*Journal of MultiDisciplinary Evaluation*
Volume 15, Issue 33, 2019

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

# Big Shoes to Fill:
## An Evaluation Journey in the Footsteps of Daniel L. Stufflebeam

Sherrie-Ann Camilli
*Athabasca University Faculty of Health Disciplines (Post Masters Student)*
*Laurentian University School of Nursing (Clinical Advisor)*

**Background:** Evaluation has evolved remarkably since the early 1960s, largely due to the innovative contributions of Daniel Stufflebeam and his colleagues. As a pioneer of evaluation methods, some of the notable achievements arising from Stufflebeam's work include the *context-input-process-product* (CIPP) model, evaluation standards, and evaluation checklists.

**Purpose:** The purpose of this paper is to explore Daniel Stufflebeam's journey beginning with the early days of evaluation through to his retirement and unfortunate passing at 80 years old in 2017. Key features of the CIPP model are considered within a context of other popular models for comparison with the goal of finding relevance for use of CIPP evaluation in education settings.

**Setting:** Not applicable.

**Intervention:** Not applicable.

**Research Design:** Literature review.

**Data Collection and Analysis:** Not applicable.

**Findings:** Stufflebeam's CIPP model and evaluation standards remain prominent in evaluation practices and his legacy will lay the foundation for future evaluators through continued professional development.

***Keywords:*** *Stufflebeam; CIPP model; standards; checklists; participatory evaluation.*

Over the past 40 years, evaluation of programs and services has evolved in response to a climate of constant change, coupled with the corresponding need to validate program effectiveness (Patton, 2017). Early evaluation methods usually involved research as the preferred approach to determine a program's effectiveness; however, academic research often ignored the utility of evaluation in favor of experimental design and quantitative data analysis that were highly technical and thus, poorly understood by decision makers (Patton, 2017). While validity, reliability, and measurability were regarded as more credible evaluation elements, it was becoming apparent that the use problem was not being addressed by greater methodological rigor since the findings were discounted as being overly complex and irrelevant (Patton, 2017).

It was during the 1960s that an evaluation pioneer, Daniel Stufflebeam, came together with his colleagues to revolutionize a model of evaluation and set of professional standards that would not only respond to the use problem, but would eventually become the gold standard of evaluation for the purpose of decision making (Alkin & Christie, 2004). Stufflebeam's accomplishments earned international recognition and endorsement from respected organizations (Patton, 2017). In fact, he even caught the attention of the World Bank who commissioned his expertise to assist them in improving their evaluation indicators used to assess educational programs (Russon & Russon, 2018).

Stufflebeam retired in his early 70s (The Oral History Project Team, 2008) and sadly, passed away in 2017 at 80 years old (Russon & Russon, 2018). His legacy remains a source of inspiration for emerging evaluation professionals, though his shoes will be very difficult to fill. This paper is a tribute to the evaluation trek by Daniel Stufflebeam, a recognition of his evaluation model and standards as souvenirs from his journey, left for the scholars of the future.

## A Stroll Through History

Daniel Leroy Stufflebeam was born in Waverly, Iowa in 1936, beginning his academic career as a substitute teacher at many Chicago public schools in 1961 (Patton, 2017; Russon & Russon, 2018; Stufflebeam, 2016). During his years as a substitute teacher, his experience of working at multiple (more than 40) problematic schools later contributed to his conceptualization of the *context-input-process-product* (CIPP) model (Patton, 2017; Stufflebeam, 2016). Stufflebeam had identified a real need for there to be systematic approaches to evaluating schools amid a crisis that was clearly present in so many institutions, yet unknown to so many others including decision makers (Patton, 2017; The Oral History Project Team, 2008).

Stufflebeam served in the U.S. Army from 1960 to1968, while founding the Evaluation Centre at Ohio State University in 1963 and unveiling the CIPP model (Stufflebeam & Coryn, 2014). In later years, Stufflebeam developed personnel standards and achievement tests for the U.S. Marine Corps (Gullickson, 2018; The Oral History Project Team, 2008).

In 1973, Stufflebeam moved the Evaluation Center to Western Michigan University where he founded the Joint Committee on Standards for Education Evaluation where served as chair for 13 years, leading to the development of professional standards for program and employee evaluations (Russon & Russon, 2018; Stufflebeam & Coryn, 2014). The success and longevity of the Joint Committee was supported by the rules and guiding principles created at the committee's inception to govern work, foster inclusion, and resolve conflict (Gullickson, 2018).

He is recognized for many other contributions to evaluation including meta-evaluation (Patton, 2017), evaluation checklists (Stufflebeam, 2001), as well as graduate education in evaluation (The Oral History Project Team, 2008). Moreover, Stufflebeam was a supporter of continuous professional development (Stufflebeam, 2016), seeking out training whenever he was unable to solve a problem; if training was unavailable, he would develop the tools needed to deal with problems (The Oral History Project Team, 2008). One such tool was facilitation and interpersonal skills, which Patton (2017) deemed as essential in order to safely navigate the rough waters of personal barriers, objections, egos, power shifts, and political variances amid some of the greatest stakes in

outcomes. Perhaps the most significant tool developed by Stufflebeam to address an unsolved problem is the CIPP evaluation model.

## The CIPP Model

Dubbed as one of the fathers, or founders of modern day evaluation, Stufflebeam achieved international notoriety through his CIPP evaluation model, professional evaluation standards, and numerous scholarly publications in as many as eight languages (Gullickson, 2018; Patton, 2017; Russon & Russon, 2018; The Oral History Project Team, 2008). A comprehensive and interrelated framework, the CIPP evaluation model was an outcome of several years of work conducted by Dan Stufflebeam in collaboration with educational organizations to improve their evaluation and decision making processes (Russon & Russon, 2018; Stufflebeam & Coryn, 2014). The model's primary orientation is to support quality improvement through a dynamic assessment strategy, designed to support decision formation and foster accountability (Stufflebeam & Zhang, 2017) rather than to prove something specific about the program (Frye & Hemmer, 2012). CIPP was designed to be simple enough for the layperson to use as a method of making judgments about worth, value, merit, feasibility, efficiency, safety, and propriety when conducting an evaluation (Patton, 2017; Stufflebeam & Zhang, 2017).

Key components of the CIPP model are the systematic use of four complimentary types of evaluation – *context, input, process, products* – to address each of the following four elements:

1. what needs to be done;
2. how it should be done;
3. if it is being done;
4. whether or not it succeeded (Stufflebeam & Zhang, 2017).

Context evaluations are conducted to assess current needs, problems, assets, and opportunities (Russon & Russon, 2018). These types of evaluations are consistent with a needs assessment where the degree of a problem is identified, along with relevant actions to address the problem (Rossi, Lipsey, & Henry, 2019). The extent to which a new program can reasonably be evaluated is also revealed through context evaluations (Frye & Hemmer, 2012). Context evaluation is a strategy that a new leader might employ when starting in a new organization or department. Input evaluations aid in the exploration of alternative approaches, proposals, and programs, (Russon & Russon, 2018). This may be ideal for when resource allocation and justification is required (Frye & Hemmer, 2012). Process evaluations are utilized when a program's implementation is being assessed (Frye & Hemmer, 2012), for example, program evaluation in education. The process evaluation model is the least disruptive to deliver while a program is in operation and provides support for accountability and continuous quality improvement (CQI) (Frye & Hemmer, 2012). Finally, the CIPP model's product evaluation serves as an assessment of program outcomes (intended, unintended, positive, negative) and the degree to which targets were met (Stufflebeam & Zhang, 2017).

Each of the four evaluations can be applied formatively from a proactive application, as well as summatively from a retrospective angle (Stufflebeam & Zhang, 2017). The flexibility afforded by the CIPP model supports its use as a full spectrum tool for evaluating programs from the point of conception to full implementation, particularly when undergoing constant change. The flexibility of the CIPP model suggests a Cinderella fit for evaluating educational programs.

## Educational Evaluation Standards

Under Stufflebeam's leadership, the Joint Committee on Standards of Educational Evaluation was established in 1975 to eventually publish the educational evaluation standards in 1981, which would later become the new multidisciplinary professional evaluation standards in 1993 (Russon & Russon, 2018). The Joint Committee was formed through a collaboration of three notable and distinguished organizations: the American Educational Research Association, the National Council on Measurement in Education, and the American Psychological Association (Gullickson, 2018). At the time of creation, the standards had four dimensions

in a sequential order respective to chronological steps in the process: *utility, feasibility, propriety*, and *accuracy* (Patton, 2017). Stufflebeam's suggested rationale for the order was that evaluations should not be conducted if they are not first of all useful to a particular audience, then feasible in terms of cost and operational ability, ethical next, and finally, accurate (Patton, 2017). In 2010, 17 years after the creation of the standards, a fifth standard was added on evaluation *accountability* (Patton, 2017). The set of five evaluation standards prompts a new perspective of responsibility by traditional academic researchers whom are otherwise detached from accountability and ownership of evaluation activities while engaged in traditional research (Patton, 2017).

## Theorist Comparison

While there is no consensus on a 'one size fits all' approach to evaluation, the CIPP model may be very close in terms of its usability. The evaluation theory tree presented by Alkin and Christie (2004) classifies theorists into three major branches, or groups: *use*, *methods*, and *valuing*. Stufflebeam's model is captured in the *use* branch and differs from the other two branches by the primary concern about the manner in which evaluation information is used and by whom (Alkin & Christie, 2004). In the *use* branch, Stufflebeam shares this domain with other theorists such as Fetterman, Patton, and Wholey (Alkin & Christie, 2004). The *methods* branch (theorists such as Rossi and Campbell) focuses on generalizations arising out of data analysis for the purpose of new knowledge and the *valuing* branch (i.e., Scriven's work) emphasizes the importance of data value as the most critical and essential part of evaluation work (Alkin & Christie, 2004).

Frye and Hemmer (2012) explore four common evaluation models, including the CIPP model in a guide produced by An International Association for Medical Education (AMEE). A comparison of these evaluation models distinguishes the CIPP model from the others in some ways, yet also reveals some similarities. For example, Kirkpatrick's model focuses on learner outcomes in training, similar to the outcomes element of product studies in CIPP (Frye & Hemmer, 2012). However, Kirkpatrick's model lacks elements of the context, input, and process studies such as considering variables that affect learning, limiting the findings of why learning did or did not occur (Frye & Hemmer, 2012). Furthermore, a comparison with CIPP and experimental/quasi-experimental models has similar limitations where only the product is evaluated with stringent focus on a narrow scope of indicators (Frye & Hemmer, 2012).

Conversely, the CIPP model shares some alignment with the logic model, in that they both share similar element taxonomy (i.e., input, activities, outputs/outcomes). However, the CIPP model is not constrained by the linear relationships that the logic model depends upon, making it superior in terms of flexibility (Frye & Hemmer, 2012) and thus, more appropriate for complex evaluation needs.

## Critics and Challenges

Controversy and debate is inherent to theoretical discussion, and Stufflebeam encountered a variety of naysayers throughout his career. Patton (2017) describes an early example of how Stufflebeam handled critics and skepticism when promoting participatory evaluation. While Stufflebeam was leading the development of evaluation standards for education in the 1970s, a committee was struck which consisted almost exclusively of "busy, opinionated, and often cantankerous" academics (Patton, 2017, p. 2). Egon Guba expressed his concern that those who were currently present at the table were not suitable to set standards for evaluations of front line education staff, arguing that those staff should also have a position at the table to voice their concerns and opinions (Patton, 2017). Stufflebeam's fear was that this diverse representation would perpetuate debate without the parties ever reaching agreement (Patton, 2017).

Despite Stufflebeam's concerns, the committee expanded to include stakeholders from both groups and the composition became evenly represented with eight on each side, and Stufflebeam as the chair or essentially, the mediator (Patton, 2017). As expected, the debate began at the first meeting, which

prompted a colleague to propose a set of rules to ensure that all voices could be heard before decisions were made, as well as giving each member the power to deny any standard (Patton, 2017). Over time, the committee members learned to value each other's point of view to the extent that a set of standards for education evaluation that everyone respected was successfully established, and veto power was never exercised (Patton, 2017). Stufflebeam learned that evaluations should reflect a participatory approach by incorporating the needs of intended users, addressing their queries, and assisting them to make constructive use of the findings (Patton, 2017).

In 2010, Stufflebeam found himself in a separate debate over the sequencing of the standards of evaluation, where it was argued by his colleagues that *accuracy* should precede *utility* in the order of mention (Patton, 2017). However, Stufflebeam felt very strongly that placing *accuracy* ahead of *utility* would revert evaluation philosophy back to the 1960s and 70s when evaluations resulted in "technically elegant but irrelevant evaluation findings" (Patton, 2017, p. 4). He further offered the reasoning that it does not make sense to plan methodology of evaluations before determining if it is worth doing in the first place (Patton, 2017). Following extensive debate with over 400 stakeholders, the standards were revised to include a fifth standard (*accountability*) and the original sequence was retained (Patton, 2017).

Stufflebeam's participatory approach has been challenged by the emergence of empowerment evaluation (Stufflebeam, 1994). David Fetterman et al. argue that the goal of the participatory approach is to address inequities and liberate those who are otherwise oppressed (Fetterman, Rodriguez-Campos, Wandersman, & O'Sullivan, 2014). They further suggest that the control of the evaluation originates with the evaluator and is shared jointly with the participants (Fetterman et al., 2014). Conversely, empowerment evaluation views all participants as having full control of the evaluation (Fetterman et al., 2014). However, in an earlier article, Stufflebeam posited empowerment evaluation as having great potential for conflict due to the variation in perceived value and objectivity, as well as skirting along an edge of unethical, even unscrupulous use of evaluation (Stufflebeam, 1994). Yet, 20 years later, Fetterman and his colleagues continue to advocate for empowerment evaluation as an approach to even distribution of control among the participants in a flattened hierarchy (Fetterman et al., 2014).

## Utility in Educational Evaluation

In current evaluation practice, participatory evaluation is commonly accepted in order to promote trust, reduce fear of evaluations, increase stakeholder support and advocacy, as well as enhance usability and credibility of findings (Russon & Russon, 2018). Stufflebeam's model of evaluation promotes participatory evaluation in the context of education, which supports the relevant application of the CIPP model in higher education evaluation. The CIPP model is flexible, expansive, and can be adopted for use in medical education by those who exercise careful planning (Frye & Hemmer, 2012). Since it can be applied for both formative and summative assessments, the CIPP model can cover a broad scope of a program's lifespan to include current and retrospective analyses (Frye & Hemmer, 2012).

Evaluation checklists were designed by Stufflebeam beginning in the 1960s and in collaboration with several colleagues, he assembled a repository of 10 distinct checklists by the early new millennium (Stufflebeam, 2001). The checklists were deemed to have a wide scope of application to guide program, personnel, and product evaluation including metaevaluation, lending themselves to effective evaluation in educational settings (Stufflebeam, 2001). In alignment with the fifth installment of the CIPP model, the checklists have been expanded to separate evaluator and stakeholder activities recommended for evaluating each of the four subdomains of *product* evaluation: *impact, effectiveness, sustainability*, and *transportability* (Stufflebeam, 2007). While several of the inaugural checklists have been retired, the Evaluation Center at Western Michigan University (2019) currently boasts a library of more than 25 checklists, some offered in three

languages, making this an excellent resource for education evaluators.

Within a similar context of teaching-learning, Stufflebeam's CIPP model and checklists can be applied to an evaluation plan of health promotion education activities because CIPP components can accommodate the constant change that encompasses most educational programs (Frye & Hemmer, 2012). Since the outcome of such an evaluation plan may include measuring the efficacy of patient and family health teaching on health decisions and compliance, a user-inclusive approach would support buy-in and confidence in the evaluation.

## Conclusion and Reflection for the Future

Evaluation is a shared responsibility, thus, it is important to enlist the participation of all of those with a vision and passion for quality (Patton, 2017). Stufflebeam envisioned a unified evaluation model, which would alleviate stakeholder confusion resulting from alternative approaches and buy-in would be enhanced (Russon & Russon, 2018). His concern over utility in lieu of accuracy stemmed from the worry that evaluations would yield a large report that would not be usable due to its technical methodology (Patton, 2017). Overall, Stufflebeam's participatory evaluation work provided the scaffolding for what has now translated into best practice in evaluation. With a growing social accountability mandate among institutions, particularly education, the CIPP model favors usability and accountability in evaluative practice.

Although this distinguished scholar's evaluation journey has moved on to a spiritual path, Stufflebeam's legacy continues to be very much alive in today's evaluation society and will continue to shape the future of professional evaluators. His emphasis on the importance of meta-evaluation based upon the evaluation standards highlights the professional responsibility of evaluators to ensure that they are critiquing their own work in a formative and summative manner (Patton, 2017; Stufflebeam & Zhang, 2017).

Finally, the skills needed for facilitation, evaluation, and metaevaluation can be acquired through emerging continuous professional development opportunities to continue along the groomed path bestowed to us by Daniel Stufflebeam. Perhaps in his absence, scholars will continue on this route or embark on a new path through their own innovation. Although Stufflebeam's shoes will be difficult to fill, there is nothing stopping the evaluators of the future from trying out a unique, new pair of their own.

## References

Alkin, M., & Christie, C. (2004). An evaluation theory tree. In Alkin, M. C., *Evaluation roots* (pp. 13-65). Thousand Oaks, CA: SAGE Publications. https://dx.doi.org/10.4135/9781412984157.n2

Fetterman, D., Rodríguez-Campos, L., Wandersman, A., & O'Sullivan, R. G. (2014). Collaborative, participatory, and empowerment evaluation: Building a strong conceptual foundation for stakeholder involvement approaches to evaluation (A Response to Cousins, Whitmore, and Shulha, 2013). *American Journal of Evaluation, 35*(1), 144–148. https://doi.org/10.1177/1098214013509875

Frye, A.W., & Hemmer, P. A. (2012) Program evaluation models and related theories: AMEE Guide NO 67. *Medical Teacher, 34*, e288-e289. http://dx.doi.org/10.3109/0142159x.2012.668637

Gullickson, A. R. (2018). In memory of Daniel L. Stufflebeam (1936–2017). *American Journal of Evaluation*, *39*(1), 146–149. https://doi.org/10.1177/1098214017739151

Patton, M. (2017). Standards of excellence from a leader of excellence: Honoring our professional heritage and remembering Daniel L. Stufflebeam. *Journal of MultiDisciplinary Evaluation, 13*(29), 1-7. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/492

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach*

(8th ed.). Thousand Oaks, CA: SAGE Publications.

Russon, C., & Russon, K. (2018). Daniel Stufflebeam and the World Bank. *Evaluation and Program Planning, 67*, 184-188. https://doi.org/10.1016/j.evalprogplan.2017.12.011

Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice, 13*(3), 321-338. https://doi.org/10.1016/0886-1633(94)90027-2

Stufflebeam, D. L. (2001). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation, 22*(1), 71-79. https://doi.org/10.1016/S1098-2140(01)00119-9

Stufflebeam, D. L. (2007). CIPP evaluation model checklist (2nd ed.). Retrieved from https://wmich.edu/sites/default/files/attachments/u350/2014/cippchecklist_mar07.pdf

Stufflebeam, D. L. (2016). Factors that influenced my conduct of evaluations and evaluation training programs. *New Directions for Evaluation, 150,* 41-49. https://doi.org/10.1002/ev.20188

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications*. Somerset: John Wiley & Sons, Incorporated. Retrieved from https://0-ebookcentral-proquest-com.aupac.lib.athabascau.ca

Stufflebeam, D., & Zhang, G. (2017). *The CIPP Evaluation Model: How to evaluate for improvement and accountability.* New York, NY: The Guilford Press. Retrieved from https://books.google.ca/books?id=5lTLDAAAQBAJ&printsec=copyright&redir_esc=y#v=onepage&q&f=false

The Oral History of Evaluation. (2008). The professional development of Daniel L. Stufflebeam. *American Journal of Evaluation*, *29*(4), 555–571. https://doi.org/10.1177/1098214008321995

Western Michigan University. (2019). Evaluation checklists [Document Repository]. Retrieved from https://wmich.edu/evaluation/checklists