International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 12, Issue 8, 2020

Models of Program Evaluation for Teacher Education Training

Ibrahim Muhammad Alhaji[a], Wun Thiam Yew[b], Nordin Abd Razak[c], [a]Department of Science Education, Sule Lamido University Kafin Hausa,Jigawa State, Nigeria, [b,c]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia,

Little emphasis has been given to evaluation research of mathematics teacher education training in Nigeria, specifically on the effectiveness of training in achieving national objectives of teacher education that groom quality preservice mathematics teachers who will shoulder the responsibility of providing quality education at secondary school level. This conceptual paper discusses the theory of program evaluation and explains when, where and how some methods apply in program evaluation and others not. The paper also explains complexity theory as the most common and widely used theory in evaluation of educational programs for different and interrelated components. The function, strengths and weaknesses of different evaluation models that are used in education were discussed and CIPO evaluation model was identified as the most appropriate and relevant evaluation model for mathematics teacher education training, as the model allow evaluators to identify the relationship between the Context, Input, Process, and Outcomes of any educational program. A description of how CIPO evaluation model and its independent components can be applied in the process of evaluation of mathematics teacher education training was discussed. The study recommended CIPO evaluation model, as the model to be used in the evaluation process for mathematics teacher education training, as the model is identified as an effective, powerful evaluation model that is not hampered by linear relationship of components of the program and it has the power of evaluation of each independent component of the program.

Key words: *Program evaluation models, evaluation theory, complexity theory, teacher education, mathematics education.*

Introduction

A theory is a collection of concepts or ideas intended to explain a phenomenon based on a set of facts that have been confirmed through observation and experiment. Reynolds (1974) has defined theory as a way of searching for invariant laws, using a definition and axioms to deduce a testable proposition by describing a causal process that mediates the relationship. The application of theory in evaluation of a program is concerned with prescribing how and when various methods should be used in the practice of evaluation. The main concern with a theory of program evaluation is to guide evaluators to identify different elements of the program and how these elements relate with one another to form a product. It also helps evaluators to specifying a chain of causal relation between components of the model on what should be done or how it should be done in order to improve the situation of the program.

Shaddish et al. (1991) posited that program evaluation theories are solely concerned with three aspects in describing and justifying why certain evaluation practice leads to a particular kind of results. They described the aspect of evaluation as follows:

a. Clarification of activities, process and goals of evaluation.
b. Explanation of relationship among the evaluative components, process and the goals they facilitate.
c. Testing empirical proposition to identify and address those that conflict with research or critically appraised the knowledge about evaluation.

In view of the above aspect of evaluation, it's evidently clear that a good evaluator needs sound theory and must understand the strategies and techniques to apply in order to evaluate a program in a logical and systematically different but organized situation. Although different evaluation models have been widely used in the evaluation of educational programs, there have been limited studies about a specific theory for a particular model of evaluation.

Stufflebeam and Shinkfield (2007) identified complexity theory as the most appropriate theory of evaluation with higher influence on program context, input, process, output and outcomes. As this conceptual paper sets out to highlight and explain the gap in evaluation research about mathematics teacher education training, complexity theory was identified as the theory which should guide educational researchers and evaluators regarding how and when an evaluation model should be applied for the evaluation of teacher education training.

Complexity theory is the most common and widely used evaluation theory in the evaluation of an educational program that has different components that re interrelated. The theory explains how the model should be applied and allows evaluators to evaluate each component of the program separately. Sourcing the theory that is most appropriate and relevant to the

model of study under evaluation is the first step in determining the type of evaluation to embark on, as the theory explains on how and when a model will represent.

Complexity Theory

Complexity theory is concerned with evaluation of different independent components of a model that form a coherent system. The theory was developed to examine the complexity generated through interaction between different resources, values and goals (Alkin, 2013; Pawson & Tilley, 1997; Weiss, 1998). Due to the amount of attention that complexity theory has gained in recent years, in contrast to other fields of study, many researchers have described the theory as a field of inquiry with apparent influence among organizational development and evaluation practitioners (Garrits & Verweij, 2015). It is the most promising evaluation theory with higher evaluation impact on intervention equity (Blackman et al., 2011; Milton et al., 2011). Complexity theory was described "*as an alternate paradigm to guide development and management of public services, embedding values into the systems and drawing upon various sources of evidence to create an adaptive and flexible management style aware of local context*" (Walton, 2016. p. 214).

As complexity theory steadily gains interest within policy and evaluation practitioner circles, it's clearly evident that the theory has gained momentum in the field of evaluation as Walton (2016) sees the use of complexity theory as appropriate and relevant and perceives other methods, or specific components or the program not working. Stern et al. (2012) and Walton (2014) have noted that opportunities for using complexity theory in evaluation are inevitable when evaluators ask questions about why and how a program works and the question may align more with complexity approach.

The works of Paton (2011), Rogers (2011), and Funnell and Rogers (2011) have identified the use of complexity theory in evaluation when intervention is intended to adapt to outcomes as develop, with implications for evaluation scope, questions, method and governance. However, some researchers (Mowles, 2014; Shiell et al., 2008) argue that complexity theory focus on the system within which intervention are implemented. Walton (2014) also argued that complexity theory focus on initial conditions of the system and intervention, when evaluation and even simple intervention and its process should be included in defining boundaries of the evaluation that will at least partially determine whether an intervention is viewed as effective or not.

Research literature shows an increased use of complexity theory in evaluation studies due to high demand of complexity approach to evaluation (Carey et al., 2015; Gerrits & Verweij, 2015; Walton, 2014). The theory was developed to describe and explain systems or programs that are too complex to understand, through linear interaction of different elements across the

system (Mitchell, 2009). The nonlinear behavior of elements of the system or program that result in causes and effects is often independent; and as such, the pattern may emerge within a complex system or program (Johnson, 2007).

In many systems or programs, non-linear behavior of elements within the system exists independently, where the behavior of non-linearity at small scale is chaotic, while the non-linearity behavior of some elements within the system or program at larger scale, are discernible (Wood & Butt, 2014). Therefore, complexity theory will be appropriate and alternative to apply in describing the linearity or non-linear behavior of certain elements of a program. Wood and Butt (2014) have described linear behavior of elements within the system or program as core to the concept of emergence in complexity theory, as a process whereby interaction within the system or program leads to change in the development of new ideas and how they are working.

However, Wood and Butt (2014) have noted that complexity theory can be used as a metaphor to create an environment or condition in order to describe complex patterns or system development from simple interactions within the system and the change can be discerned. Consequently, complexity theory as a guide to program evaluation is used as a metaphor to create an environment or condition to describe when and how the context of the program (i.e. national policy of education, NUC guideline, program vision and mission, and intended objectives of the program) influences inputs of the program to provide adequate and relevant resources in achieving intended objectives of the training.

Models of Program Evaluation

The main idea behind evaluation of any program is to provide information to policy and decision makers who have responsibilities for the existing program and or are designing or proposing a new educational program. The information provided to policy and decision makers may be useful in making decisions concerning whether to propose a new program (placement evaluation), or to develop a program (need assessment), how to develop the existing program (formative evaluation) and whether to modify or to continue with the existing program (summative evaluation). These are the basic questions in program evaluation that each evaluator should ask when evaluating a program; answers to these questions could give evaluators a direction to determine the purpose of evaluating the program (Adelejen & Oguine, 1988).

Stufflebeam (2003) defined evaluation based on his CIPP model as: "*the process of obtaining judgmental information about the merit and worth of some object's goals, design, implementation, and outcomes. The information provides accountability reports, informs, dissemination decisions, and improves understanding of the involved phenomena*" (p. 7).

Program evaluation models have been classified by different authors into different categories. These authors classified models based on their areas of emphasis, purpose of program evaluation and their relevant to educational program and theories. Some classified models based on what they considered to be the major task of program evaluation. Among the authors who classified program evaluation models based on similarities and area of emphasis is Ubanya (1997). His classification is:

a. The judgment-oriented model
b. The decision-management model
c. The decision objective-oriented model

Saylar, Alexander and Lewis (1981) further classified models of program evaluation into five groups based on the area of purpose of program evaluation as follows:
a. Behavioral objective model
b. Decision making model
c. Goal-free evaluation model
d. Responsive model
e. Accreditation model.

CIPO Evaluation Model

Van Petegem et al. (2008) have identified Scheeren's (1990) CIPO evaluation model as the most appropriate and relevant evaluation model to educational program which identify the relationship between the context, input, process and outcomes in education within a certain context.

Scheeren's (2000) has described CIPO model of evaluation as a school effectiveness model that has been widely used in school effectiveness research. The model takes into account the relationship and intermediate causal effects on the association between indicators within the inputs into the system, the process through the system, and the output (Scheerens, 1992). These characteristics of CIPO model makes it as the most appropriate and relevant evaluation model for monitoring school effectiveness (Scheeren's, 1992). The function of each element of CIPO model of evaluation has been described by Scheerens (1990) as follows:

Context Element: The context element of CIPO is concerned with environment of the school, needs for which the program was designed, goals and objectives of the program. The context element has direct effects on the process indicators. These indicators could be measured based on the current direction of the program in comparison with initial or original guidelines of the program and intended objectives or needs. Different instruments are used in

collection of information on context evaluation. This depends on the demand of the situation of collecting data. The following are some of the instruments (Frye & Hemmer, 2012):

i. Document analysis
ii. Interview
iii. Surveys
iv. Demographical data analysis
v. Focus group
vi. Records analysis (e.g., learner performance data, test result)

Input Element: This component is concerned with education resources that could meet and satisfy the needs of the program which include, student characteristics, student initial experience and education, teacher qualification, curriculum content, learning environment, instruction material and so on. These indictors could be measured by adequacy and availability of input based on the guideline of the program, needs and stated objectives of the program. In assessing input component of the program, different approaches are used in the cause of data collection and some of instruments are (Frye & Hemmer, 2012):

i. Consulting experts
ii. Literature review
iii. Visiting exemplary program
iv. Inviting proposal from persons interested in addressing the identified needs

Process Element: This component of CIPO is concerned with the implementation process of the program in achieving intended objectives by inputs into practice, in line with basic guidelines of the vision and mission of the program. Scheerens et al. (2003) has identified the following as the basic concern of the process evaluation element of CIPO:

a. Educational leadership refers to the amount of time spent on educational matters as well as appraisal of educators and the amount of time dedicated to instructional matters during staff meetings.
b. Consensus and cooperative planning of educators are articulated in terms of the type and frequency of meetings, nature of cooperation as well as importance attributed to cooperation.
c. Quality of curricula is seen as the cornerstone of the most important function of education. Quality of curricula includes indicating clear targets, formal structure, and the degree to which the specified content is covered.
d. Orderly environment refers to the school climate in which there is good discipline and the learner behavior is considered acceptable. (p. 68)

The indicators of this component can be measured by examining the extent to which the input of the program provided for and satisfied the needs of the program in line with basic guidelines in achieving intended objectives of the program. Different approaches are employed in the case of data collection in process evaluation. Some of the instruments are as follows (Frye & Hemmer, 2012):

a. Observation
b. Document review
c. Participant interview

Outcome Element: This component is concern with student achievement in line with intended objectives of the program. The indicator of this component is measured by comparing program outcomes with intended objectives of the program. Different techniques are employed in outcomes evaluation in the cause of data collection. Some of them are (Frye & Hemmer, 2012):

i. Surveys
ii. Assessment of achievement of program objectives
iii. Group interviews about the full range of program outcomes
iv. Stakeholders judgment of the project program
v. Case studies of selected participant experience
vi. Comparative studies of outcomes with those of similar programs
vii. Participant report of project effects.

The main concern with CIPO model of evaluation is to explore possible differences in student achievement, whether these differences are related to context, input or process of the program (Reynolds, Sammons, De Fraine, Van Damme, Towsend, Teddie, & Stringfield, 2014). The model has been identified by Schereen (2015) as a school system model that can be applied to any level of education. The quality and effectiveness of school system are assessed through CIPO model (Cuyvers, 2002). For quality evaluation of any educational program, Scheeren (2005) has argue that education must be seen as a production process whereby context gives input and process result into output.

Creemers and Reezigt (1997) have identified some of the characteristics of school effectiveness model in educational research as follows:

a. Identification of the needs to be changed in schools in order to become more effective
b. Program evaluation
c. Searching for stable causes and effects
d. Searching for objective knowledge

e. Focus on student learning / classroom level

While some researchers have criticized the school effectiveness evaluation model in educational research because of its inability to pay more attention to the overall goals of school and education, it is mainly centred on the assumption of cause and effects and it neglects the diversity of school aims (Van Petegem, 1997). The model focuses on the performance tests that reflect a narrow knowledge domain in which only cognitive skills are measure; the model emphasises examination results and test scores as key performance indicators in drawing a conclusion of whether a school is good or not (Morley & Rassool, 1999).

However, Frye and Hemmer (2012) have identified four evaluation approaches: experimental/ quasi experimental approach to evaluation; Kirkpatrick's approach; Logic model and the Context/Input/Process/Product (CIPP) models. They describe them as the most common approaches used in educational evaluation, providing a clear picture of the program and informed evaluators about the situation of the program. Table 1 below illustrates the comparison between the four models based on their influence on educational program and independent studies on each component of the program.

Table 1: Comparison of evaluation models based on appropriateness and relevant to education program

CIPP Model	Context studies	Input studies	Process studies	Product studies	
Logic Model		Input	Activities	Output	Outcomes
Linear related outcomes					
Kirkpatrick's Model		element →	element→	element	linear
relationship of intended					
Experimental Model					program outcomes
to program element					

Experimental/Quasi Experimental Approach

This approach to educational evaluation has been described as one of the earliest used to evaluate educational in the mid 1960's (Stufflebeam & Shinkfield, 2007). The experimental approaches focuses on a linear relationship between program elements and desired program outcomes by isolating independent approaches of the program and its' study. The experimental approach was enormously useful in advancing biological science over two

centuries (Stufflebeam & Shinkfield, 2007). However, the experimental approach to evaluation was proven less useful in the educational field due to the nature of complexity of the educational environment program, in which true experiments with high control are very difficult to implement in the educational field which is as complex as in medical education (Fryes & Hemmer, 2012).

Kirkpatrick's Four level Evaluation Model

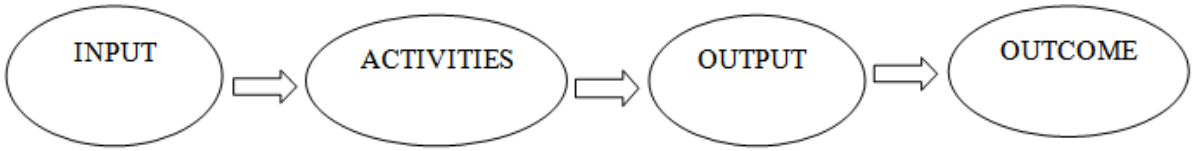
This approach is one of the most commonly used approaches, with widespread popularity for evaluation of learner outcomes in training program (Kirkpatrick, 1996). The Kirkpatrick's approach has differentiated itself with other evaluation models because of its power to clearly describe training outcomes beyond learner satisfaction. The Kirkpatrick approach has four hierarchical levels of assessing program outcomes as identified by its founder. The hierarchies that each level assesses are as follows:

a. Assess learners' satisfaction or reaction to the program.
b. Measures of learning attributed to the program (e.g. knowledge gained, skills improved and attitudes changed).
c. Changes in learner behavior in the context for which they are trained.
d. Program final result in its longer context.

Logic Model of Evaluation

The logic model of program evaluation has been identified as one of the evaluation models with strong influence in both educational planning and evaluation approach (Fechting, 2007). The model focuses on the relationship between program components and relationship between program components and context of the program. Figure 1 below depicts the clear picture of the model and linear relationship of the components.

Figure 1. Linear relationship of logic components.



The structure of the logic model shares some characteristics with Stufflebeam's CIPP model of evaluation, but the logic model focuses on the change process and the system within which the change occurs. The simplicity of logic model is attractive to many evaluators as the model clearly explains the relationship between each component of the model and outcomes. The

functions of each component of the model have been explained by Frechting (2007) as follows:

i. **Input**: The input component of logic model is concerned with all resources expected to be included in the program which comprises facilities, staff, students, funding resources and relevant facilities to the program.
ii. **Activities**: The second component of logic model focuses on the activities of program or set of treatment such as innovation, strategies or change planned for educational program.
iii. **Output**: The output component is an indicator of the program which revealed the completion of activities of the program or some components have been completed.
iv. **Outcomes**: This component focus on the short-term, medium-term and long term change intended as a result of the program activities. In the educational environment, program outcomes reveal the attainment of program goals by demonstrating learner knowledge or skills acquisition. The outcomes revealed learner outcomes in terms of meeting a performance standard on a relevant knowledge test or demonstrating specified skills.

Stufflebeam's CIPP Model

In response to the shortcomings of some program evaluation models, CIPP model was created for the purpose of improvement of the dominant experimental design of its time (Stufflebeam & Shinkfield, 2007). The CIPP model focuses on program improvement instead of providing information about the program. CIPP refers to the Context, Input, Process, and Product evaluation model. CIPP is an evaluation model that is concerned with decision making (Boutmetis & Dutwin, 2005).

CIPP model was first developed in 1971 by the Phi Delta Kappa Evaluation Committee (Smith, 1980). Stufflebeam (1971a), as a pioneer of CIPP model, has described the model as a process of obtaining information to describe the situation in order to provide useful information for making decisions about an educational program.

Different researchers have described CIPP model with different names, as a model of providing information for decision making (Stufflebeam, 1971b), as the best model for decision making (Boulmetis & Dutwin, 2005), as a system best model (Esreyel's, 2002), and Hew et al. (2004) describes the model as a macro level model for educational evaluation. Smith and Freeman (2002) and Williams (2000) are of the opinion that each one of the four types of evaluation that form CIPP model have an important role to play in the evaluation of any educational program with their functions, described by Stefflebeam (1971a) as follows:

a. **Context evaluation** serves planning decisions by identifying unmet needs, unused opportunities and underlying problems that prevent the meeting of needs or the use of opportunities.
b. **Input evaluation** serves structuring decisions by projecting and analyzing alternative procedural designs.
c. **Process evaluation** serves implementing decisions by monitoring project operations;
d. **Product evaluation** serves recycling decisions by determining the degree to which objectives have been achieved and by determining the cause of the obtained results (p. 268).

CIPP model of program evaluation have been identified as the most powerful approach to program evaluation for evaluators who understand educational programs in terms of elements of complex dynamics and nonlinear relationships of program component (Frye & Hemmer, 2012). CIPP approach to program evaluation is not hindered by assumptions of linear relationship, as it is a logic approach to evaluation and is constrained. The CIPP model for program evaluation consists of four complementary sets of evaluation studies that allow evaluators to evaluate each component of an educational program as independent.

The model accommodates the ever-changing nature of most educational programs and also focuses on program improvement. The first three components of the approach (context, input and process) are very useful components for the improvement of educational programs (formative evaluation) while the fourth component of the model is useful for summative evaluation.

Conclusion

In this conceptual paper, the aim was to explain complexity theory and its function as the most common and widely used theory in evaluation of educational programs. The theory explains to and guides educational evaluators on when and how the model should be applied to evaluate a particular program and each component of the program independently. Complexity theory was identified as the most appropriate and relevant theory in examining the complexity generated through interaction between different resources, values and goals, which is the peculiarity of teacher education training, as the study aims to highlight the functionality, strengths and weaknesses of models of program evaluation in evaluation of the effectiveness of any education program and its achievement of intended objectives.

The following conclusion can be drawn from the present paper, namely that the use of complexity theory in program evaluation is inevitable when evaluators ask questions about why and how programs work; in the context where the questions are asked, there is graeater alignment with complexity approach. Models of program evaluation are used to provide

information to policy and decision makers who have the responsibility for the existing program and or are designing or proposing new program and the information can be useful in decision making regarding whether to propose a new program or to develop an existing program or to modify or continue with the existing program. The evidence from this study suggests that CIPO evaluation model is the most appropriate and relevant evaluation model for educational programs as it identifies the relationship between the context, input, process and outcomes in education within a certain context. This research will serve as a base for future evaluation of mathematics teacher education training to evaluate effectiveness of the training in achieving national teacher education training objectives.

REFERENCES

Alkin, M. C. (2013*). Evaluation Roots: A Wider Perspective of Theorists' Views and Influences,* (2nd ed.) Thousand Oaks, CA: SAGE.

Blackman. T., Wistow, J., & Byrne, D. (2011). A Qualitative Comparative Analysis of factors associated with trends in narrowing health inequalities in England. *Social Science and Medicine 72*, 1965–74.

Boulmetis, J., & Dutwin, P. (2005). *The ABCs of evaluation: Timeless techniques for program and project managers* (2nd ed.), San Francisco: Jossey-Bass.

Carey, G., Malbon, E., Carey, N. et al. (2015). Systems science and systems thinking for public health. *A systematic review of the field*, BMJ Open 5.

Creemers, B. P., & Reezigt, G. J. (1997). School effectiveness and school improvement: Sustaining links. *School effectiveness and school improvement*, *8*(4), 396-429

Cuyvers, G. (2002). *Kwaliteitsontwikkeling in het onderwijs.* Apeldoorn: Garant

Eseryel, D. (2002). Approaches to evaluation of training: Theory & practice. *Educational Technology & Society, 5(2*), 93-99.

Frechtling, J. (2007). *Logic modeling methods in program evaluation*. San Francisco: John Wiley & Sons

Frye, A. W., Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE guide no. 67. *Med Teach, 34*(5), 288-99.

Funnell, S. C., & Rogers, P. J. (2011*). Purposeful Program Theory*. San Francisco, CA: Jossey-Bass.

Gerrits, L. & Verweij, S. (2015). Taking stock of complexity in evaluation: A discussion of three recent publications. *Evaluation, 21*, 481–91.

Hew, K. F., Liu, S., Martinez, R., Bonk, C., & Lee, J. Y. (2004). Online education evaluation: What should we evaluate? *Association for Educational Communications and Technology, 27 (*IL), 19-23.

Johnson, N. (2007). Simply complexity: A clear guide to complexity theory. Oxford: One world Publications.

Kirkpatrick, D. (1996). Revisiting Kirkpatrick's four-level model. *Train Dev,* 1,54–59.

Milton, B., Moonan, M., Taylor-Ronbinson, D., et al. (2011). *How Can Health Equity Impact of Universal Policies be Evaluated? Insights into Approaches and Next Steps*. Liverpool: World Health Organization.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Morley, L., & Rassool, N. (1999). *School Effectiveness: Fracturing the Discourse*. London & New York: Falmer press.

Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation 20*, 160–75.

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: The Guilford Pres.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: SAGE.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, *25*(2), 197-230.

Reynolds, John C. (1974). "Towards a theory of type structure." In *Programming Symposium*, pp. 408-425. Springer, Berlin, Heidelberg.

Rogers, P. J. (2011). Implications of complicated and complex characteristics for key tasks in evaluation. In: Schwartz R., Forss, K., & Marra. M. (Eds.), *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers, 33–52.

Saylor, J. G., Alexander. W. M., & Lewis, J. 1. (1981). *Curriculum planning for better teaching and learning.* Japan: Holt, Rine hart and Winston In.

Scheerens, J. (1990). School Effectiveness and the Development of Process Indicators of School Functioning. *School Effectiveness and School Improvement, 1,* 61-80. doi: 10.1080/0924345900010106.

Scheerens, J. (1992). *Effective schooling: research theory and practice*. Onderwijs Organisatie en management.

Scheerens, J. (2000). Improving School Effectiveness; Fundamental of Educational Planning, *Vol. 68,* Paris: UNESCO: International Institute for Educational planning. Retrieved from http://www.unesco.org//iieg

Scheerens, J. (2015). School Effectiveness Research. *International Encyclopedia of the Social & Behavioral Sciences, 21,* 80-85. doi: 10.1016/B978-0-08-097086-8.92080-4

Scheerens, J., Glas, C. A., Thomas, S. M., & Thomas, S. (2003). Educational evaluation, assessment, and monitoring*: A systemic approach*, *1*(3), 20. Taylor & Francis.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Sage.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ 336*, 1281–3.

Smith, C. L., & Freeman, R. L. (2002). Using continuous system level assessment to build school capacity. *American Journal of Evaluation, 23*(3), 307–319

Smith, K. M. (1980*). An analysis of the practice of educational program in terms of the CIPP model*. Unpublished PhD thesis, submitted to Loyola University of Chicago.

Stern, E., Stame N., Mayne, J., et al. (2012). *Broadening the Range of Designs and Methods for Impact Evaluations*. London: Department for International Development.

Stufflebeam, D. L. (1971a). The use of experimental design in educational evaluation. *Journal of Educational Measurement, 8*(4), 267-274.

Stufflebeam, D. L. (1971b). An EEPA interview with Daniel L. Stufflebeam. *Educational Evaluation and Policy Analysis, 2*(4), 85-90.

Stufflebeam, D. L. (2003). The CIPP model for evaluation. *International handbook of educational evaluation*, 31-62. Springer, Dordrecht.

Stufflebeam, D. L. (2007). CIPP Evaluation model checklist: A tool for applying the CIPP Model to assess long-term enterprises. *Evaluation Checklists Project*, 1-16.

Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, & applications*. San Francisco: Jossey Bass/John Wiley & Sons, Inc.

Van, Petegem, K. (1997). Scholen op zoek naar hun kwaliteit. Effectieve- scholenonderzoek als inspiratiebron voor de zelfevaluatie van scholen. Unpublished PhD Thesis Submitted to Universiteit Gent.

Van, Petegem, K., Aelterman, A., Van Keer, H., & Rosseel, Y. (2008). The influence of student characteristics and interpersonal teacher behaviour in the classroom on student's wellbeing. *Social Indicators Research*, *85*(2), 279-291.

Walton, M. (2016). Expert views on applying complexity theory in evaluation: opportunities & barriers: *SAGE 22*(4), 410-423.

Weiss, C. H. (1998). *Evaluation. Upper Saddle River*, NJ: Prentice Hall.

Williams, D. D. (2000). Evaluation of Learning Objects and Instruction using Learning Objects. In D. Wiley. (Ed.), *The instructional use of learning objects*. Retrieved from http:///reusability.org/read/

Wood, P., & Butt, G. (2014). Exploring the use of complexity theory and action research as frameworks for curriculum change. *Journal of Curriculum Studies*, *46*(5), 676-696.