# Assessment of Financial Savings From Peer Reviews of In-Progress Projects: A Case Study from the Department of Energy's Hydrogen Program

Prepared By:

Yaw O. Agyeman, TMS, Inc.

with

Jeff Dowd, DOE Office of Energy Efficiency & Renewable Energy

**Prepared for the Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy**

October 2008

**Acknowledgments:**

Special thanks go to the following people who provided inputs to the study or review comments on an earlier draft of this paper.

- EERE Hydrogen, Fuel Cells and Infrastructure Technologies Program – JoAnn Milliken, Nancy Garland, Sunita Satyapal, Antonio Ruiz, Roxanne Garland, and John Garbak.
- Office of Planning, Budget and Analysis – Darrell Beschen and Randy Steer.
- National Renewable Energy Laboratory – Neil Snyder and Bill Babiuch.

*"It is the policy of the Federal Government to spend taxpayer dollars effectively and more effectively each year. Agencies shall apply taxpayer resources efficiently in a manner that maximizes the effectiveness of Government programs in serving the American people."*
*[Executive Order: Improving Government Program Performance, November 13, 2007]*

## Executive Summary


The importance of peer review as a tool for improving the efficiency and effectiveness of programs is well established. However, the correlation between management efficiency improvements from peer reviews and direct financial savings has not been examined.

This study was undertaken to assess the financial savings to the Department of Energy's (DOE) Hydrogen Program due to decisions made from peer reviews. Data were gathered from the program's peer reviews and additional information solicited from program staff. The timeframe of observation was 2003–2007. The assessment of financial savings involved identifying competitively selected program-planned projects that were discontinued due at least in part to peer review ratings and comments. The total planned or proposed budgetary outlay for these projects was summed, and the amount spent until the date of discontinuation was subtracted to arrive at a savings estimate.

The analysis found that, in the period covered (2003–2007), the Hydrogen Program invested about $1.8 million in peer reviews. In return:

- The Program obtained expert validation that 95 percent of program-planned project funding was appropriately allocated to productive projects that supported the program's goals and planned targets. These competitively selected projects — for which the collective expert judgment was positive — were successfully completed, continued as is, or continued with some recommended modifications for scope refinement, realignment, or focus adjustment.

- The remaining five percent of competitively selected program-planned projects were discontinued, due in part to the ratings and comments projects received at the peer reviews.

- The Program avoided spending approximately $29 million in investments in projects not judged to be fully productive or aligned with program goals, and where funds could be reallocated for use in other activities with higher goal-achievement potential — a greater than 15-fold return for the investment in peer reviews.

Thus, for research & development (R&D) programs with long-term research goals, peer reviews not only improve overall management efficiency and effectiveness, but they can also greatly improve financial efficiency, specifically.

## 1. Purpose and Background

The purpose of this study is to identify and quantify financial savings from investments in evaluation activities in the Department of Energy (DOE) Hydrogen Program, as a case study in management efficiency as practiced in a federal agency. The focus is on investment "savings" (actually, planned investments subsequently averted) owed in large measure to recommendations from peer reviews.

Several policy directives require all programs within DOE, and throughout the entire federal government, to undertake evaluations to improve their efficiency and effectiveness. Examples include:

- The HEWD FY2007 appropriations Conference Report (H.Rpt. 110–185; p.59) directive, in which "The Committee directs the Department to quantify and track the progress and impact of the substantial investments in R&D."[1]
- The Office of Management and Budget's (OMB) Program Assessment Rating Tool (PART), which assesses the performance of programs on a set of criteria deemed relevant for program success.[2] This includes assessing each program's application and utilization of program evaluations.[3]
- Executive Order: Improving Government Program Performance, November 13, 2007, which requires federal agencies to implement clear annual and long-term goals, and have a means to measure progress toward achievement of goals through the efficient use of resources.

In addition, DOE's Office of Energy Efficiency and Renewable Energy (EERE) has its own requirement for Programs to undertake peer reviews.[4] Taken together, these directives set a clear expectation that programs will undertake evaluation activities to generate information that enable them to better manage and more efficiently achieve desired results.

For R&D Programs with medium- to long-term outcomes that may take several years to realize, management efficiency can be attained through the use of evaluation activities,

[1] HEWD is the House Subcommittee on Energy and Water Development Appropriations.

[2] PART includes assessment of "Program Purpose and Design," "Strategic Planning," "Program Management," and "Program Results/Accountability."

[3] For example, there are three evaluation-specific questions in the PART.

- *"Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?"*
- *"Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?"*
- *"Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?"*

[4] EERE Peer Review Guide, August 2004.

such as peer reviews. Peer review findings can be used to make critical project and program decisions, such as discontinue, continue as is, or continue with scope refinement, realignment, or even a shift in the focus of the research.

Such evaluation activities help programs become more efficient overall, as evidenced by the wide-ranging recommendations from peer reviews. However, there also appear to be financial "benefits" due to these evaluation activities, which have rarely been documented. In this particular framing, the financial benefit does not refer to funds brought in as a result of the evaluation activity, but rather to funds that might have been misallocated but for the investment in the evaluation activity. This avoided direct and opportunity cost represents tangible evidence of management efficiency, especially for R&D programs, where lines of research may be pursued over extended periods.

When using peer review to help make decisions to continue or discontinue projects, the Hydrogen Program balances numerical review ratings with written reviewer comments. The Hydrogen Program considers the opinions and recommendations of experts from across several evaluation platforms, including the annual Peer Reviews (hereafter referred to as "Peer Reviews"), the periodic government-industry Technical Team meetings, Technology Development Manager (TDM) on-site reviews, and the FreedomCAR and Fuel Partnership program-level reviews. These provide the Program with sufficient information from a variety of partners and stakeholders to enable well-informed management decisions.

## 2. Methods

DOE's Hydrogen Program has a well-established record of conducting peer reviews. This study covers the peer reviews of competitively selected program-planned projects[5] conducted from 2003–2007, inclusive.

There were three sources of data for the study. The primary sources were the Hydrogen Peer Review reports for 2003 through 2007.[6] These were complemented by information obtained from interviews with program staff. Specifically, the Program's TDMs provided information that was missing from the reports. For a few of the 2003 projects, the Principal Investigators provided some information for the analysis.

[5] The focus of the current paper is on program-planned projects. "Program-planned" projects, which are selected through a competitive process, are projects planned as part of the Multi-year Program Plan and Annual Operating Plan processes, in support of technical targets and goals. Over the 2003–2007 period, 811 reviews of projects were performed – 695 on program-planned projects and 116 on congressionally-directed projects. Congressionally-directed projects were excluded from the dataset used for this analysis, since decisions on congressionally-directed projects are out of the purview of the Program.

[6] 2007 Annual Merit Review, held May 15–18, 2007 in Arlington, VA.
2006 Annual Merit Review, held May 16–19, 2006 in Arlington, VA.
2005 Annual Merit Review, held May 23–25, 2005 in Arlington, VA.
2004 Annual Merit Review, held May 24–27, 2004 in Philadelphia, PA.
2003 Merit Review and Peer Evaluation, held May 19–22, 2003 in Berkeley, CA.

Project ratings and reviewer comments were examined and combined with information from the interviews. The Hydrogen Program uses a rating scale of 1–5 to rate projects, with the numerical increase signifying favorability. A project's quantitative score on a given peer review is rarely considered a sufficient basis, in isolation, for understanding the decision made with regards to that project. Rather, the decisions reached on projects are better understood in consideration of both qualitative critiques and the quantitative scores from the peer reviews. Funding data was collected for discontinued projects. This included funding for the fiscal year when the project was reviewed as well as planned or proposed out year funding. Project start and end date information was also collected. With a few exceptions, the decision made by the Program to continue or discontinue a project was obtained directly from the peer review reports. The details of the decision-making process were obtained from interviews with program staff.

For the analysis, the overall rating distribution was explored, followed by an assessment of ratings according to the decisions that the Program took to continue or discontinue a project. As a qualitative complement, the analysis also examined the full range of considerations that factor in the decisions the Program makes with regards to each project. The comments from the peer review reports, along with those from staff interviews, provided additional insight for understanding anomalous situations where some projects that rated relatively low on peer reviews were nonetheless continued.

Additionally, the analysis sought to confirm the likelihood that peer review ratings do influence the decision to continue or discontinue a project. For this purpose, three categories of peer review ratings were generated: low, moderate, and high, corresponding to a rating of less than 2.75, between 2.75 to 3.3, and greater than 3.3, respectively. The low category was chosen to include all projects whose peer review rating was 0.25 points below the lower bound of the 99.5 percent confidence interval around the mean rating. For symmetrical effect, the high category was chosen to include all projects whose peer review rating was 0.25 points above the higher bound of the 99.5 percent confidence interval around the mean rating. The moderate category encompasses the ratings between those two bounds.[7]

The calculation of peer review cost was based on actual data provided by the program for the five years under review. The financial savings calculations for discontinued projects involved two steps. First, the total budget for all the discontinued projects was summed. From this total, the amount spent for each project, up to and including the year of review, was subtracted. The difference represents the total amount "saved" (i.e., avoided continued investment) by the Hydrogen Program.

[7] The use of the 1–5 scale for rating projects means that distinctions occur in very small increments. The 99.5 percent confidence bound captured the majority of project ratings. The .25 expansion at both ends of the confidence interval was based on a judgment of the distribution of scores, and what appeared to be a natural cut-off point below which a project could be considered to be underperforming relative to other projects.

## 3. Findings

### 3.1 Financial Cost of the Peer Reviews

The cost to conduct the peer reviews over the 5 years (2003–2007) was about $1.8 million (Table 1), for an average cost of approximately $357,400 per year, or slightly over $2,200 per project reviewed. This figure represents the DOE Hydrogen Program's overall cost. EERE covered the vast majority of the cost, with small contributions from Fossil Energy (FE), Nuclear Energy (NE), and Basic Energy Sciences (BES) — the other three offices that support the President's Hydrogen Fuel Initiative. Beginning in 2004, all four DOE offices provided overview presentations during the opening plenary session of the Reviews, but only EERE, FE (since 2005) and NE (since 2005) have presented projects that are peer reviewed. The BES's presentations and posters have not been subject to reviews.

The increase in overall cost over the years is at least in part due to the increase in number of reviewed projects and rise in attendance each year. The lower cost in 2003 and 2004 was partly because registration fees were charged in those two years.

**Table 1. Cost of Peer Reviews, DOE Hydrogen Program, 2003-2007**

| Year of Review | Number of Reviewed projects | Peer Review Cost |
|---|---|---|
| 2003 | 129 | $250,000 |
| 2004 | 164 | $272,000 |
| 2005 | 191 | $350,000 |
| 2006 | 167 | $400,000 |
| 2007 | 160 | $515,000 |
| Total | 811 | $1,787,000 |

### 3.2 Financial Savings from Peer Reviews

The overall cumulative budget for the Hydrogen Program over the 2003–2007 period of this study was approximately $608 million (excluding congressionally-directed projects).

Approximately eight percent of this total, or $56 million, was the total budget outlay for all the program-planned projects that were discontinued (Table 2). These are presented in the "Total Funding" column, reflecting the funding that was planned or proposed for each project at inception. Some projects, for several technical and management reasons, are identified from inception as requiring annual determination of funding, and are thus funded for that year alone. By definition, if such a project is discontinued after a review, there can be no claim of avoided investment because no funds would have been planned

**Table 2. Financial Savings from Discontinued Projects, DOE Hydrogen Program, 2003–2007.**

| Project ID | Year | Sub-Program | Project Name | Project Rating | Total Funding | Funds Spent | Funds Saved |
|---|---|---|---|---|---|---|---|
| 120 | 2003 | Fuel Cells | PEMFC power system on ethanol | 2.8 | $882K | $632K | $250K |
| 123 | 2003 | Fuel Cells | Testing of fuel cell reformers | 3.04 | $250K | $250K | $0K |
| 133 | 2003 | Fuel Cells | DOE compressor/expander module development program | 1.8 | $930K | $500K | $430K |
| 87 | 2003 | Fuel Cells | Carbon foam for fuel cell humidification | 1.93 | $100K | $100K | $0K |
| 89 | 2003 | Fuel Cells | Sulfur removal from reformate | 2.7 | $200K | $200K | $0K |
| 99 | 2003 | Fuel Cells | Diesel reforming | 2.33 | $100K | $100K | $0K |
| FC-P4 | 2004 | Fuel Cells | Microchannel reformate cleanup: water gas shift & preferential oxidation | 2.72 | $2100K | $700K | $1400K |
| FC-P5 | 2004 | Fuel Cells | Effects of fuel composition on fuel processing | 2.56 | $600K | $300K | $300K |
| FC-P24 | 2004 | Fuel Cells | Graphite-based thermal management | 2.96 | $398K | $129K | $269K |
| FC-P25 | 2004 | Fuel Cells | CO sensors for fuel cell applications | 3.1 | $400K | $200K | $200K |
| FC-24 | 2005 | Fuel Cells | Water gas shift catalysis | 2.65 | $7200K | $4200K | $3000K |
| FC-25 | 2005 | Fuel Cells | Catalysts for autothermal reforming | 2.68 | $6000K | $4000K | $2000K |
| FC-31 | 2005 | Fuel Cells | DOE Hydrogen Program sensor dev. | 1.91 | $1521K | $831K | $690K |
| FC-34 | 2005 | Fuel Cells | Direct methanol fuel cells | 3.23 | $650K | $650K | $0K |
| FC-36 | 2005 | Fuel Cells | Bipolar plate-supported solid oxide fuel cell “tuffcell” | 2.63 | $1600K | $800K | $800K |
| FCP-11 | 2005 | Fuel Cells | Modeling and control of an SOFC APU | 3.1 | $1666K | $1000K | $666K |
| FCP-15 | 2005 | Fuel Cells | Plate-based fuel processing system | 2.73 | $8160K | $6999K | $1161K |
| FCP-19 | 2005 | Fuel Cells | Fuel cells vehicle systems analysis | 2.54 | $1020K | $204K | $816K |
| FCP-26 | 2005 | Fuel Cells | Fore court fuel processing | 3.03 | $1200K | $400K | $800K |
| FC-06 | 2007 | Fuel Cells | Development of transition metal/ chalcogen based cathode catalysts for PEM fuel cells | 2.77 | $1580K | $1033K | $547K |
| 48 | 2003 | Storage | Low permeation liner for H2 gas storage tanks | 2.75 | $50K | $50K | $0K |
| 52 | 2003 | Storage | H2 storage using complex hydrides | 2.2 | $750K | $248K | $502K |
| ST-4 | 2004 | Storage | Radiolysis process for the regeneration of sodium borate to sodium borohydride | 2.32 | $5000K | $50K | $4950K |
| STP-02 | 2007 | Storage | Conducting polymers as new materials for H2 storage | 2.62 | $664K | $269K | $395K |
| 5 | 2003 | Prod & Del | Biological water gas shift | 3.23 | $750K | $750K | $0K |
| 13 | 2003 | Prod & Del | Reformer model dev. For H2 production | 2.27 | $1599K | $150K | $1449K |
| 31 | 2003 | Prod & Del | Supercritical water partial oxidation | 2.57 | $7200K | $313K | $6887K |
| 38 | 2003 | Prod & Del | Low cost, high efficiency reversible FC systems | 2.8 | $819K | $614K | $204K |
| 39 | 2003 | Prod & Del | High-efficiency steam electrolyzer | 2.37 | $873K | $873K | $0K |
| PDP-39 | 2005 | Prod & Del | H2 production – increasing the efficiency of water electrolysis | 2.41 | $803K | $241K | $562K |
| PDP-31 | 2007 | Prod & Del | Corrosion studies of metallic materials for thermochemical cycles | 3.39 | $989K | $630K | $359 |
| | | | | TOTALS | $56055K | $27418K | $28637K |

*There were no reviewed program-planned projects discontinued for the Safety, Codes & Standards sub-program, or for the Technology Validation sub-program. Figures are rounded to the nearest $1,000.

for that project past the year of review. For most of the projects, however, the expected period of performance is longer than a year, and funds are appropriated accordingly. The "Total Funding" column is the sum of all the funds budgeted for the reviewed projects that were discontinued.

Of this amount, approximately $27 million was spent, up to and including the year of the review. Those figures are presented in the "Funds Spent" column (Table 2). They reflect the funds spent on all the projects reviewed, including the budget for the year of review, since discontinued projects were allowed to run their course to the end of the year of review.

The difference between the $56 million total budget for all discontinued projects, and the $27 million that was spent, results in approximately $29 million of the total budget (that was planned or proposed for investment) being saved due to the influence of the peer reviews. Thus, for an investment of approximately $2 million for peer reviews, the Hydrogen Program avoided a direct and opportunity cost of approximately $29 million.

For context, this $29 million in averted (saved) investments is roughly equivalent to the total FY2007 budget appropriations of $28 million for the Hydrogen Program's research in Transportation Fuel Cell Systems, Distributed Energy Fuel Cell Systems, and Safety Codes and Standards.

### 3.3 A Closer Examination of the Hydrogen Program's Decision Making

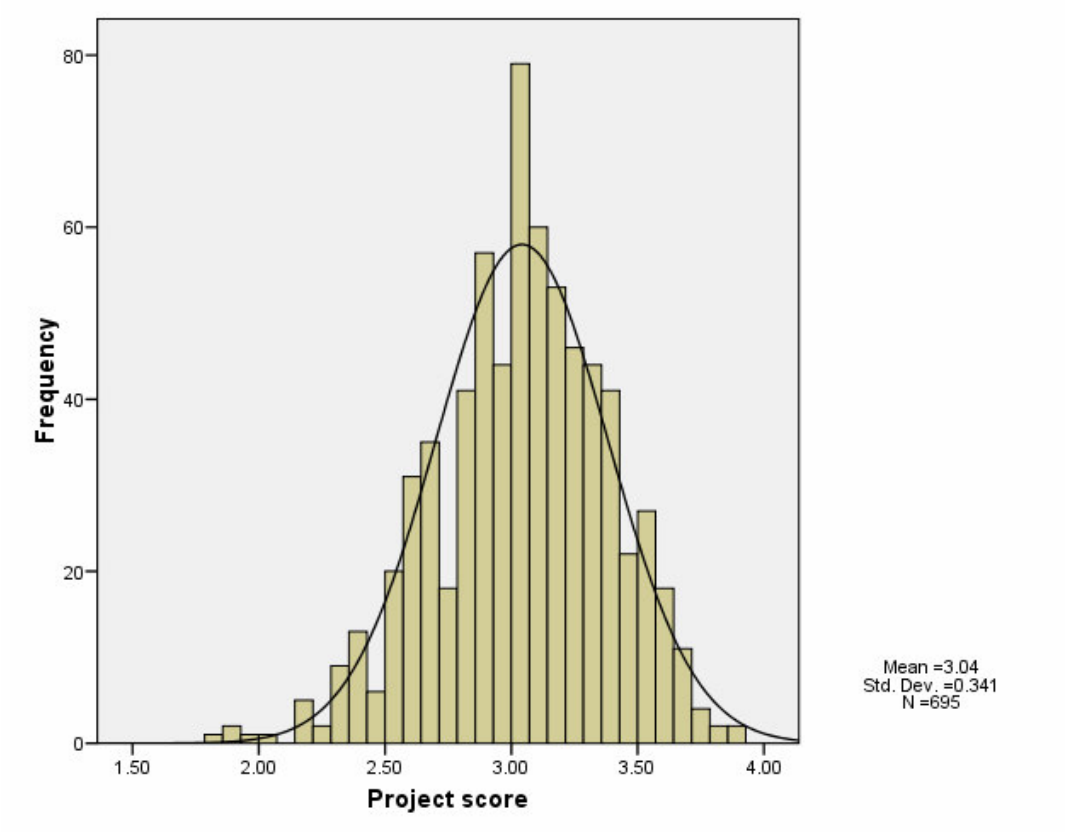
Six hundred and ninety-five (86 percent) of the total number of reviews (811) from 2003–2007 were reviews of program-planned projects.[8] Figure 1 presents the distribution of peer review ratings for the program-planned projects.

The mean rating on the 1–5 scale was 3.04, with a minimum of 1.8, a maximum of 3.92, and a standard deviation of 0.34.

Decisions taken to continue or discontinue projects were clearly related to review ratings. Table 3 presents the average project ratings according to the decision taken. Compared to discontinued projects, both completed projects and continued projects had higher peer review ratings.

[8] The remaining 116 reviews were for congressionally directed projects.

Figure 1. Distribution of project scores from peer reviews, DOE Hydrogen Program, 2003-2007.




A statistical comparison of the pairwise difference in means between the categories showed that the difference in average rating between continued versus discontinued projects is statistically significant. Likewise, the pairwise difference in average rating between continued projects and completed projects is statistically significant.

**Table 3. Project Ratings According to Decision Taken, for Program-planned Projects, DOE Hydrogen Program, 2003–2007.**

| Decision | Mean Peer Review Rating | Number of Projects | Std. Deviation |
|---|---|---|---|
| Continued | 3.08 | 570 (82%) | .32 |
| Discontinued | 2.66 | 34 (5%) | .38 |
| Completed | 2.97 | 91 (13%) | .35 |
| Total | 3.04 | 695 (100%) | .34 |

Note: Pairwise Significance - Continued vs. Discontinued: p<.000; Continued vs. Completed: p<.05; and Discontinued vs. Completed: p<.000.

Table 4 and Figure 2 present additional evidence in support of the influence of peer review rating on decisions taken to discontinue projects. The projects are presented in terms of the designated categories of low (less than 2.75), moderate (between 2.75 and 3.3), and relatively high (greater than 3.3) ratings, as described in Section 2. A total of 133 (19 percent) of the reviews of projects were rated in the relatively low category; 417

(60 percent) were rated in the moderate category, and 145 (21 percent) were rated in the relatively high category.

A significantly higher proportion (14 percent) of low-rated projects was discontinued, compared with 3 percent of the moderate-rated projects, and only 1 percent of the relatively high-rated projects.

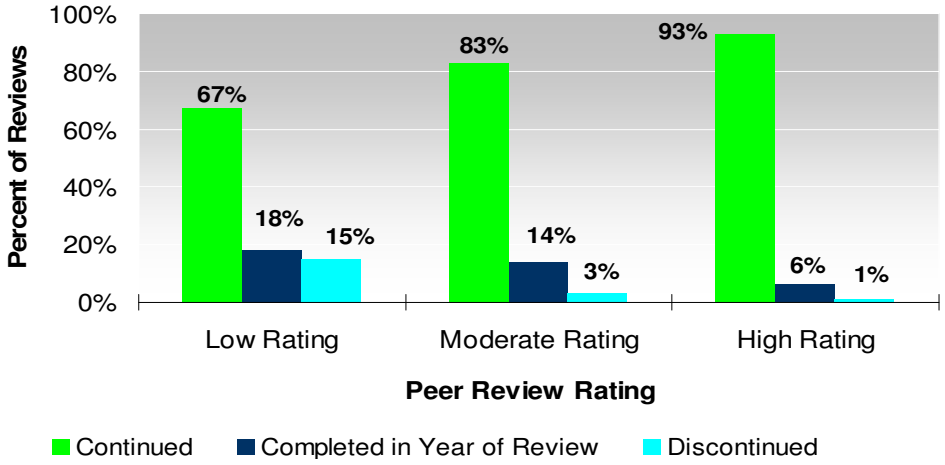
**Table 4. Decision Taken According to Project Rating, DOE Hydrogen Program, 2003-2007.**

| | Decision | | | |
|---|---|---|---|---|
| **Project rating** | **Continued** | **Discontinued** | **Completed** | **Total** |
| Low (Less than 2.75) | 89 (67%) | 20 (15%) | 24 (18%) | 133 (Overall percent = 19%) |
| Moderate (Between 2.75 and 3.3) | 346 (83%) | 13 (3%) | 58 (14%) | 417 (Overall percent = 60%) |
| High (Greater than 3.3) | 135 (93%) | 1 (1%) | 9 (6%) | 145 (Overall percent = 21%) |
| **Total** | 570 (82%) | 33 (5%) | 92 (13%) | 695 (100%) |

Note: Significance – Pairwise: Low vs. Moderate, p<.004; Pairwise: Low vs. High, p<.000; and Pairwise: Moderate vs. High, p<.01.

Conversely, 93 percent of high-rated projects were continued, 83 percent of the moderate-rated projects were continued, and 67 percent of low-rated projects were continued.

**Figure 2. Decisions Taken According to Review Ratings, DOE Hydrogen Program, 2003-2007.**



### 3.4 Examination of the Qualitative Complement to the Quantitative Ratings[9]

As noted in Table 4, 67 percent of low-rated projects were continued. To further understand the decision making process, an examination of qualitative information was undertaken. Primarily, this examination was motivated by the desire to understand why some low-rated projects were nevertheless allowed to continue, as well as why some moderate- and high-rated projects were discontinued.

*Reasons Why Some Low-Rated Projects Were Continued*

In this category, there was particular interest in projects scoring below 2, as well as, more broadly, projects scoring below the mid-point of the 1-5 rating scale (that is, below 2.5). Three projects were identified that rated below 2 on the peer review. All three of them were discontinued.

Looking a bit more broadly at projects that rated above 2, but below the mid-point of the scale (2.5), 15 were identified, of which 4 were completed in the year of review. Of the remaining 11 projects scoring below 2.5, 4 were discontinued, and 7 were continued. One example of a continued project is Project #109, Fuel Cells, 2003, "OnBoard Vehicle, Cost-Effective Hydrogen Enhancement Technology for Transportation Polymer Electrolyte Membrane Fuel Cells (PEMFCs)." This project was rated as follows on the criteria used:[10]

- Relevance to overall DOE objectives – 2.33
- Approach to performing the R&D – 2.33
- Technical accomplishments and progress – 2.33
- Technology transfer and collaboration – 1.33
- Approach to and relevance of future proposed research – 2.33

The overall score for the project was 2.13, with the criteria of "Technology transfer and collaboration" proving a particularly weak point. The reviewers rated the project strong on innovative potential, but weak on other aspects, such as an absence of clarity on projected cost, and absence of a timeline, as well as specific technical weaknesses including the inability to prove CO tolerance and the absence of thermodynamic analysis. The decision appears to have been to give the project at least one additional year to address its structural weaknesses.

For this example, and for the other continuing projects scoring below 2.5, additional reasons given for continuing are listed below:

[9] Detailed explanations are provided in the Program's Peer Review Reports. For this paper, program staff provided additional information during interviews.

[10] 2003 Merit Review and Peer Evaluation, held May 19–22, 2003 in Berkeley, CA.

- The project had a late start, or was new. The Program recognized a need to give the R&D an appropriate timeframe to succeed.
- Some projects were allowed to continue with very specific recommendations on the focus desired and the modifications needed to address identified weaknesses in one or more of the review criteria areas.
- In a few instances, despite the rating, the Program decided that there were additional benefits to allowing the project to conclude; that is to say, the research was deemed critical to achievement of program goals.

*Reasons Why Some Moderate-Rated Projects Were Discontinued*

Three percent of moderate-rated projects were discontinued. For the projects whose ratings were in the moderate range (where the majority of the projects were clustered), there were considerable nuances behind the decision to discontinue. For example:

- Although these projects scored well on some criteria, some were discontinued because they were unable to meet their technical targets, or a determination was made that their approaches would not meet cost targets, or succeed in overcoming critical technical barriers.
- Some were discontinued due to a decision to redirect the research based on a re-prioritization of program focus.
- Some projects were discontinued because better approaches had been found (for example more cost-effective methods for doing the same thing).
- Some national laboratory projects were discontinued during the switch from informal to formal competition of lab R&D.
- In some instances, a programmatic decision was made to discontinue good work because of program budget reductions, due to shortfalls in appropriations or to congressionally-directed projects.

## 4. Conclusion

The primary purpose of this paper is to document the financial savings to the Hydrogen Program due in considerable part to decisions made as a result of peer review ratings and comments. Peer reviews are well established as effective tools for improving management efficiency and effectiveness, but the analysis presented in this paper takes a rare step to identify their potential for providing financial savings.

The findings suggest considerable financial savings to the Hydrogen Program for its investment in peer reviews. The current paper sought simply to identify the savings, without specifically seeking to examine the uses to which the savings were diverted. For the Hydrogen Program -- and other R&D programs, in general -- the argument for sustained rigor in conducting peer reviews would appear evident. This paper demonstrates the financial benefit of peer reviews, and their usefulness as a management tool for ensuring Government program resources are spent efficiently and more effectively.