

UNNATURAL SELECTION:
AN EXAMINATION OF THE FRASER
INSTITUTE'S PERFORMANCE INDEX FOR
THE DONNER CANADIAN FOUNDATION
AWARDS FOR EXCELLENCE IN THE
DELIVERY OF SOCIAL SERVICES

Jerry DeMarco
Toronto, Ontario

Abstract: Using Darwinian metaphors, this article examines a relatively new quantitative performance evaluation tool for Canadian non-profits: the Fraser Institute's Performance Index for the Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services. Several aspects of the Index, including strategic management, outcome monitoring, and income independence, are analyzed in light of their ability to accurately evaluate excellence in performance and their potential unsuitability for many non-profits. A theme throughout this discussion is the interaction of the "business-like" approaches favoured by the Index and the fundamentally different nature of non-profit perspectives. Particular attention is paid to the difference between organizations that share a profit motive and non-profits that have diverse organization-specific missions. Context-specific mission-driven approaches to non-profit performance evaluation are recommended.

Résumé: Utilisant des métaphores darwiniennes, cet article examine un outil quantitatif d'évaluation du rendement des organismes à but non lucratif canadiens relativement nouveau: le Performance Index du Fraser Institute pour les prix d'excellence de la fondation canadienne Donner (Donner Canadian Foundation Awards) pour la prestation de services sociaux. Plusieurs aspects de l'Index, dont la gestion stratégique, le suivi des résultats et l'autonomie financière sont analysés afin de déterminer leur capacité à évaluer avec exactitude l'excellence du rendement et leur inadéquation potentielle pour de nombreux organismes à but non lucratif. L'interaction entre les approches «commerciales» préconisées par l'Index et la nature fondamentalement différente des objectifs des organismes à but non lucratif constitue un thème récurrent de la discussion. On accorde

Corresponding author: Jerry DeMarco, Sierra Legal Defence Fund,
30 St. Patrick Street, Suite 900, Toronto, ON, M5T 3A3; <bear@sierralegal.org>
(The views expressed herein are solely those of the author.)

une attention particulière à la différence entre les organismes qui ont un souci de rentabilité et les organismes à but non lucratif qui ont des missions spécifiques diverses. Pour évaluer le rendement des organismes à but non lucratif, on recommande des approches axées sur un contexte et une mission spécifiques.

“What gets measured gets done.”

David Osborne & Ted Gaebler (1992, p. 146)

“In every move, in every decision, in every policy,
the non-profit institution needs to start out by asking,
Will this advance our capacity to carry out our mission?”

Peter F. Drucker (1990, pp. 87–88)

“Change is not necessarily progress. Change is adaptation.”

Michael Quinn Patton (1997, p. 106)

Non-profit performance evaluation and Darwinism are two quite unrelated topics, it is clear, but I will attempt to weave them together to help convey my concerns about a relatively new Canadian quantitative non-profit evaluation tool. To begin, I wish to set out the general parameters of the discussion to follow. First, I recognize that non-profits are without the benefit of simple quantitative evaluation tools that could be used to assess their overall performance. Second, I believe that, despite the many complexities within business models, the common evaluative denominator remains focused on quantifiable profit, which makes those models not wholly transferable to the non-profit arena. Third, I believe that the lack of a simple quantitative evaluative tool for non-profits inspired the creation of the Fraser Institute’s Performance Index for the Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services (the “Donner Awards”), and that as a new entrant to the field of quantitative evaluation tools for non-profits, the Index holds a “first-on-the-scene” advantage. Fourth, I believe that there are inherent difficulties in attempting to create a one-size-fits-all evaluation tool for non-profits and that these difficulties mitigate in favour of considering a range of utilization-focused evaluation tools (Patton, 1997) in a context-specific and mission-driven manner.

THE IMPORTANCE OF PERFORMANCE EVALUATION

Critical to any organization is the question of how well it is performing. As noted by Patton (1997, p. 15), “[c]orporations, philanthropic foundations, and nonprofit agencies are increasingly turning

to evaluators for help in enhancing their organizational effectiveness.” Evaluation is normally intended to serve one or more of the following purposes: “rendering judgments, facilitating improvements, and/or generating knowledge” (Patton, 1997, p. 65). While the choice of which measures and factors to use in such evaluation exercises is open to debate, it is evident that a non-profit organization should strive to perform to the best of its abilities. Critical to such performance is knowing how well an organization is doing vis-à-vis its mission, and how it can improve itself. How one goes about this process of evaluation in the non-profit arena is not the subject of any consensus and many different approaches to evaluation are available (see Hall, Phillips, Meillat, & Pickering, 2003, p. 38; Patton, 1997, pp. 387–413; Wood & Clemens, 2002, pp. 80–84). In the for-profit world, there are at least some common denominators revolving around measures of profit that can be used to compare performance across organizations, though it is recognized that the for-profit arena is not simple to evaluate.

In the non-profit world, no obvious common denominator provides even the basic measuring stick that profitability measures offer to for-profit enterprises. But this should not be used to excuse an organization from evaluating performance. While non-profit performance evaluation may be difficult, it still needs to be done if we are to understand whether a non-profit is “doing well while doing good” (Kanter & Summers, 1987, p. 154).

THE PERFORMANCE INDEX AND THE DONNER AWARDS

The Donner Awards are a collaborative initiative of the Donner Canadian Foundation and the Fraser Institute. The Awards are based on a two-stage process. First, non-profit organizations in nine different areas of social services (i.e., alternative education, child care, counselling, crisis intervention, prevention and treatment of substance abuse, provision of basic necessities, services for people with disabilities, services to seniors, and traditional education) complete a quantitative-based questionnaire regarding ten areas of performance management (i.e., financial management, income independence, strategic management, board governance, volunteers, staff, innovation, program cost, outcome monitoring, and accessibility) (Wood & Clemens, 2002, pp. 5–6), which are then analyzed by the Institute according to a Performance Index (see Appendix 1). Second, based on the results of this quantitative exercise, three or four finalists in each category of social service are selected to respond in essay-style

format to five questions dealing with strategic management, outcome monitoring, program delivery, innovation, and affirmation (Wood & Clemens, 2002, p. 7). Based on these answers as well as independent letters of support, \$5,000 awards for excellence are issued in each of the nine categories of social services and a single \$20,000 award is issued to the overall winner. Accompanying the awards is an annual performance report (Wood & Clemens, 2002) that provides aggregate results from all applicants and graphs of how each class of organization scored in each area of performance measurement.

From the inception of the Donner Awards in 1998 to the present, between 233 and 299 applications have been received annually from non-profit organizations (Wood & Clemens, 2002, p. 79). Though the awards are the key public outcome to the process, the performance evaluation also results in organization-specific "Confidential Performance Reports," which each applicant receives "to identify particular areas of weakness" and "to discover ways to improve performance" (Wood & Clemens, 2002, p. 14). Thus, this evaluative process appears to focus on two of the typical general purposes of evaluation set out by Patton (1997, p. 65), namely, rendering judgement and facilitating improvement.

The Performance Index is relatively uncommon in its attempt to provide a quantitative performance evaluation system to measure the excellence of social service programs in many very different areas of endeavour. Wood and Clemens (2002, p. 9) consider their approach to be a "comprehensive" system that is a "major step forward in the ability of non-profit organization managers to assess performance on an objective and quantifiable basis." Considering that it seeks to provide a comprehensive approach to assessing the excellence with which social services are delivered in nine completely different areas, the Performance Index system is quite simple. The application form is a self-assessment questionnaire that is four pages in length and contains just over 50 questions (Appendix 2). A four-page guide for applicants is also provided to assist in completing the questionnaire (Fraser Institute, 2003d). The literature associated with the Donner Awards refers to the general concepts of performance, effectiveness, excellence, and efficiency as the matters being evaluated by the Performance Index.

Given that a key aspect of the Performance Index is to provide non-profit organizations with information that they can use to improve

their performance, it makes sense to first examine whether the guidance provided by the Index is well suited to the organization and the non-profit sector. This is a key aspect of “utilization-focused evaluation” (Patton, 1997) and “practical program evaluation” (Newcomer, Hatry, & Wholey, 1994, p. 2). Consistent with the notion of “evaluability assessment” (Wholey, 1994, p. 17), one must be cognizant of how an evaluation will be used and whether it will be useful and practical for the organization or other stakeholders. This involves examining whether the tool is well suited to the particular milieu and, if it is not, whether acting on the results may provide little added value or even detract from the organization’s mission (Hall et al., 2003, p. 2).

To help examine the potential impact of an evaluation tool on an organization’s evolution, I will use a Darwinian metaphor. Will the Performance Index cause non-profits to *adapt* to the wrong *environment*? Will participating organizations *select* characteristics that help them better accomplish their mission or better emulate business, or both? Does the process assist the organizations in becoming more *fit* for their mission-driven *environment*, or more *fit* for another type of *environment* (e.g., a business-like world)?

It is worth noting that the Fraser Institute’s program director, who oversees the Donner Awards, invokes a Darwinian metaphor in his call for the non-profit sector to become more “business-like.” With reference to the current debate surrounding this alleged need, he remarks:

Not all non-profit organizations embrace change. Some continue to believe that good words speak for themselves. They believe the language and practice of business have no place in the world of volunteers and compassion. Others go further, blaming business for creating a Darwinian world of winners and losers — and then turning its back on the losers. (Clemens, 2000, p. B11)

Because the use of the Performance Index is in many ways a specific example of the larger debate about becoming more “business-like,” it is particularly relevant (see Drucker, 1977, p. 19). The Index exemplifies the collection of traits generally reflective of business approaches, which some advocate should be used in the non-profit arena. Whether one views the increased emphasis on business-like approaches as unwanted colonization or useful cross-pollination, the

phenomenon itself is very present in the current non-profit environment. Legitimate calls for more accountability, better disclosure of outcomes, increased efficiency, and so on can all lead a non-profit organization to look to the business world as a source of guidance.

ON THE USE OF A DARWINIAN METAPHOR

I recognize that there are some problems with choosing a Darwinian metaphor and that it is not a perfect fit for all aspects of evaluation. Nevertheless, in choosing this metaphor, I was influenced by my own predisposition for learning by analogy and inspired in part by Patton's (1997, p. 37) comment: "As we look for new metaphors in evaluation, we would do well to do so in the spirit of Thoreau, who observed, 'All perception of truth is the detection of an analogy.'" Here, as others have done (Aldrich, 1979, pp. 26–55; Weick, 1977, pp. 210–211), I will use the standard hallmarks of Darwinism (evolution, natural selection, adaptation, survival of the fittest, competition, etc.) as a metaphorical lexicon — as a broad analogy.

In so doing, one obvious danger is to misuse Darwinism, as others including social Darwinists have done, so as to give some pretense of scientific validity to what is simply a point of view like any other. In light of that caveat, my purpose here is not to apply Darwinism holus-bolus to the chosen subject but simply to borrow some of the language of Darwinism in an effort to find a colourful lens through which one can view the subject.

Natural and Artificial Selection

The centrepiece of Darwin's theory of evolution is the notion of natural selection: "the fittest survive and spread their favored traits through populations" (Gould, 1977, p. 40). How this process of selection is carried out is succinctly described by Livingston (1994, p. 205):

Darwinian natural selection is about the environmental constraints or opportunities that cause extinctions and allow the emergence of new species. If a species "fits" its environment it will survive. If it does not, it will disappear. Those that survive have been selected in; those that do not have been selected out.

I realize that organizational change is not the product of randomness in the same way as a species' evolution is affected by random factors and that organizational change, in many ways, is consequently closer to the notion of artificial selection. In natural selection, randomness and inherent genetic variability play crucial roles in affecting the eventual fitness of a species (like an emergent outcome). On the other hand, the fitness of an organization can be managed and planned in an active manner (like an intended outcome) (Zimmerman & Dart, 1998, p. 13). Organizations are actively changing themselves ("artificially" selecting certain characteristics and discontinuing others) and so are evolving, but not by the same natural mechanisms at play in the evolution of species. In utilizing Darwin's concepts in this essay, I will borrow ideas from both natural selection and artificial selection.

Darwin used pigeon breeding as an obvious example of artificial selection, in which humans intentionally select certain traits in pigeons, whether for appearance, flight speed, or so on. Is there an analogy to this type of selection in the present discussion of non-profit evaluation? Possibly. Breeders mould organisms according to their own image of what is desirable. If an evaluator believes that certain characteristics are synonymous with a particular desired state of affairs for an organization, then the evaluation and its recommendations will tend to influence the organization in that direction. In the context of evaluation specifically, Patton (1997, p. 179) warns of the "distortions that result when program staff pay too much attention to what an evaluator decides to measure, essentially giving the evaluator the power to determine what activities become primary in a program."

The analogy is not perfectly suited to organizations, however, because in artificial selection, the outside agent (the breeder) controls much of the system and the subject is an involuntary one. In evaluations such as those that rely on the Performance Index, the organization is voluntarily choosing to participate and itself controls whether it will choose to change its characteristics based on the results of the evaluation. To make this self-determination more meaningful, the organization should examine what the influence of the outside agent (the Performance Index) is, what assumptions it has made, and whether the agent's picture of a desired future state for the organization is one with which the organization agrees. If the organization goes into the process without such a critical eye, it will be at risk of being "bred" into a form that it may not find, in the end, to be desirable at all.

The External Environment and the Survival of the Fittest

Along with random factors, the importance of the external environment on evolution was key to Darwin's theory (Gould, 1977, p. 33). The organism that is a better fit with its environment survives according to Herbert Spencer's often quoted phrase "survival of the fittest." The notion of fitness is key to understanding Darwin's theory. Fitness is relative to the current suite of environmental influences and is not judged against any other standard of desirability. But what about fitness in the Performance Index model? Under the model, survival of the fittest non-profit seems to be a "good" thing — something to strive for. Under the model, the organization that retains revenues, eschews unstable funders, and perpetuates itself into the future will score higher. Are such measures, however, appropriate indicators of fitness in all situations? I believe that the suitable definition of fitness for an organization may be so context-specific that no overarching measurable metric of fitness is possible. It may be that fitness can only be defined relative to both the external environment and the mission of the organization itself. For those reasons, one should look at recipe-style evaluation tools with circumspection.

Competition

Darwin's theory of evolution was heavily influenced by the concept of competition (Darwin, 1859, p. 323):

The theory of natural selection is grounded on the belief that each new variety, and ultimately each new species, is produced and maintained by having some advantage over those with which it comes into competition; and the consequent extinction of less-favoured forms almost inevitably flows.

Whether one agrees with the Darwinian emphasis on competition (and indeed Darwin himself did not always focus on competition throughout his writings), the point is that the metaphor itself can be a powerful interpretive force, providing a particular and always limited perspective on that which it seeks to explain. The constitutive power of the competition metaphor is important to keep in mind when discussing the use of business models in the non-profit sector.

The competition metaphor is endemic to business models, and its influence on the Performance Index is quite evident. The Donner

Awards brochure (Fraser Institute, 2003e) states that the results of the Performance Index provide “a useful tool for corporations, foundations and individuals to evaluate the organizations *competing* for their charitable dollars” (emphasis added). The byline from a media release announcing the merger of the Donner Awards with another awards program (which was never finalized) noted that it is the “only program in Canada to measure performance *against* peer groups” (emphasis added) (Fraser Institute, 2000). As well, the annual performance report (Wood & Clemens, 2002, p. 6) notes: “Data from the application is then used to objectively assess the agency’s performance on a comparative basis in ten performance areas ... Agencies are, therefore, *rated against each other* rather than assessed on the basis of an imposed standard” (emphasis added).

The veracity of the claims to objectivity and to not measuring against an imposed standard will be discussed further below, but the above passages do help underline the important role of competition in the Performance Index. Non-profits are competing against each other to perform best, win awards, and attract funders.

Progress

The idea of evolving to a higher or more progressed state of being and how those terms were defined was an issue that Darwin struggled with. For Darwin, progress simply meant achieving a better state of adaptation to the current environment (Gould, 1977, p. 37). This provided no guarantee of long-term prosperity or survival, of course. A rapid change in the environment could easily leave behind those forms that were more highly adapted to the previous conditions. Without expressly invoking a Darwinian metaphor, Patton (1997, p. 106) speaks about the fact that changes within organizations are not necessarily improvements, either: “changes are dictated by your current perceptions, but the commitment to change doesn’t carry a judgment that what was done before was inadequate or less effective. Change is not necessarily progress. Change is adaptation.”

Evaluating performance and improving an organization are, however, geared toward progress. If the point of an organization is to accomplish a given mission and it is not yet completely doing so, then there is room for progress — to reach a higher level. But how one defines progress is key. Applying this to the Performance Index, one can see that organizations are being judged against an ideal

“higher” state. Despite the model’s contention that the organizations are only being judged against each other, it is quite obvious that they are being judged against each other according to a particular set of standards. Admittedly, using statistical methods, the relative distribution can be expressed in any manner thought desirable — a scale of 0 to 10 (which is what the Index does use), as a percentage of the highest scoring organization, and so on — but who scores high and who scores low is very much correlated with what exactly the standard is. An organization that buys into the standards set by the model will do better on the Index, whether it be by, for example, reducing staff turnover, building up an accumulated surplus, or avoiding unstable government funding. “Higher” performance means progressing toward the Performance Index’s view of what is desirable. If the Index’s definition of progress were compatible with a definition based on progress relative to mission accomplishment, then no difficulties would likely arise. However, these two types of progress are not synonymous. High performance under the Performance Index could help an organization perform better, but many other important variables are also at stake. Questions that might arise in process and impact evaluations (Scheirer, 1994, p. 40) such as: “What impact did the services have on the beneficiaries?”; “Were resources put towards the areas of greatest effectiveness?”; “Was the cause of the problem giving rise to the need addressed?”; and “Was the organization’s mission fulfilled in whole or in part?” are not the focus of the Index.

To say that doing well on one indicator of success (the Performance Index) translates into overall excellence in non-profit service delivery is a large leap of faith. The fact is that only the top 10% of scores in the Index are eligible to carry on to the second stage of the Donner Awards process (the qualitative essay questions and references process), and thus, the 90% of organizations that may be excelling on perhaps a wider definition of performance are weeded out — they are “lower” and less fit for the environment defined by the model.

THE INFLUENCE OF BUSINESS PERSPECTIVES

As noted above, the Performance Index approach is rooted in a business view of the world. Consistent with the general approach of the Fraser Institute (Fraser Institute, 2002), the Index borrows much from the lessons learned in the competitive world of the market, and the awards program does not hide the fact that it helps non-profits become more “business-like” (Clemens, 2000). Being more

business-like and efficient is assumed to be naturally better. Several authors challenge this assumption, however, as it applies in other fields (Ehrenfeld, 1978, pp. 246–247; Evernden, 1992, pp. 5–6; Livingston, 1994, pp. 72, 91, 176–197; Worster, 1985, p. 315).

There is a difference between a very utilitarian business-like view of the world and one that places value on the experience of a particular beneficiary of social services instead of just the numbers served. There are legitimate differences in what is deemed to be good and desirable according to different perspectives (Zimmerman & Dart, 1998, p. 15). The Performance Index tries to respond to these differences by creating an external authority to which we can look for direction. That authority is rooted not in the values of the non-profit sector but in the perspective of market solutions to social problems. Given that the driving forces behind the two sectors are very different, it is naïve to believe that a business model is appropriate across the board in the non-profit arena or vice versa. I share the concern of Zimmerman and Dart (1998, pp. 16–17) that business models are presently permeating society, including the non-profit sector:

If we accept the assumptions underlying the commercial model without examining the context or limitations of their application, we could blindly or unconsciously apply inappropriate decision models to societal choices ... The market ideology is powerful because it has 1) the support from groups that control the resources; 2) a consistent, coherent language; and 3) measurement tools. The market rules and logic will “decide” which charitable activities are valuable in society.

The Performance Index is clearly a “measurement tool” steeped in the business model mindset and this general warning seems particularly apt to any discussion of the Index’s potential influence on non-profits. The ability to measure is alluring. Stein (2001, pp. 151–152) discusses the attraction of simple efficient measurement within two main public services (health and education):

What is measured matters, because what is measured is what people do ... It matters profoundly which measures of effectiveness are chosen. It matters because these measures — once they are socially accepted — feed back to, and drive, the performance of those who provide public goods.

In the non-profit context, Kanter and Summers (1987) (see also Hayes, 1996, pp. 86–91; Kerr, 1995, pp. 7–14; and Patton, 1997, pp. 159–161, 316) warn that a tendency can arise to manage for the indicators of effectiveness, instead of performance itself, which is particularly likely for organizations that serve very broad, and thus difficult to define, missions:

When goals are vague or ill defined, effectiveness criteria may themselves become substitutes for goals, particularly when they are more precise and suggest concrete actions. ... When immediate effectiveness measures set the standards for the organization, a tendency can arise to favor the short term over the long term — to maximize the score on indicators of today's performance. (Kanter & Summers, 1987, p. 156)

Knowing how effective an organization is depends very much on what definition of effectiveness or performance is being used. As noted by Kanter and Summers (1987, p. 156), attention needs to be paid to the “larger mission” and to temporal aspects as well (short-term and long-term effectiveness measures may be very different). Generating this type of knowledge is necessarily subjective in part, because it will depend on the nature of the organizational mission. No “true” overarching measure like profit will be found for all non-profits to gauge their overall performance.

Simple and attractive measurements (such as the Performance Index) pose a risk: organizations may manage for them because there is no easy alternative. Rather than recognizing that an organization's mission can be translated into some specific measurable (some indeed quantifiable) goals against which performance can be judged, the simplistic one-size-fits-all approach may become the convenient choice — especially where it leads to a direct monetary reward. To borrow a term from Gould (1981), simple measurements may in some cases simply be mismeasurements. In his discussion of biological determinism, Gould (pp. 24–25) examines how simplistic measurements and the fallacies of reification (i.e., converting “abstract concepts into entities”) and ranking (“ordering complex variation as a gradual ascending scale”) have helped perpetuate historical discrimination within the human species.

The Performance Index suffers from both types of fallacies. Because it converts an abstract notion (or at least a diverse one), that is, the

accomplishment of many different types of missions in the non-profit sector, into one simple entity, a performance excellence score, it suffers from reification. More obviously, by ordering many types of organizations (with different missions, even though they are in similar fields) into one ascending scale, it is clearly based on ranking. But reification and ranking are not conducive to an examination of diverse, complex matters such as judging performance across entities that have different missions. The Performance Index may be an efficient attempt to judge performance with the resources its proponents have assigned to the project, but it is not likely accurate in gauging overall performance. As such, its utility in actually accomplishing what it sets out to do (i.e., recognizing and rewarding excellence in non-profit program provision and providing useful, pertinent performance information to assist non-profits to improve service delivery [Fraser Institute, 2003e]) is questionable.

Alongside the notion of competition, efficiency occupies a prominent place in the business world, and the Performance Index is heavily influenced by the concept of efficiency. The Donner Awards place significant emphasis on efficiency of service delivery. But filling a non-profit role may have little to do with the numbers that provide the handy metric in the business world, as noted by Zimmerman and Dart (1998, pp. 17–18) in their study on charities undertaking commercial ventures. They emphasize, in particular, the temporal differences in approach between the non-profit and business sectors (e.g., long-term vs. short-term). Another simple example would be the many important services for the needy for which there is no prospect of profit. Clients with no capacity to pay may create an important need for non-profit services but create no profit-oriented niche in the market economy. These long-term and wider-view needs, typically serviced by the non-profit sector, may be given less attention if that sector adopts more of a business-like mentality. Zimmerman and Dart (1998, p. 18) warn:

If a commercial mentality is essentially based on self-interest, charities undertaking commerce run the inherent risk of losing their focus on the public interest. Also, and at a more macro level, the public risks losing its understanding of the distinct moral metrics of the public and charitable sectors.

Many of the general conclusions reached by Zimmerman and Dart (1998) shed light on business-like approaches to performance evalu-

ation. Just as undertaking commercial ventures can cause a fundamental shift in the culture of an organization (p. 42), so undue attention paid to the Index's individualized Confidential Performance Reports may unnecessarily or unwisely lead to decisions to change the nature of an organization. Before the recommendations implicit in these reports are acted upon, I believe that an organization must first return to its mission, goals, and objectives as a starting point for deciding whether to evolve in the direction pointed to by the Performance Index (Drucker, 1990, pp. 87–88; Scheirer, 1994, p. 40; Wholey, 1994, p. 15; Zimmerman & Dart, 1998, p. 42).

THE PERFORMANCE INDEX AS AN ANSWER TO THE “MEASUREMENT CHALLENGE”

The “Measurement Challenge,” a section of the Wood and Clemens (2002, p. 9) report, sets out part of the rationale for the development of the Performance Index:

Unlike the for-profit business sector, the non-profit sector is hampered in its ability to assess performance due to the lack of an objective, quantifiable performance measure. The for-profit sector relies on a number of objective measures to assess performance, including profitability, market share, and return on assets. The existence of standard, objective performance measures in the for-profit sector allows for comprehensive and comparative performance analysis. Unfortunately, there is no such parallel for the non-profit sector. ... The performance measurement system developed for the Awards Program represents a major step forward in the ability of non-profit organization managers to assess performance on an objective and quantifiable basis.

The suitability of more quantifiable measures to the non-profit sector, however, is open to question. As Kanter and Summers (1987, p. 155) contend, non-profits define themselves around their mission or services, not around their financial returns, and it is this “centrality of social values over financial values” that makes measurement complicated. Similarly, Drucker (1990, pp. 81, 107–108) notes that “the results of a non-profit institution are always outside the organization, *not* inside” (see also Patton, 1997, p. 13). Rather than inward-looking empire-building institutions, non-profits are outward looking in their approach — seeking to effect change pursuant to their particular mission. In other words, the question of whether a

non-profit is excelling at carrying out its mission will not be conclusively answered by posing questions about the internal workings of the organization. This leads us to the question of whether the Performance Index and the Donner Awards are intended to measure the excellence with which an organization carries out its mission or, rather, whether they are intended to measure only the efficiency aspects of performance.

The literature published in conjunction with these awards is somewhat confusing on this point. Different words such as excellence, efficiency, performance and effectiveness are used to describe the thrust of the initiative (Fraser Institute, 2003b, 2003c; Wood & Clemens, 2002, p. 6). There are important differences between, for example, efficiency and effectiveness as noted by Osborne and Gaebler (1992, p. 351). The Performance Index's ten areas of measurement (each of which gets equal weight in the final overall performance score) emphasize efficiency measures (e.g., cost containment, program spending versus overall spending, size of accumulated surplus relative to expenses, use of volunteers relative to staff, turnover, cost per hour of programming provided, etc.). The areas of measurement include some components that could, in theory, help measure the important outcomes pointed out by Drucker (1990), which relate to mission. In particular, the "Strategic Management" and "Outcome Monitoring" areas appear to be good candidates. However, both areas of measurement constitute very rudimentary inquiries into these critical subject matters. Under "Strategic Management," five simple questions are asked of the organization (see Appendix 2), which could honestly be answered in a way that yields high scores even if the type of strategic management that is undertaken is failing to achieve good results. More specifically, it is possible to have a concise mission statement, annual goals and objectives for programs, staff, and volunteers, a structured mission updating process, and staff involvement in strategic management and goal setting, but such would not necessarily equate with either effective strategic planning or achievement of desired outcomes. What is being measured is whether strategic management tasks are being carried out, not whether they are being carried out well. Most competent organizations would meet these criteria and, in fact, the annual performance report concludes that a substantial proportion of organizations demonstrate "superior performance" in this area:

There were strong performances in all service categories in the area of Strategic Management, with the average and median scores ranging between 7.8 and 9.9

[out of 10]. Almost one-third of agencies received a score of 10 for Strategic Management, which indicates superior performance. (Wood & Clemens, 2002, p. 21)

While such high scores might indicate “superior performance” in some circumstances, it is more likely that what is actually being measured is the ease with which an organization can get a high score when such basic indicators of good “Strategic Management” are used as the evaluation tool. This demonstrates the importance of choosing valid indicators of performance.

Similar concerns arise in the “Outcome Monitoring” area of measurement, also comprising five simple questions (see Appendix 2). What the questions measure are not actual outcomes but rather whether outcomes are monitored within an organization. The problem, of course, is that one could diligently monitor outcomes, but fail miserably year after year in reaching them and yet still receive a high score in all five categories. Seemingly oblivious to the effects of having chosen such basic measures of performance, the annual performance report (Wood & Clemens, 2002, p. 33) notes:

The scores for Outcome Monitoring are relatively high with the average and median scores for all categories falling in the range of 7.1 to 9.5. This indicates a relatively high level of average performance in terms of managing and pursuing outcomes. One-third of all agencies received a score of 9 or 10, which indicated high performance.

The graphs set out in the annual performance report (Wood & Clemens, 2002, pp. 20, 32) confirm that the vast majority of organizations are clustered in the top performance classes for both “Strategic Management” and “Outcome Monitoring.” Given the lack of statistical differentiation amongst organizations in these two areas, they end up having little effect on the overall outcome of which organizations will be considered high performers under the index. The remaining eight areas reveal much greater ranges in performance, and their relative impact on the overall comparative results is consequently higher (that is, though each of the 10 areas gets equal weight, if some categories yield nearly identical results for all organizations because they are so simple to pass, the relative weight of the other scores thereby increases as they provide better means to differentiate amongst organizations). The result is that these two

critical areas of performance (“Strategic Management” and “Outcome Monitoring”) have, mathematically, only a 20% effect (2 out of 10 equally weighted areas in the overall score). Further, their actual impact on which organizations emerge as finalists is likely much lower because the shallow measures being used are not uncovering critical differences amongst organizations.

Does the Performance Index simply suffer from asking too easy questions in these critical areas? Should the search begin for better questions that will better reflect good or excellent performance? Better questions will inevitably be context-specific. If other factors were measured, similar to those suggested by Drucker (1990) (e.g., did performance improve, did we put resources in the areas of high results, what percentage of beneficiaries achieved the outcome sought, to what degree was the mission accomplished?), a more informative picture of excellence in these two areas might be painted. Unfortunately, meaningful results could be elusive. Indeed, the reason that such basic measures were chosen for the Performance Index may be that this is all that is common across non-profits in these critical areas of measurement. Delving further might require actually undertaking organization-specific inquiries, but the Performance Index notes that “the scope for comparison is limited and costly” (Wood & Clemens, 2002, p. 9) in this kind of subjective inquiry. By searching for the common denominator on critical performance issues, the developers of the Index may have inadvertently proved a very key point — that is, that meaningful performance evaluation does not lend itself to a simple one-size-fits-all recipe based on a few basic measures, which on their own may have little correlation with those organizations that perform their missions well.

It is the vast area of context and mission-specific questions that will likely reveal more about performance. To ask these types of questions, one would have to abandon the “common denominator” approach. This is not to say that quantitative approaches are to be avoided. Both quantitative and qualitative approaches have an important place in evaluation (Patton, 1997, pp. 273–277). Rather, rigorous performance evaluation, whether quantitative or not, will require approaches tailored to the organization’s mission and its many constituencies. Indeed, mission-driven context-specific performance evaluation questions can still be quantitative, as Drucker (1990, pp. 107–108) notes. Yet it is difficult to conceive of any overarching generic questions that will accurately gauge performance and be applicable across a wide range of non-profit organiza-

tions. The common denominators that might emerge, such as “for the public or common good” or “for the public trust” yield no measurable metric that can be used across the entire sector.

In essence, with the Performance Index, non-profits embark on a journey (toward finding the common objective measure of non-profit excellence) whose destination might not exist. Its proponents can assert that “until now” the non-profit sector has been “hampered in its ability to assess performance due to the lack of an objective, quantifiable performance measure” and that the Index “represents a major step forward in the ability of non-profit organization managers to assess performance on an objective and quantifiable basis” (Wood & Clemens, 2002, p. 9). But alas, it is only an assertion — strongly stated but without much evidence.

The question then remains: What does the Performance Index measure? Is it possible that aspects of the model are worth retaining? The Performance Index may assist in measuring efficiencies and financial performance, two areas that are critical to high performance in the business world. Many of the criteria would help an organization gauge its financial stability, its growth, its efficient use of staff, volunteers, and resources, and so on, and lend themselves to easily quantifiable measures. If one accepts the assumptions behind the standards set, one can judge performance in some of these areas. For example, if one agrees with the proposition that the “optimal size of the accumulated surplus is equal to one year’s annual expenses” (Wood & Clemens, 2002, p. 19), the index will assist in evaluating performance toward that goal. Of course, it is far from certain that each non-profit should accept that this is the optimal situation for it. Circumstances vary, and in some cases non-profits may rightly choose to adopt a high discount rate and to spend an accumulated surplus (for instance, when an unusual opportunity to help fulfill its mission arises). Nevertheless, as long as non-profits are aware of the assumptions underlying some of the simple measures in the Performance Index and check their validity against their current circumstances, some of these measures may be helpful in improving efficiency, financial management, and so on in the organization. They may be useful in a targeted way.

Even so, this generalized approach and the assumptions that underlie it are open to question. For example, the Index examines sources and concentration of revenue, in particular with respect to government sources, and in so doing assumes both that long-term funding stability is necessary and that reliance on government fund-

ing runs contrary to this goal. These assumptions may be questionable, however, for many non-profits.

First, with regard to government funding, while at the macro-level in recent times in Canada, government funding of many non-profits has not been stable, it has been stable or increasing for others. Why should the latter organizations be penalized in the Performance Index simply because their individual circumstances are not reflective of the general trend? Indeed, for them, it is possible that long-term stability is better achieved through government funding than through other sources. The Index's approach to government funding is very much in line with the Fraser Institute's desired state of affairs of less government spending. It is an open question as to whether the Index was intentionally designed to cause organizations to evolve and adapt to such an environment. But insofar as the model singles out unstable government funding, and not unstable funding generally, it leaves itself open to criticism that its assumptions may be ideologically driven (i.e., toward lower government spending) instead of logically driven to penalize all types of unstable sources.

Second, the assumption that unstable sources (including government sources and others) are not worth pursuing is questionable as well. If an organization is cognizant of the instability of certain sources, it can pursue them but be on the ready to live without them as well. Planning and living with many uncertain sources may in the aggregate allow the organization to better fulfill its mission than relying on smaller stable sources. Yet those organizations that pursue unstable sources of revenue (such as some government sources) and plan accordingly will be penalized under the Performance Index. Further, since this measure looks at the total amount of revenue from all levels of government, an organization that is funded by many different types of government agencies would be penalized, even though it would be very unlikely that its entire portfolio of government funders would dry up simultaneously. While such an organization would be rewarded for having a diversified income portfolio, it would be penalized under the government funding criterion.

THE EVALUATION PURPOSES OF THE INDEX AND AWARDS: JUDGEMENT AND IMPROVEMENT

The Performance Index is designed to accomplish two purposes: render judgement and facilitate improvement. Generally, judgement-

rendering evaluations tend to be spawned by calls for “accountability” and are aimed at determining “overall, merit, worth, or value.” They often use “comparative ratings or rankings of programs” (Patton, 1997, p. 65), though there are significant obstacles to comparing effectiveness across organizations (Hannan & Freeman, 1977, p. 107). Patton (1997, p. 66) notes the importance of criteria selection in judgement-oriented evaluations: “In judgment-oriented evaluations, specifying the criteria for judgment is central and critical. Different stakeholders will bring different criteria to the table.” While the Donner Awards and the Performance Index constitute a “judgment-oriented” evaluation, it is not clear how reflective the chosen criteria are of the various stakeholders involved. The evaluation protocol was developed with input from four non-profit organizations (Fraser Institute, 2003c, 2003e), but the literature provides no information about whether these or other non-profits or funders endorsed the final protocol or whether social service beneficiaries were involved in criteria selection. Though the Donner Canadian Foundation has sought the endorsement of other funders (Fraser Institute, 2000), there is no evidence that such endorsements have materialized. With little information provided on how the criteria were selected vis-à-vis the most obvious stakeholders (non-profits, funders, beneficiaries), the evaluation protocol may be reflective only of the views of the two proponents (i.e., the Fraser Institute and the Donner Canadian Foundation).

The importance of the evaluation design to the intended users is underscored when an evaluation also includes an improvement facilitation purpose (Wholey, 1994, p. 17), as this model explicitly intends. These “formative”-type goals (as opposed to “summative” goals that render judgement) focus “on ways of improving and enhancing programs” (Patton, 1997, p. 67). The Fraser Institute was ambitious in trying to satisfy equally these two very different types of purposes, given the different characteristics of each type of evaluation. As Patton (1997, p. 68) explains: “Using evaluation results to improve a program turns out, in practice, to be fundamentally different from rendering judgment about overall effectiveness, merit, or worth.”

While attempting to satisfy both types of purposes, this particular evaluation tool is notably weighted more toward the judgement-oriented approach. And yet it is far from evident that the judgement-oriented aspects of the evaluation protocol have employed defensible and supported criteria. Indeed, the criteria are notably lacking in the critical results-based measurements that would be needed to

render judgement about whether services were in fact delivered with excellence. The critical questions of how the life circumstances of beneficiaries changed as a result of the service that was delivered and how well the organization's mission was carried out vis-à-vis those beneficiaries go unanswered in the evaluation.

The Performance Index and Donner Awards suffer from two key design weaknesses vis-à-vis their stated purposes of recognizing and rewarding excellence and providing information to improve operations. First, the program shows little evidence of having chosen criteria that are direct indicators of excellence in service delivery or of having employed criteria that are broadly supported by users other than the program's proponents and those organizations that succeed in obtaining an award. Second, the program fails to include nearly all the approaches typical of evaluations intended to improve operations as discussed by Patton (1997, pp. 68–70). It is rigid and judgement-oriented and focuses on providing direction on how an organization can improve in the next year's judgement-rendering exercise. This tends to create an incentive for non-profit users to manage for the indicators of success rather than managing for actual results. This is especially true in the design of this particular evaluation protocol since the two components that may provide some guidance on managing for actual results (i.e., "Strategic Management" and "Outcome Monitoring") are too simplistic to provide meaningful organization-specific guidance. Figure 1 summarizes what I see as the main deficiencies in the Performance Index.

Figure 1
Key Deficiencies of the Performance Index

- Its one-size-fits-all approach, which compares and ranks performance across organizations, is not adapted to the diversity of missions in the non-profit sector.
- It focuses on efficiency and process, with little emphasis on actual service delivery effectiveness or outcomes.
- It may cause users to manage for performance indicators in order to improve their score for the awards rather than to manage directly for excellence in accomplishing the organizational mission.
- It makes unsubstantiated claims to objectivity and to not ranking non-profits against an imposed standard.
- Its emphasis on simple quantifiable measures leads to superficial assessment in key areas like strategic management and outcome monitoring.
- Through its desire to encourage non-profits to be more business-like, the Index appeals to values that are often at odds with the charitable activities and public interest outlook of many non-profits.

We now turn to a consideration of the practical implications for non-profit organizations.

IMPLICATIONS FOR NON-PROFIT ORGANIZATIONS

The Donner Awards and the Performance Index have not attained a very high profile as yet. However, they do typically receive nearly 300 applications annually and have enjoyed the credibility associated with having the awards given out by the Lieutenant Governor of Ontario each year. Award recipients have also received, for example, positive coverage in the media and recognition in the Hansard of legislatures. As well, according to the testimonials from non-profits provided by the Fraser Institute, the Performance Index is having a direct impact on these organizations and also the funding community (Fraser Institute, 2003a).

As well, the Donner Canadian Foundation has made it clear that it hopes the Awards will attain greater currency. The proponents hope to have an effect on other funders by providing “a useful tool for corporations, foundations and individuals to evaluate the organizations competing for their charitable dollars” (Fraser Institute, 2003e, p. 1). Indeed, the latest performance report profiles many non-profits that have performed well in the Awards Program and includes detailed information on each organization, presumably to affect, in part, the decisions of other funders (Wood & Clemens, 2002, pp. 39–78).

Given the intended influence of the Index and Donner Awards on the decisions of non-profits and funders, it is critical that the approach benefit these intended users. The evaluation must be designed so that it exerts a positive influence on the evolution of the non-profit organizations being evaluated. The difficulty, however, is that what is to be considered positive will very much depend on the individual characteristics of each non-profit. The nature of one non-profit may lend itself to adding more volunteers, reducing government funding, having a substantial accumulated surplus, and so on, but for many others, this may not be the case. By generalizing what it means to be doing well, the model misses the mark on what is most important to every non-profit: whether the decision will improve its ability to carry out its specific mission.

If organizations do not keep these questions in mind and too easily adopt the blunt and simple evaluation instrument offered by the Performance Index, they may embark on a series of changes to im-

prove their performance according to the Index's methods. Evolution can happen very quickly in an organization if it explicitly embarks on a new path. If the Index gets widespread support (in this case perhaps because of the lack of any other allegedly comprehensive quantitative non-profit evaluation tool as a competitive force), rapid changes to participating non-profits may occur. But will they be more adapted to their overall environment or just more fit to the microenvironment of an artificial breeder?

I believe that a diversity of approaches is preferable — that non-profits explore a range of evaluation tools and employ those that are best suited to their circumstances and best able to provide guidance on better accomplishing missions. There is no single sequential path like the Performance Index that is suitable for all organizations to follow. As Gould (1977, p. 62) notes in describing humankind's place in the process of evolution: "*Homo sapiens* is not the foreordained product of a ladder that was reaching toward our exalted estate from the start. We are merely the surviving branch of a once luxuriant bush." The danger of having only one artificial breeder acting as the evaluation environment is that we would lose the diversity of the non-profit "bush" and evolve toward a monocultural approach, under the mistaken notion that the "ladder" will lead us to the promised land. But if we have a range of evaluative tools, each acting as a different selective force on different organizations, diversity might be retained or fostered.

I agree with Zimmerman and Dart (1998, pp. 11–13) that, at a minimum, diversity is needed across the societal sectors and within the non-profit sector. The question is how that diversity within the non-profit sector might be affected by the use of one dominant tool for quantitative evaluation.

The quantifiable aspects inherent in business models make them attractive to many audiences. Zimmerman and Dart (1998, p. 45) put it this way:

Part of the appeal of commerce is the tidiness of the models. There are excellent measurement tools, evaluation tools and decision models within the market framework. We have not put the same effort into developing alternative measurement concepts, language or evaluation concepts for community or public good actions.

Will these alternatives ever be developed? If an equivalent approach could be found that succeeded in effectively evaluating non-profit performance, it would be inappropriate to reject it simply because it shared some similarities with business models. However, I doubt whether simple common denominators like profit or return on investment have analogues waiting to be found in the non-profit sector. The non-profit sector is so diverse that the evaluative tools that could be seen to be transferable sector-wide could only be very general in nature. The search for very specific quantifiable tools is probably one that should be undertaken only at the organizational or subsectoral level (i.e., within a group of similarly situated organizations) for the tools to be meaningful and appropriate for their contexts.

Part of the answer may lie in the words we use to describe the sectors themselves. One sector is the profit sector, the other the non-profit. The former is incredibly diverse, of course, but shares the common thread of making a profit, which is quantifiable (though I admit that performance evaluation is not, therefore, “easy” in the for-profit sector). The latter is the non-profit sector, which by definition can encompass almost anything else. Can this residual “anything else” be characterized by any common thread or common denominator? I would argue that the sector does share a common thread (for example, seeking the betterment of society) but that this very general common thread is neither amenable to a common metric nor even amenable to a common definition. Indeed the missions of two non-profits could have polar opposite goals because the members being represented or the constituency being served hold opposite views of what constitutes the betterment of society. One of the consequences of a heterogeneous society and social relativism is that different groups may have completely different concepts of “what constitutes the good life” (Evernden, 1992, pp. 5–6). Just as there is a diversity of views about what standards of excellence should be striven for at the individual level, so too is there a diversity across non-profits. So long as that diversity persists, the one-size-fits-all non-profit excellence evaluation tool will never be found, because it will not exist.

Gould (1977, p. 84) noted that “Darwin’s theory of natural selection requires that evolutionary changes be adaptive — that is, that they be useful to the organism” (see also Patton, 1997, pp. 382–385). Would succeeding according to the Fraser Institute’s criteria really be a useful adaptation to the individual organization or to the non-

profit sector? Or would it just succeed in causing the organization and sector to “evolve” in the direction that the Fraser Institute wants them to go? “Darwinian evolution decrees that no animal shall actively develop a harmful structure, but it offers no guarantee that useful structures will continue to be adaptive in changed circumstances” (Gould, 1977, p. 90). Non-profits and their funders should be careful in determining how the Performance Index model will influence them. They may inadvertently change themselves in a harmful way, rendering them hyper-adapted to the model (Livingston, 1973, pp. 79–81), but not particularly fit for the overall environment in which they operate.

If simplistic convenient measures such as the Performance Index are allowed to shape the development of non-profits, then non-profits will have allowed themselves to be artificially bred to thrive in the environment defined by business and efficiency. They may become superbly engineered to act in a world that in many ways is completely antithetical to the caring, compassionate, and altruistic beliefs that draw individuals to the non-profit sector to start with — artificially bred and unnaturally selected to be more adapted to live in an environment that in many ways created the need for the multitude of non-profits to start with — a world where costs can be externalized to the powerless; a world where quality is subordinate to quantity. Accepting the business model as the source for evaluating non-profits would be tantamount to accepting a mismeasure.

Just as the prevailing imperialism of Britain in the 19th century influenced the approaches of Darwin, especially with his emphasis on competition, a focus on business models and the attendant commandments like efficiency is permeating society now. That prevalence is exerting an influence on the way non-profits manage themselves and how they evaluate performance. A new lexicon will eventually replace it. In the meantime, if the “cult of efficiency” (Stein, 2001) is embraced by the non-profit world, what will it then evolve toward, or, to use a Darwinian term in a more expansive manner than he perhaps intended, what will it *descend* to? (See Worster’s [1985, p. 181] discussion of the title of one of Darwin’s books: *The Descent of Man*.) It will be more fit within an efficient business-like environment, but will it be better, higher, lower? Will it have simply *descended* with modification, as Darwin would say, or will it have truly *declined* against the measuring stick of helping achieve the betterment of society?

ADAPTING EVALUATION TO AN ORGANIZATION'S ENVIRONMENT

On the subject of performance evaluation for non-profits, I would like to suggest a path that is more similar to Gould's branching bush than to a ladder. Different evaluation questions should be posed for different branches (subsectors) within the non-profit field. By knowing what adaptations are potentially useful to their specific missions, organizations can try to evolve and better adapt to the circumstances within which they are embedded. Patton's (1997, pp. 131–133) repeated focus on making evaluation useful for the intended users is especially apt in this context: "Utilization-focused evaluation is a problem-solving approach that calls for creative adaptation to changed and changing conditions, as opposed to a technical approach, which attempts to mold and define conditions to fit preconceived models of how things should be done."

Patton's (1997, p. 135) general call for more "situational responsiveness" fits well with a more context-specific and mission-driven approach to non-profit performance evaluation. Patton (pp. 137–138) goes on to state that part of being situationally responsive includes being respectful of the stakeholders by not intimidating or manipulating intended users. The Performance Index points the way along the "business-like" ladder without ever acknowledging that businesses are just as susceptible to failure or underperformance, even with their objective measures of profit. What makes a particular business prosper has little to do with the availability of a common denominator like profit, but everything to do with its specific strengths and weaknesses vis-à-vis the particular market environment in which it is acting (which makes performance evaluation in the for-profit sector much more involved than simply looking at profit margins). These important context-specific matters are left out entirely from the Performance Index and enter into the equation only for the handful of organizations that score high and are eligible for the final stage of the Donner Awards process, during which essay questions are posed and letters of support provided (Wood & Clemens, 2002, p. 7).

In his call for useful evaluations, Patton (1997, pp. 16–17) discusses the standards developed by the Joint Committee on Standards for Education Evaluation: utility, feasibility, propriety, and accuracy. He concludes: "Implementation of a utility-focused, feasibility-conscious, propriety-oriented, and accuracy-based evaluation requires

situational responsiveness, methodological flexibility, multiple evaluator roles, political sophistication, and substantial doses of creativity, all elements of utilization-focused evaluation.” The Performance Index lacks many of these characteristics.

While I can offer no simple alternative to the Performance Index for non-profits, I agree that testing performance remains a key task to be undertaken, for performance “is the ultimate test of any institution [and] every non-profit institution exists for the sake of performance in changing people and society” (Drucker, 1990, p. 107). Finding or creating evaluation tools that help measure just that — how an organization’s actions have effected real change — is what is needed. For each organization, fashioning such a tool should not be exceedingly difficult, though for the long-term goals included in some missions, very long time horizons may be involved, making evaluation challenging.

Returning to Drucker’s (1990) notion of focusing on the outside rather than the inside, it makes sense to look at what change is occurring as a result of the non-profit’s performance:

The non-profits are human-change agents. And their results are therefore always a change in people — in their behavior, in their circumstances, in their vision, in their health, in their hopes, above all in their competence and capacity. The non-profit institution therefore needs to set specific goals in terms of its *service* to people. (Drucker, 1990, p. 85)

This outward-looking perspective nests well within mission-driven approaches. As Drucker notes (1990, pp. 87–88, 108–109), the mission of non-profits is not simply to build a well-run non-profit organization but to build a better society. In addition, performance measurement needs to be cognizant of the many constituencies at play (funders, donors, clients, the public, etc.) (Hall et al., 2003, p. 1; Kanter & Summers, 1987, p. 164). Because, for example, most non-profits have very different stakeholders when it comes to funders and beneficiaries (except in commercial fee-for-service areas), it is important to demonstrate performance to funders and the broader public. Non-profits, especially charities, are given certain privileges (such as those relating to taxation and charitable tax receipts), and that means that members of the public (through government but not exclusively so) need to be able to gauge performance if inter-

ested in doing so. The diversity within the sector makes such communication difficult, and as a result there will always be those that state that the sector is therefore unaccountable. In truth, the sector is not defined by an obvious quantifiable common measure, but a non-profit can still be accountable by establishing a mission that provides a public benefit, carrying it out well, and providing evidence that it is doing so through performance measures that are appropriate to it.

Returning to the Darwinian metaphor, I believe that being mindful of several of Darwin's evolutionary terms could help organizations design and implement appropriate performance evaluation exercises. Organizations that keep their mission at the forefront and that wish to assess and improve their performance could ask themselves some of the following questions:

- What does it mean for our organization to be *fit*? What standards should be set to help gauge *fitness* vis-à-vis our mission, goals, and objectives? Does *fitness* in our context necessarily equate with maximizing chances of organizational *survival*?
- How do we define *progress*? What measures outside our organization should be used? To what extent does the timeframe being considered affect the definition of *progress*? Having regard to our mission, how will we assess when we are enjoying *high* performance or *low*?
- What *adaptations* can be made to make ourselves more *fit*? What *external environment* should we examine in helping to define what it means to be *fit* and thus what *adaptations* should be *selected*? To what extent does an environment characterized by the notions of *competition* and *efficiency* apply to our organization or parts thereof?
- To what extent are the assumptions that help define the *external environment* of an evaluation directly applicable to our context? Are the *adaptations* being recommended a *natural fit* with respect to our mission?
- How will we critically assess recommendations for change to make sure that we do not allow ourselves to be *artificially bred* into a different form that offers no advantage to accomplishing our mission?

Organizational change is clearly a matter of choice. In many ways, it is about deciding how to evolve, or more passively, how to be bred

and moulded by others. While acknowledging the risks of pushing the Darwinian metaphor perhaps too far, I submit that adopting diverse mission-driven evaluation processes is a more *natural* fit to the non-profit setting than the one-size-fits-all business-like approach advocated by the Performance Index.

ACKNOWLEDGEMENTS

I wish to thank the late Bruce Miller and my other colleagues in the McGill-McConnell Team Integrative Project (Shawn Bayes, Karen Flannery, Marc Johnson, and Valerie Wilder), whose collective idea it was to critically analyze the Performance Index as a group project. In the course of preparing our team report and presentation, they developed many of the insights that I present herein. I also wish to thank Anne Bell, Michael Quinn Patton, Janice Gross Stein, Frances Westley, Brenda Zimmerman, and the CJPE anonymous reviewers.

REFERENCES

- Aldrich, H.E. (1979). *Organizations and environments*. Englewood Cliffs, NJ: Prentice-Hall.
- Clemens, J. (2000, November 3). Non-profit organizations need to be more business-like. *Toronto Globe and Mail*, B11.
- Darwin, C. (1859). *The origin of species*. London: Penguin Books (Penguin Classics Edition 1985).
- Drucker, P.F. (1977). Managing the public service institution. In D. Borst & P.J. Montana (Eds.), *Managing nonprofit organizations* (pp. 16–31). New York: Amacom.
- Drucker, P.F. (1990). *Managing the non-profit organization: Practices and principles*. Oxford: Butterworth-Heinemann.
- Ehrenfeld, D. (1978). *The arrogance of humanism*. Oxford: Oxford University Press.
- Evernden, N. (1992). *The social creation of nature*. Baltimore: Johns Hopkins University Press.

- Fraser Institute. (2000, October 13). *Merger creates largest non-profit awards program in Canada: Only program in Canada to measure performance against peer groups* (media release). Vancouver: Author. Retrieved January 17, 2003, from <http://oldfraser.lexi.net/media/media_releases/2000/20001013-2.html>.
- Fraser Institute. (2002). *Annual report 2001*. Vancouver: Author.
- Fraser Institute. (2003a). *Award recipient testimonials*. Vancouver: Author. Retrieved February 10, 2003, from <<http://www.fraserinstitute.ca/donner/pdf/awards-testimonials1.pdf>>.
- Fraser Institute. (2003b). *Donner Award ceremony and recipients*. Vancouver: Author. Retrieved January 30, 2003, from <<http://www.fraserinstitute.ca/donner/award.asp>>.
- Fraser Institute. (2003c). *Donner Canadian Foundation Awards*. Vancouver: Author. Retrieved January 30, 2003, from <<http://www.fraserinstitute.ca/donner/index.asp>>.
- Fraser Institute. (2003d). *Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services 2003 application form and guide*. Vancouver: Author. Retrieved February 10, 2003, from <<http://www.fraserinstitute.ca/donner/pdf/application.pdf>>.
- Fraser Institute. (2003e). *Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services* (brochure). Vancouver: Author. Retrieved February 10, 2003, from <<http://www.fraserinstitute.ca/donner/pdf/brochure.pdf>>.
- Gould, S.J. (1977). *Ever since Darwin: Reflections in natural history*. New York: W.W. Norton.
- Gould, S.J. (1981). *The mismeasure of man*. New York: W.W. Norton.
- Hall, M.H., Phillips, S.D., Meillat, C., & Pickering, D. (2003). *Assessing performance: Evaluation practices & perspectives in Canada's voluntary sector*. Toronto: Canadian Centre for Philanthropy.
- Hannan, M.T., & Freeman, J. (1977). Obstacles to comparative studies. In P.S. Goodman, J.M. Pennings, & Associates (Eds.), *New perspectives on organizational effectiveness* (pp. 106–131). San Francisco: Jossey-Bass.

- Hayes, T. (1996). *Management, control and accountability in nonprofit/voluntary organizations*. Aldershot, UK: Avebury.
- Kanter, R.M., & Summers, D.V. (1987). Doing well while doing good: Dilemmas of performance measurement in nonprofit organizations and the need for a multiple-constituency approach. In W.W. Powell (Ed.), *The nonprofit sector: A research handbook* (pp. 154–166). New Haven, CT: Yale University Press.
- Kerr, S. (1995). On the folly of rewarding A, while hoping for B. *Academy of Management Executive*, 9(1), 7–14.
- Livingston, J.A. (1973). *One cosmic instant: A natural history of human arrogance*. Toronto: McClelland and Stewart.
- Livingston, J.A. (1994). *Rogue primate: An exploration of human domestication*. Toronto: Key Porter.
- Newcomer, K.E., Hatry, H.P., & Wholey, J.S. (1994). Meeting the need for practical evaluation approaches: An introduction. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 1–14). San Francisco: Jossey-Bass.
- Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector from schoolhouse to statehouse, city hall to the Pentagon*. Reading, MA: Addison-Wesley.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Scheirer, M.A. (1994). Designing and using process evaluation. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 40–68). San Francisco: Jossey-Bass.
- Stein, J.G. (2001). *The cult of efficiency*. Toronto: House of Anansi.
- Weick, K.E. (1977). Re-punctuating the problem. In P.S. Goodman, J.M. Pennings, & Associates (Eds.), *New perspectives on organizational effectiveness* (pp. 193–225). San Francisco: Jossey-Bass.
- Wholey, J.S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15–39). San Francisco: Jossey-Bass.

Wood, K., & Clemens, J. (2002). *2002 non-profit performance report: An analysis of management, staff, volunteers, and board effectiveness in the non-profit sector. Report based on the analysis undertaken as part of the Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services*. Vancouver: Fraser Institute.

Worster, D. (1985). *Nature's economy: A history of ecological ideas*. Cambridge: Cambridge University Press.

Zimmerman, B., & Dart, R. (1998). *Charities doing commercial ventures: Societal and organizational implications*. Toronto: Trillium Foundation.

Appendix 1

Components of Performance Measurement in the Performance Index

(Wood & Clemens, 2002, pp. 10-11, used with permission of the Fraser Institute)
See <www.fraserinstitute.ca/donner> for more information on the awards program.

Area of Measurement and Components

1. Financial Management
 - annual surplus—composite measure of the 4 year average and most recent year
 - revenue increase—composite measure of the 3 year average and most recent year
 - cost containment—composite measure of the 3 year average and most recent year
 - program spending versus overall spending—composite measure of the 4 year average and most recent year
 - financial reporting
2. Income Independence
 - number of sources of income adjusted for the average size of the donation
 - percentage of revenue provided by largest revenue source
 - percentage of revenue provided by government
 - size of accumulated surplus relative to expenses—composite measure of the 4 year average and most recent year
3. Strategic Management
 - use and prevalence of a mission statement
 - level of objective and goal setting
 - depth of involvement
4. Board Governance
 - independence
 - financial contributions
 - level of involvement as measured by frequency of meetings
 - level of participation as measured by attendance at meetings
 - policy guidelines to avoid conflicts of interest
5. Volunteers
 - use of volunteers relative to staff—composite measure of agency total and program total
 - recruiting activities
 - management and development of volunteers
 - donations other than time by volunteers
 - turnover

6. Staff

- level of programming provided by employees
- percentage of employees working in programs
- turnover
- management and development of staff

7. Innovation

- uniqueness of agency's program
- level of restructuring/change
- use of alternative delivery systems/technology in the delivery of services

8. Program Cost

- cost per hour of programming provided
- cost per client—*information only*
- hours per client—*information only*

9. Outcome Monitoring

- defining desired outcomes/goals for program
- measured actual outcomes
- desired versus actual outcome comparisons
- plans to deal with divergences

10. Accessibility

- process of assessing need and targeting assistance
- measurement of the level of usage by clients
- determination of the cause of a client's difficulties

OVERALL SCORE Composite of ten areas of measurement

Appendix 2

2003 Application Form for the Donner Canadian Foundation Awards for Excellence in the Delivery of Social Services

(Fraser Institute, 2003d, used with permission of the Fraser Institute)

See <www.fraserinstitute.ca/donner> for more information on the awards program.

SECTION ONE: AGENCY INFORMATION (Source of information: Whole agency)

- 1 a) Name of Organization _____
- 1 b) Street Address / Postal Box _____
- 1 c) City, Province, Postal Code _____
- 1 d) Phone Number (_____) _____ Fax Number(_____) _____
- 1 e) E-Mail / Web Site _____
- 1 f) Contact Person _____ Title _____
- 1 g) Year Organization (Local) was Founded _____
- 1 h) Indicate the program area (**one only**) to which your organization is applying for recognition (see category definitions in the Guide).
 - Child Care Services
 - Alternative Education
 - Provision of Basic Necessities
 - Crisis Intervention
 - Traditional Education
 - Services for People with Disabilities
 - Counselling Services
 - Prevention and Treatment of Substance Abuse
 - Services for Seniors

SECTION TWO: FINANCIAL MANAGEMENT (Source of information: Whole agency)

- 2 a) **INCOME STATEMENT.** Please complete the following tables using the **four** most recent years available. (Revenue and expenditure information should be available in your agency's income statement.)

	Most Recent Year	Previous Year	2 Years Previous	3 Years Previous
(i) Total Revenue (\$)				
(ii) Total Expenses (\$)				
(iii) Total Program Spending (\$)				
(iv) Total Non-Program Spending (\$)				

NOTE: (iii) plus (iv) should approximately equal (ii)

- 2 b) **BALANCE SHEET INFORMATION.** Please complete the following table using the **four** most recent years available. (All information should be available in your agency's balance sheet.)

	Most Recent Year	Previous Year	2 Years Previous	3 Years Previous
(i) Total Assets (\$)				
(ii) Total Liabilities (\$)				
(iii) Accumulated Surplus/Deficit or Fund Balance (\$)				

NOTE: (ii) plus (iii) should approximately equal (i)

- 2 c) To what extent does your agency regularly report its financial performance to its members and donors? (Please use a scale of 1 to 10, where 1 = No Reporting and 10 = Detailed Periodic Reporting Beyond the Annual Report) _____ (#)
- 2 d) Are your agency's financial statements either "audited" or "prepared under review engagement"? (circle one) Yes / No

SECTION THREE: INCOME INDEPENDENCE (Source of information: Whole agency)

- 3 a) How many separate sources of revenue does your agency maintain? _____ (#)
- 3 b) What percentage of your operating revenue is provided by your single largest revenue source? _____ (%)
- 3 c) What percentage of your agency's operating revenue is provided by government sources? Include all levels of government. _____ (%)

SECTION FOUR: STRATEGIC MANAGEMENT (Source of information: Whole agency)

- 4 a) To what extent does your organization have a clear, concise statement of purpose that is on display and known by staff, volunteers, and clients – commonly referred to as a mission statement? (Please use a scale of 1 to 10, where 1 = No Statement and 10 = Concise, Written Statement) _____ (#)
- 4 b) Does your agency establish, on the basis of the mission statement, annual goals and objectives for: (circle one)
- (i) programs? Yes / No
- (ii) staff? Yes / No
- (iii) volunteers? Yes / No
- 4 c) To what extent is there an annual process for updating the mission statement and consideration of its appropriateness in the face of changes in the environment, the clients, and/or community needs? (Please use a scale of 1 to 10, where 1 = No Process and 10 = Frequent Structured Process) _____ (#)
- 4 d) What percentage of the program staff, managers and senior support staff are involved in the strategic management process? _____ (%)
- 4 e) To what extent are the staff mentioned in question 4 d) involved in the setting of goals and objectives? (Please use a scale of 1 to 10, where 1 = No Participation and 10 = High Participation) _____ (#)

SECTION FIVE: BOARD GOVERNANCE (Source of information: Whole agency)

- 5 a) What percentage of voting board members are also employed staff members? _____ (%)
- 5 b) What percentage of board members are also financial contributors? _____ (%)
- 5 c) How many times did the Board of Directors (or the executive committee) meet last year? _____ (#)
- 5 d) As a percentage, what was the average attendance of Board members at Board of Director meetings (or the executive committee) throughout the last year? _____ (%)
- 5 e) To what extent does your organization have policy guidelines in place to avoid conflicts of interest involving Board or staff members? (Please use a scale of 1 to 10, where 1 = No Policy Guidelines and 10 = Formal, Documented Policy Guidelines) _____ (#)

SECTION SIX: VOLUNTEERS (Source of information: Whole agency)

- 6 a) How many full-time-equivalent (F.T.E.) volunteers does your agency have in total? _____ (#)
- 6 b) To what extent are volunteers screened for eligibility? (Please use a scale of 1 to 10, where 1 = No Screening and 10 = Detailed Screening Process Including Reference Checks) _____ (#)
- 6 c) To what extent are volunteers assessed for job allocation? (Please use a scale of 1 to 10, where 1 = No Assessment and 10 = Detailed Initial Assessment and Ongoing Reassessment) _____ (#)
- 6 d) To what extent does your agency have a volunteer training program in place? (Please use a scale of 1 to 10, where 1 = No Training Program and 10 = Ongoing Training Program) _____ (#)
- 6 e) To what extent does your agency have a regular evaluation process in place for volunteers? (Please use a scale of 1 to 10, where 1 = No Evaluation Process and 10 = Detailed Periodic Evaluation Process, Beyond an Annual Evaluation Process) _____ (#)
- 6 f) To what extent does your agency attempt to recruit **eligible** past adult clientele for volunteering? (Please use a scale of 1 to 10, where 1 = No Attempt Made and 10 = Actively Promotes Volunteering to Eligible Past Clientele) _____ (#)
- 6 g) What percentage of your agency's **eligible** adult clientele is asked to volunteer? _____ (%)
- 6 h) What percentage of your agency's volunteers donated money or goods and services in-kind, in addition to the donation of their time? _____ (%)
- 6 i) What was the percentage turnover of your agency's volunteers in the previous year? _____ (%)

SECTION SEVEN: STAFF (Source of information: Whole agency)

- 7 a) How many full-time-equivalent (F.T.E.) staff does your agency have in total? _____ (#)
- 7 b) What percentage of the total full-time equivalent agency staff work primarily in the delivery of all programs? _____ (%)
- 7 c) What was the percentage turnover of your agency's **total staff** in the previous year? _____ (%)
- 7 d) What was the percentage turnover of your agency's **total program staff** in the previous year? _____ (%)
- 7 e) To what extent are staff regularly evaluated? (Please use a scale of 1 to 10, where 1 = No Evaluation Process and 10 = Detailed Periodic Evaluation Process, Beyond an Annual Evaluation Process) _____ (#)
- 7 f) To what extent does your agency maintain a formal hiring process? (Please use a scale of 1 to 10, where 1 = No Formal Hiring Process and 10 = Extensive, Formalized Hiring Process) _____ (#)
- 7 g) To what extent does your agency have a staff training program in place? (Please use a scale of 1 to 10, where 1 = No Training Program and 10 = Ongoing Training Program) _____ (#)

SECTION EIGHT: PROGRAM COST & PROFILE (Source of information: Program)

- 8 a) Name of the program being submitted for recognition: _____
- 8 b) Started in: _____ (Year)
- 8 c) Total cost of the program in the most recent year of operations: \$ _____
- 8 d) Number of full-time equivalent (F.T.E.) volunteers involved in the program: _____ (#)
- 8 e) Number of full-time equivalent (F.T.E.) staff involved in the program: _____ (#)
- 8 f) Number of clients served in the most recent year: _____ (#)
- 8 g) Number of program hours (or units for the Provision of Basic Necessities category) provided in the most recent year by the program applying for recognition: _____ (#)

SECTION NINE: INNOVATION (Source of information: Program; Questions 9e and 9f: Whole agency)

- 9 a) To what extent does the government provide or finance programs in your community similar to the one your agency offers? (Please use a scale of 1 to 10, where 1 = No Level of Provision or Financing and 10 = High Level of Parallel Provision or Financing) _____ (#)
- 9 b) To what extent does your agency's program resemble similar programs in your community? (Please use a scale of 1 to 10, where 1 = High Level of Similarity and 10 = Completely Unique) _____ (#)
- 9 c) What percentage of your current program was considered for restructuring or changing in the last year in order to improve the program? _____ (%)
- 9 d) What percentage of your current program was actually restructured or changed in the last year in order to improve the program? _____ (%)
- 9 e) To what extent (percentage) are computers and similar technology incorporated in your agency's operations? Include both non-program and program functions. _____ (%)

9 f) If an innovation in your program has been implemented in the last year, what approximate percentage of the whole agency's operations was affected by the innovation? _____ (%)

ONLY EDUCATIONAL SERVICES AND CHILD CARE PROVIDERS SHOULD ANSWER QUESTION 9 G). ALL OTHER AGENCIES PLEASE SKIP QUESTION 9 G).

9 g) To what extent (percentage) are computers and similar technology incorporated in the delivery of your agency's program? _____ (%)

SECTION TEN: OUTCOME MONITORING (Source of information: Program)

10a) To what extent has your organization defined the desired outcomes for the program? That is, what is it that the program is attempting to achieve? (Please use a scale of 1 to 10, where 1 = No Desired Outcomes Defined and 10 = Detailed Statement of Desired Outcomes) _____ (#)

10b) Given the program's desired outcomes, what type of assessment is made of the actual outcomes? (Please use a scale of 1 to 10, where 1 = No Measurement of Actual Outcomes and 10 = Formal, Documented Assessment) _____ (#)

10c) How often does your agency compare the actual outcomes [question 10 b)] against the desired outcomes [question 10 a)] of its programs? (Please use a scale of 1 to 10, where 1 = No Comparison Made and 10 = Regular, Documented Comparisons) _____ (#)

10d) To what extent is there a process for dealing with differences between the actual and the desired outcomes? (Please use a scale of 1 to 10, where 1 = No Process Defined and 10 = Documented Periodic Process, Including Action Plans, Monitoring, and Follow-Up) _____ (#)

10e) To what extent does your agency attempt to track past clientele for progress? (Please use a scale of 1 to 10, where 1 = No Attempt to Track Past Clientele and 10 = Regular, Documented Tracking) _____ (#)

SECTION ELEVEN: ACCESSIBILITY (Source of information: Program)

11a) To what extent does your agency restrict the receipt of services based on an assessment of need? (Please use a scale of 1 to 10, where 1 = No Restrictions and 10 = Restrictions Applied According to a Standardized Process of Needs Assessment) _____ (#)

EDUCATIONAL SERVICES AND CHILD CARE SERVICES SKIP QUESTIONS 11 B) TO 11 D).

11b) To what extent does your organization attempt to ascertain the cause of a client's present circumstance? (Please use a scale of 1 to 10, where 1 = No Inquiry and 10 = Detailed Inquiry) _____ (#)

11c) To what extent does your agency measure the frequency of usage by clients? (Please use a scale of 1 to 10, where 1 = None and 10 = High) _____ (#)

SERVICES FOR SENIORS AND SERVICES FOR PEOPLE WITH DISABILITIES SKIP QUESTION 11 D).

11d) To what extent does your agency restrict access to services based on the frequency of usage? (Please use a scale of 1 to 10, where 1 = No Restrictions and 10 = High) _____ (#)

11e) To what extent does your agency charge a fee for use of its services? (Please use a scale of 1 to 10, where 1 = No Fee Charged, 5 = Flat Fee Charged and 10 = Fee Charged that is Geared to Client Income (Adjusted Based on Income Levels)) _____ (#)

END OF APPLICATION

Jerry DeMarco, B.A. (Windsor), M.E.S. (York), LL.B. (Toronto), M.M. (McGill), R.P.P./MCIP, has worked for and served on the boards of Canadian non-profit organizations for the past 10 years. This article is based on his final paper for the McGill-McConnell Master of Management Program for National Voluntary Sector Leaders.