一方、2点（やや低い）以下が2件あるなど、一部評価の低い案件もある。このように、非常に成功したプロジェクトもある一方で、少数ではあるが評価の低いプロジェクトも存在する。

プロジェクトの不成功の要因としては、外部条件の急変、プロジェクト運営上の不手際などもあるが、立案、企画が甘く、そもそも効果を期待できないものや、効果が何かの理解が十分でないプロジェクト

も見受けられる。最初の計画そのものがどうであったかということをも十分検証しなければならない。

問題のあるプロジェクトは、計画段階で目標が明確化されていない、または、プロジェクト・デザインに問題があるものもある。カウンターパートの選択、プロジェクト対象としての適切性、プロジェクトのタイトルと内容の不一致、パートナー国の主体性の弱さなど、そもそもプロジェクトを立案していく初期の過程で問題があったと思われる。

妥当性について、パートナー国の開発計画やわが国の国別援助計画の中にプロジェクトの目標が含まれていることだけで妥当であるという事例が多いが安易である。目標から見て、手段としてそのプロジェクトを採用することが適切かどうか、妥当性の基準になるのではないだろうか。手段として理屈が通らないプロジェクトは妥当性が低いということになる。また、その国全体の開発とセクター全域の開発プログラムをどのようにしていくかということ、日本だけでなく、各国で知恵を絞り、それに日本の開発がフィットしているかを明確にすることが重要である。

## 1-4 まとめと提言

以上の計量的分析及びコメント等の定性的分析結果を総合して、今後の終了時評価及び2次評価のあり方に関する提言をまとめれば以下ようになる。また、それに関連したプロジェクトの企画・運営の改善についても述べる。

### (1) 評価に耐えるプロジェクトの企画・運営

#### 1) 目標の明確化

終了時評価にあたって情報を分析するために必要なデータが不十分である報告書がみうけられた。これはプロジェクトの開始時にPDMの作成が十分吟味されていないことに帰する部分が多い。

PDMの作成は、終了時評価での検討手段として指標を用いることを念頭に置き、プロジェクトの計

画段階で目標設定や指標に用いるデータの入手手段を十分に検討することが重要である。そのため、PDMの作成にあたってはパートナー国の関係者にもPDMを十分に理解してもらい、指標となる統計資料が入手可能か否かをその時点で入念に確認することが必要である。

数値を用いた目標の設定は重要である。あらかじめ研修回数、受講者数、作成数、収穫・収入の増加率などに、可能なものについては具体的な数値目標を設定することで、目標に達成しているか否かの評価を容易にすることができる。

もちろん、どのような数値目標を設定するかは重要である。技術の向上を目標にしている場合、研修の参加者数という量的な結果だけを示して達成したとすることが多い。受講者がどの程度習得したか、彼らの習得程度で果たして実践に役に立つのか、という質の面を見ることが技術の向上を評価する場合必要である。研修前、研修後の受講者の技術レベルを調査し、数値による客観的指標を用いることが必要であろう。アンケートを用いる場合はイエス/ノーによる評価ではなく、5段階評価法などの評価が望ましい。

#### 2) 中間評価とモニタリングによるプロジェクト運営の改善

プロジェクトを実際に始めてからでないで得られない情報は数多くある。プロジェクト開始前に十分吟味したPDMであっても、プロジェクト開始後得られた情報に基づき、修正する必要があるのは当然である。しかし、PDMはプロジェクトの行動指針であるため、中間評価などプロジェクトの節目に公式の手続きに従って変更すべきである。終了時評価は直近の、通常は中間評価の際に改訂されたPDMに基づいて行うのが基本であろう。終了時評価のためにさらにPDMを作り替えた評価事例があるが、PDMが形式的に作られているだけで役立っていなかったことの証明でもある。

PDMは活用しなければならない。プロジェクト開始時に派遣専門家やカウンターパートなどプロジ

エクトの関係者にPDMを十分理解させておき、専門家やカウンターパートが各自の活動目的を明確に把握することによってPDMを自己評価に用いることができ、プロジェクトの効率性を高めることが可能となる。日頃のモニタリングもPDMに即して行い、そのつど必要な記録を残しておけば、事後評価作業は容易に行える。

## (2) 評価報告書の改善

### 1) 評価調査団員の構成

評価者を構成する際には専門性の高い者を選ぶ、また専門性に偏りがないように配慮する必要がある。さらに兼任することにより調査メンバーを少なくして調査経費のコストダウンを図ることを考えなければならない。

ここでいう高い専門性とは、当該分野における専門的知識のほか、開発問題に関する専門性、評価に関する専門性も含まれる。一人でこれらすべての専門性を備えた人材はなかなか得られないであろうが、評価調査団全体としてバランスがとれていることが重要である。

評価にはパートナー国側も当然参加するが、パートナー国の評価者に前もってPDMや評価手法を習得してもらうことにより、合同評価を効率的に行うことができる。

### 2) 情報収集の範囲と方法

目標の達成度を測るために、統計資料ばかりでなく幅広くデータを収集しなければならない。インパクトを図るためには直接的受益者（農民、受講者、患者など）からの情報の収集が必要であるが、インタビュー調査や質問紙調査では自由回答のほかに、理解度や考えに関する評価を行う工夫も必要である。なお、質問紙調査は評価調査団が行く前に実施し、調査時には補足調査にするのが効率的である。また、調査対象者数を多くする、対象者を派遣専門家、カウンターパート、受益者など幅広くして客観性を高めることが必要である。1次評価で十分な情

報が得られていなければ、それを基にした2次評価の十分な成果は期待できない。

### 3) 分析法

評価の判断基準を明確にするために、可能な限り数量的分析を行うことが望まれる。そのためにはプロジェクト開始時から目標を指標として明確に設定しておく必要がある。活動目的の達成度や5項目評価を文章で説明するとともに、5段階評価法等で評価することが考えられる。また阻害要因の分析を十分行わなければならない。

### 4) 報告書の記述法

読み手の理解が容易となるような書き方が必要である。また、実施経過を記述する際には、正負両面の経過を記述するなどの配慮が必要である。また、アンケートを実施した場合は質問項目及び結果を添付することが必須である。基本的に、PDMの目標・指標・活動目的にそって、結果や評価を記述するのが良い。調査結果全てを表にして記述しているのは、それぞれの関連性は分かりやすいが、字が小さい、図表が無いなどにより読みにくく、理解しにくいという欠点がある。重要なポイントは文章化すべきである。また、評価内容に関連する統計的資料や調査結果を図表にし、文中に提示して評価の判断理由を分かりやすく提示することが必要である。

## (3) 2次評価の意義

### 1) 評価のガイドブック

評価のために様々なガイドブック、マニュアルが作成されている。しかし、一般論を述べるだけでは具体性に欠け、なかなか理解ができない。ここで行ったような2次評価によって、質の高い評価報告書を客観的に示すことができれば、それらの報告書は一種の手本として利用できる。数年続ければ、分野や援助スキームごとに手本がそろい、その手法や内容を踏襲しながら調査報告書を書けば、質の担保も期待できる。

今回の40件の評価報告書の中で、報告書の質に関する27評価項目の合計得点が高かった評価報告書は「高生産性稲作技術研究計画（フィリピン）」、「中小企業鑄造技術向上プロジェクト（ブラジル）」などである。これらは、報告書から読みとれるプロジェクトの評価でも非常に高い評価を得ている。一方、「環境センタープロジェクト（チリ）」は報告書の質としては上記に劣らない高い評価がされているが、報告書から読み取れるプロジェクトの評価は平均より低い。一般的には、報告書の質の評価（特に、情報の整理分析、5項目評価の有用性）結果と報告書から読みとれるプロジェクトの評価結果との間に、いくらか強い相関が見られるが、この場合のように、プロジェクトそのものの評価は高くなくとも、報告書の質としては高いものもある。

## 2) 評価内容の不偏性

終了時評価は各事業部で行っている。内容を十分理解した部署による評価は重要であるが、それを外部の有識者が2次評価を行うことによって、評価の透明性を高めると共に、評価内容の客観化が可能であることを示した。このような活動を通して、終了時評価の質的向上を図ることができる。

しかし、有識者が行う外部評価についても同じ事がいえる。内部の人間ではなく、プロジェクトと利害が伴わない意味で評価結果は客観的であるといえるが、内容に信頼性があるかどうかは別の問題である。本稿で明らかにしたように、有識者にも評価傾向の偏りがそれぞれあるからである。本来であれば、複数の有識者に同一プロジェクトを評価してもらい、どの評価項目が一致し、どれが一致しないかを明らかにするのが望ましい。しかし、時間と費用の面で実用的ではない。だからといって、有識者が評価した結果は黙って受け取れというのは乱暴である。有識者の評価結果といえども、他の多数の有識者の目を通して、個人の評価傾向に偏らない、一般的知見を抽出する努力は必要であろう。

## 3) 評価結果のフィードバックメカニズムの構築

評価報告書が読まれない、使われないと言われる理由の一つは、評価した時点でプロジェクトは終了しており、直接そのプロジェクトの改善には役立たない事にある。そこで、将来の類似プロジェクトに知見を役立てることが期待されるが、それには教訓・提言の一般化が必要である。

多くの評価が行われ、教訓、提言が出されれば、それらが相互に矛盾する可能性は捨てきれない。そうであれば、個々の評価結果とその利用との中間に、一般化と権威付けの作業が必要である。さらに、そのようにして権威づけられた提言が実行される組織的な仕組みが重要である。提言されたものが実行されたかどうかのチェックも行わなければならない。従来、評価結果が生かされなかったのは、評価結果の内容、提示の仕方に問題があったと同時に、関係者の良識に頼るだけで、改善のための組織的な仕組みが十分ではなかったからである。

評価結果から導出された教訓・提言の実行を確認する仕組みを作り、提言を言いつばなしにしないことが重要である。外部有識者評価委員会がその一翼を担えればと願うものである。