The Tao of Evaluation

Craig Russon
W. K. Kellogg Foundation
Craig.Russon@wkkf.org

Abstract

In his treatise on strategy entitled, A Book of Five Rings, famed samurai, Miyamoto Musashi, puts forward a nine-step method for learning Kendo or "the Way of the sword." Musashi’s words also speak to those of us who practice the evaluation profession. The same nine-step method can be applied for learning "the Way of evaluation."

Keywords

evaluation, tao, Japan, samurai, strategy

1. The Tao of Evaluation

"I have been many years training in the Way of strategy, called Ni Ten Ichi Ryu, and now I think I will explain it in writing for the first time." It is with these words that Miyamoto Musashi began his treatise on strategy entitled, A Book of Five Rings. Born in 1584, Musashi was destined to become one of Japan’s most famous warriors or samurai (Harris, 1974).

Samurai were easily distinguishable from the other classes of sixteenth century Japanese society (i.e., farmers, artisans, and merchants) by the two swords that they wore in their belts. The longer of the two was carried out of doors only, while the shorter was worn at all times. Dueling and tests of arms were common, with both real and practice swords. These took place within fencing halls (Dojo), before shrines, behind castle walls, and even in the streets. Duels were fought to the death or until one of the contestants was disabled.

By the age of 30, Musashi had fought and won more than 60 tests of arms by killing his opponents (Harris, 1974). Satisfied that he was invincible, Musashi turned to formulating his philosophy of “the Way of the sword" or Kendo (Harris, 1974). Over the next 30 years, Musashi refined his philosophy. A few weeks before his death in 1645, he committed his philosophy to paper.

A Book of Five Rings consists of the following sections: Ground, Water, Fire, Tradition (Wind), and Void. Ground contains an explanation of strategy and provides the context
for the other four books. Water contains instruction on the fundamentals of long sword fighting such as stance, gaze, grip, footwork, and attitudes. Fire contains an exposition of some 28 strategies for long sword fighting. (Many contemporary Japanese and American businessmen use these to guide their business practices.) Wind provides a critique of other schools of warfare and other traditions of strategy. Void instructs students to practice the Way until "the clouds of bewilderment clear away."

Musashi's words also speak to those of us who practice the evaluation profession. In Ground, he prescribes a nine-step method for learning his strategy. These nine steps, listed below, also apply to those who seek to learn "the Way of evaluation."

1. Do not think dishonestly.
2. The Way is in training.
3. Become acquainted with every art.
4. Know the Ways of all professions.
5. Distinguish between gain and loss in worldly matters.
6. Develop intuitive judgment and understanding for everything.
7. Perceive those things which cannot be seen.
8. Pay attention to even trifles.
9. Do nothing which is of no use. (pg. 49)

2. Do Not Think Dishonestly

In an effort to promote honesty, many regional and national evaluation organizations around the world have put forward guidelines for ethical conduct. The American Evaluation Association's (AEA) Guiding Principles (1994) is an early attempt to promote ethical behavior. The document contains five principles: (a) Systematic Inquiry; (b) Competence; (c) Integrity/Honesty; (d) Respect for People; and, (e) Responsibilities for General and Public Welfare. The Integrity/Honesty principle reads, "Evaluators ensure the honesty and integrity of the entire evaluation process." This is followed by seven statements that explain how the principle should influence every step of the evaluation process. The statements deal with contract negotiations, changes to the contract, conflict of interest, misrepresentations, misleading conclusions, and financial sources.

The Canadian Evaluation Association (CES) adopted its Guidelines for Ethical Conduct in 1996. The document contains three guidelines: (a) Competence; (b) Integrity; and (c) Accountability. The Integrity guideline reads, "Evaluators are to act with integrity in their relationships with all stakeholders." This is followed by four statements that operationalize the concept. These statements deal with accurate representation, conflict of interest, cultural sensitivity, and contract negotiations.

A very extensive effort to promote ethics was made by the Australasian Evaluation Society (AES). In 1997, AES endorsed its Guidelines for the Ethical Conduct of Evaluations. The preface of the document is devoted to the background, purpose, scope, audience, relationship to other guidelines, approach, and terminology. The document contains five principles that deal with critical phases of the evaluation process: (a) Commissioning and preparing for an evaluation; (b) Conducting an evaluation; and (c) Reporting the results of an evaluation. Each principle has several guidelines associated with it. Guidelines deal with contract negotiations, changes to the contract, potential risks, competence, conflict of interest, conducting one's self in an honorable manner, etc. Altogether there are 22 guidelines.
Even newly formed regional and national evaluation organizations have addressed this issue. The International Network of Program and Evaluation Specialists (Georgia, Moldova, Russia, and Ukraine) has adopted a set of five Program Evaluation Principles: (a) Program evaluation can be based only on accurate and reliable information; (b) Evaluation should be competent; (c) Program evaluators guarantee honesty and transparency throughout evaluation; (d) Program evaluators respect safety and dignity of people with whom they cooperate in the course of their professional activity; and (e) Evaluators have professional obligations determined by public interests and public welfare. The honesty and transparency principle is followed by three statements dealing with contract negotiations, communication with clients, and conflict of interests.

(While Musashi advocated for honesty and integrity, there are aspects of the ethical codes explained above that he did not practice. For example, it probably could not be said that Musashi respected the safety and dignity of the people with whom he competed during the course of his professional activity.)

3. The Way is in Training

Evaluation is generally acknowledged as an emerging profession with an expanding body of knowledge and skills that require special training (Mertens, 1994). In general, there are three types of evaluation training: short, medium, and long-term. An example of short-term training would be the informal, non-degree, professional development sessions that are often offered prior to conferences sponsored by regional and national evaluation organizations.

An example of medium-term training would be the World Bank’s International Programme for Development Evaluation Training (IPDET). In 2001, the Bank partnered with Carleton University in Canada to offer about 160 hours of intensive training in development evaluation. Eighty hours of core training focused on basic evaluation knowledge/skills and development evaluation issues. Afterwards, another 80 hours of electives were offered. In 2003, the World Bank and Carleton University will be offering their third IPDET session.

An example of long-term training would be degree programs offered by colleges and universities. A directory of such programs was published in 1994 (Altschuld, Engle, Cullen, Kim, and Macce). The authors are in the process of updating this directory, with special emphasis on degree programs at colleges and universities outside the United States of America. The World Bank has also developed an on-line directory of evaluation training programs. The URL is http://wbln0018.worldbank.org/oed/evalcat.nsf?OpenDatabase These are both valuable information resources.

Patton (2002) believes that training can be made more effective through ongoing “reflective practice.” Evaluators should take what they learn in training and put into practice. Then, they should follow-up on their practice and reflect on the results. This reflection should be the basis for improving the evaluators’ future work. Of course, this process could be done vicariously by reflecting upon the practice of other evaluators. This important way of deepening evaluator competence is very consistent with Musashi’s own self-training.

Some regional and national evaluation organizations have undertaken efforts to link training with systems of licensure, certification, and/or credentialization. For example, Altschuld
Craig Russon

(1999) described the Canadian Evaluation Society’s system that requires evaluators to successfully participate in four daylong workshops in order to ensure that credentialed evaluators have been trained at a basic level in key evaluation areas. Worthen (1999), pointed out the serious challenges that must be overcome. They include:

- Determining the basic approach certification should take
- Reaching agreement on a definition of evaluation and core competencies
- Constructing a professionally and legally defensible system
- Garnering support for the process

4. Become Acquainted with Every Art

The author of this paper served for a time as editor of Western Michigan University, The Evaluation Center’s Occasional Papers Series. During this period, Daniel Stufflebeam (1999) published an Occasional Paper entitled, *Foundational Models of 21st Century Program Evaluation*. In his publication, Stufflebeam critically appraised 22 evaluation approaches in order to determine which were worthy of continued application and further development. Twenty of the approaches were found to have legitimate value for evaluating programs. Stufflebeam (1999) divided these approaches into three categories: questions/methods-oriented approaches, improvement/accountability approaches, and social agenda/advocacy approaches. The approaches are listed by category below, along with the briefest of descriptions and a reference for further information. Evaluators following the injunction of Musashi to become acquainted with every art should seek to become familiar with as many of these approaches as possible.

**Questions/Methods-Oriented Evaluation Approaches**

- Objectives-Based Studies-seeks to determine if a program’s objectives have been achieved (Madaus & Stufflebeam, 1988)
- Accountability (particularly payment by results studies)-used to ascertain responsibility for good and bad outcomes (Lessinger, 1970)
- Objective Testing Programs-assesses student achievement as compared to norms and/or standards (Lindquist, 1951)
- Outcome Monitoring/Value-Added Assessment-recurrent objective testing in order to assess trends and partial out effects (Sanders & Horn, 1994)
- Experimental Studies-compares subjects randomly assigned into groups that receive different interventions to establish causality (Campbell & Stanley, 1966)
- Management Information Systems-supplies managers with information to guide programs (Deming, 1986)
- Benefit-Cost Analysis Approach-factors program costs discounted over time with accrued benefits (Tsang, 1997)
- Clarification Hearing-uses a judicial model to put a program on trial (Wolf, 1975)
- Case Study Evaluations-a focused, in-depth description and inductive analysis of a program (Stake, 1995)
- Criticism and Connoisseurship-expert appraisal akin to art criticism and wine connoisseurship (Eisner, 1983)
- Program Theory-Based Evaluation-uses a program’s theory as basis of evaluation. Operationalized through logic-modeling (Chen, 1990)
Mixed Methods Studies-eclectic mix of approaches (National Science Foundation, 1997)

Improvement/Accountability-Oriented Evaluation Approaches

- Decision/Accountability-Oriented Studies-provides managers information for making decisions that promote cost-effective services (Stufflebeam, 1967)
- Consumer-Oriented Studies-uses evaluation to determine worth and merit of products and services offered to consumers (Scriven, 1967)
- Accreditation/Certification Approach-institutions are studied for approval to perform specified functions

Social Agenda-Directed (Advocacy) Approaches

- Client-Centered Studies (or Responsive Evaluation)-evaluators interact continuously to respond to clients’ evaluation needs (Stake, 1983)
- Constructivist Evaluation-the evaluator weaves together the stakeholders’ various constructions of reality (Guba and Lincoln, 1989)
- Deliberative Democratic Evaluation-applies the principles of democracy to evaluation (House and Howe, 1998)
- Utilization-Focused Evaluation-users of evaluation are identified and a variety of methods employed to ensure that evaluation meets their needs (Patton, 1997)

5. Know the Ways of All Professions

There are two basic approaches to professional knowledge. One can know much about a single profession or one can know a little about many professions. Based upon this injunction, Musashi appears to favor the latter. This issue has been debated in the evaluation profession as well. Worthen and Sanders (1984) question whether evaluators should be trained as generalists, whose knowledge and skills are applicable to a broad range of professions. Or is it better to train them as specialists, with knowledge and skills in a single profession and less extensive evaluation training? To answer the question, they recommend taking six factors into account:

- Difficulty and uniqueness of the content-if the content is neither difficult nor unique, then a specialist may not be necessary.
- Reference groups and impartiality-the training and experience of specialists may influence them to exclude important stakeholder groups.
- Evaluation roles and tasks-do the job requirements call for the knowledge and skills of an evaluation generalist or a professional specialist?
- Scope of work-does the evaluator’s scope of work require that he or she deal with a variety of professional specializations?
- Implication for training-is it possible to provide the specialist with the training necessary to do a quality evaluation?
- Professional status and rewards-the more time specialists spend on evaluation, the less they are able to devote to their profession. Evaluators who limit their work to one profession will have low career security.

Based upon the above factors, Worthen and Sanders (1984) conclude that it is often better for evaluators to be trained as generalists, who know a little about many professions, rather than as specialists, who know much about one profession.
6. Distinguish Between Gain and Loss in Worldly Matters

Musashi believed that ‘worldly matters’ should not cloud the intuitive judgment of the swordsman, which is always the single source of his strength; for that, he should in daily practical dealings resort to simple calculation of gain and loss. Evaluators must do the exactly the same thing. We refer to this in technical terms as controlling bias. Bias is systematic error introduced into evaluation by selecting or encouraging one outcome or answer over others. According to Scriven (1975), there may be economic or social incentives that can cloud the judgment of evaluators. He offers four methodological approaches to controlling bias:

- Standardization or routinization of qualitative aspects of the procedures
- Upgrading the training procedures for evaluators, especially in the qualitative dimension
- Focus on the organizational and economic sources of bias external to the evaluator
- Take a team approach to generating alternative plans which can then be comparatively evaluated

In addition to these four approaches, Scriven (1975) also recommends the use of metaevaluation as a way of controlling bias. Metaevaluation is simply an evaluation of an evaluation. Often times the code of ethics discussed in a previous section of this paper or the standards to be discussed in a later section serve as the basis for such metaevaluation.

7. Develop Intuitive Judgment

The life of a samurai like Musashi depended in large measure on the judgments that he made. Likewise, a fundamental part of the evaluator’s job involves making judgments. Unlike the samurai, however, there is some difference of opinion among evaluators over who should make the judgments. Scriven (1974) believes that the evaluator is the best qualified to judge. Stake (1973), on the other hand, believes that the evaluator’s role is to process the judgments of others. Musashi’s injunction suggests that he would have been Scriven’s ideal type of evaluator.

The process of judging consists of three steps: (1) determining the criteria for making judgments, (2) collecting necessary data, (3) applying the criteria to the data. For Musashi, the criterion for making judgments was clear-cut: whether or not he could kill his opponent. It should also be noted that Musashi’s judgments had a high degree of finality, so in a sense he could be considered as a specialist in summative evaluation. For many contemporary evaluators the criteria are not so clear cut nor their judgments as final.

Evaluators typically make judgments of two kinds: merit and worth (Western Michigan University, The Evaluation Center, 2002). Merit is defined as “the excellence of an object as assessed by its intrinsic qualities or performance.” Worth is defined as “the value of an object in relationship to a purpose.” For example, an expensive silk kimono might have a lot of merit—but little worth if one is trying to stay warm outdoors during the winter in northern Hokkaido.

8. Perceive Those Things Which Cannot be Seen/Develop an Understanding for Everything

The injunction to perceive those things
which cannot be seen is influenced by one’s worldview. The traditional scientific worldview holds that there is a single, objective, tangible reality that one can see, know, and measure. The alternative worldview holds that there exist many, unseen, socially constructed realities (Guba and Lincoln, 1989). Evaluation is an iterative process of analysis leading to the emergence of a joint construction of that which is being evaluated.

The evaluator’s worldview will tend to influence the approach (see the section on knowing every art) that he or she adopts. For example, an evaluator who believes the traditional scientific worldview would tend to use approaches such as Accountability, Objective Testing Programs, Outcome Monitoring, Performance Testing, Experimental Studies, and Cost-Benefit Analysis. An evaluator who believes in the alternative worldview would tend to use approaches such as Case Study, Criticism and Connoisseurship, Client-Centered Studies, and Constructivist Evaluation.

According to Stake (1979), the products of evaluations using traditional scientific worldview are different from those of using the alternative worldview. The product of the former is explanation that emphasizes formalistic generalizations from sample to population. The product of the latter is understanding with emphasis on particulars and experiential knowing. Stake counsels us not to confuse explanation and understanding.

9. Pay Attention Even to Trifles

The injunction to pay attention even to trifles, for the evaluator, is an exhortation to produce quality work. One way that evaluators ensure the quality of their work is by conforming to professional standards. One of the most enduring and broadly accepted sets of evaluation standards is The Program Evaluation Standards (Joint Committee for Educational Evaluation, 1994). The Standards contain 30 principles that can help improve evaluation of programs, projects, and materials. The principles fall under four categories: Utility, Feasibility, Propriety, and Accuracy.

The Joint Committee has always maintained that The Standards are uniquely American and may not be appropriate for use in other countries without adaptation. Several regional and national evaluation organizations from around the world have undertaken projects to create standards suitable for use in their own cultural contexts. Many think that standards could provide a model to guide practice.

In an Occasional Paper entitled, The Program Evaluation Standards in International Settings, Widmer, Landert & Bachmann (2000) put forward the evaluation standards recommended by the Swiss Evaluation Society (SEVAL). The SEVAL standards are, in part, based on The Program Evaluation Standards. This was an historic first. These were the first evaluation standards to be developed in a country outside of the USA.

Patel & Russon (2000) described a 1998 evaluation workshop of the United Nations Children’s Fund (UNICEF) held in Nairobi that initiated a discussion of the appropriateness of The Program Evaluation Standards for use in African cultures. During the discussion, it appeared that, in some cases, The Standards were culturally inappropriate or misleading. The results of the discussion were presented at the Inaugural Conference of the African Evaluation Association (AfrEA) as a draft document.
In December, 2000, the board of the Australasian Evaluation Society voted to authorize the development of Australian and New Zealand standards of evaluation (Fraser, 2001). A special committee is developing the standards for endorsement by the membership. These standards may, ultimately, be registered with the standards authorities of both countries (Fraser, 2001).

10. Do Nothing Which is of No Use

Utilization is widely seen as the sine qua non of evaluation. It is often said that it is better not to conduct an evaluation at all, than to conduct an evaluation that is not used. This is because an evaluation that is not used wastes money that could have been devoted to programming. The professional literature historically has recognized three main types of evaluation utilization: instrumental, conceptual, and symbolic (Kirkhart, 2000).

Instrumental utilization occurs when major stakeholders immediately implement the recommendations from an evaluation report. Conceptual utilization occurs when an evaluation report serves to alter the way that stakeholders think about the object of the evaluation. Symbolic utilization occurs when the evaluation is used for political purposes, for example, to muster support for the continuation of a program.

Utilization is such an important matter that an entire approach has been devoted to it. Michael Quinn Patton’s Utilization-Focused Evaluation is a widely practiced evaluation approach. According to Patton, “Utilization-focused program evaluation . . . is evaluation done for and with specific, intended primary users for specific, intended uses” (p. 23). Patton’s writings have recently been translated into Japanese and published by G. Pam Communications (2000).

11. Conclusion

To conclude this article, I wish to paraphrase Musashi. It is important to start a profession in evaluation by setting these broad principles in your heart. With your spirit settled, you must accumulate practice in evaluation day by day, year by year. As you grow in the art and the science of evaluation, eventually you will come to think of things in a wide sense and, taking the void as the Way, you will see the Way as Void. “In the void is virtue, and no evil. Wisdom has existence, principle has existence, the Way has existence, spirit is nothingness” (p. 95). Study this well.

Notes

1. This article represents the views of the author and not necessarily those of the W. K. Kellogg Foundation.
2. The author wishes to thank Michael Quinn Patton and Blaine Worthen for their valuable feedback on an early draft of this manuscript.

References


Patton, M. Q. (2002). *Personal communication*.


Japan: G. Pam Communications.
Stake, R. (1979). Naturalistic study of such things as science teaching. Paper delivered at the May 17 meeting of the North Carolina Association for Research in Education.

(2003.7.7受理)