A COMPARATIVE ANALYSIS OF THE EFFICACY OF THREE PROGRAM-EVALUATION MODELS –A REVIEW ON THEIR IMPLICATION IN EDUCATIONAL PROGRAMS

Zafar Iqbal[1*], Muhammad Anees[2], Rahim Khan[3], Abdul Wadood[4], Shakila Malik[5]

[1*,5]Ph.D., Scholar, Qurtuba University of Science and Information Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan; [2]Lecturer in English Govt. College, Peshawar, Khyber Pakhtunkhwa, Pakistan; [3]Assistant Professor, Teacher Education Department, Qurtuba University of Science and Information Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan; [4]Assistant Professor, Department of Education, Sarhad University of Science and Information Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan.

Email: [1*]zafariqbal101275@gmail.com, [2]aneesjaffarshah@gmail.com, [3]dr.rahim43@gmail.com, [4]abdul.wadood3376@gmail.com, [5]shakila.malik224@gmail.com

Abstract

Purpose of the study: This article reviews the comparative efficacy, theoretical and practical background of three program evaluation models (Stufflebeam's CIPP model, Kirkpatrick's model, and outcome-based evaluation models) and their implications in educational programs. The article discusses the strengths and limitations of the three evaluation models.

Methodology: Peer-reviewed and scholarly journals were searched for articles related to program evaluation models and their importance. Keywords included program evaluation', 'assessment', 'CIPP model', 'evaluation of educational programs, 'outcome-based model, and 'planning'. Articles on Stufflebeam's CIPP model, Kirkpatrick's model, and outcome-based evaluation models were particularly focused because the review aimed at analysing these three models. The strengths and inadequacies of the three models were weighed and presented.

Main Findings: The three models –outcome-based evaluation model, the Kirkpatric model, and the CIPP evaluation model –discussed in this review, have some strengths and weaknesses. Among the compared models, the CIPP model seems more appropriate for its implantation in evaluating educational programs because it is broader, comprehensive, flexible, cost-effective, and feasible.

Applications of this study: Like other programs and projects, evaluation of educational programs is necessary to achieve high standards, better outcomes, and meet the objectives. Evaluation is employed before designing a particular educational program or during the already designed program. This review concludes that among different evaluation models, the CIPP evaluation model is more appropriate in evaluating educational programs because it is more comprehensive, efficient, and feasible. Employment of the CIPP model for evaluating educational programs can achieve plausible results about the overall progress of the educational programs.

Novelty/Originality of this study: This review highlights the importance of different program evaluation models. It concludes that the CIPP evaluation model offers an excellent mechanism to evaluate educational programs at different stages.

Keywords: Program Evaluation, CIPP Model, Educational Assessment, Accountability, Objective-oriented Outcomes, Formation, Summation, Planning, Decisions.

INTRODUCTION

Based on societal needs, different programs and projects are initiated by policymakers after a comprehensive planning considering their structure, design, costs, and intended outcomes. The programs may either be designed for the short term when the fulfillment of the desired outcomes is emergently required, or they may be designed for the longer terms when the goals are broader. Both short-term and long-term program may not necessarily be efficient and perfect. To monitor the progress of a particular program, its evaluation is necessary. Evaluation of programs provides sound background about the direction of progress of the programs, their functionality, and goals achievement. Without continuous monitoring and evaluation, opportunities for growth and development of a program remain overlooked. Evaluation of programs helps make decisions about the continuity, modification, or termination of the program based on linkages with finance, career, and welfare of the people (Butler, 2020). In general, the program-evaluation efforts have been long recognized. A systematic expansion and development were observed during the 1960s in the USA where the evaluation of different programs in the military, health care services and social organizations, and educational institutions were encouraged to attain the accountability, competency, and reforms (Stufflebeam, 2001).

Educational programs and organizations are important drivers in social, behavioural, professional, and economic development, and their evaluation is likely necessary as other programs and projects do. The evaluation of educational programs and organizations helps in making the right decisions, modifying the previous decisions, attitude establishment or modification, and building the capacity of organizations (Alkin and King, 2017). The contextual benefits of program evaluation have been recognized throughout the world, and stakeholders employ a variety of evaluation models to

enhance the efficacy of different programs. Several evaluation models such as the logic model, outcome-based evaluation, Stufflebeam's CIPP model, and Kirkpatrick's model, among others, have been developed and extensively used during the last few decades to assess the progress of different programs (Schalock, 2001a; Stufflebeam, 2001). Implementing a particular evaluation model for measuring the progress of educational programs depends on the structure of the evaluation model, its objectivity, feasibility and cost affectivity, intentions of the evaluation, and nature of the program to be evaluated. From theoretical and empirical studies, it is evident that different program evaluation models help in improving decision making, which leads to the quality enhancement and goal achievement of educational programs (Mizikaci, 2006; Rooholamini et al., 2017; Darma, 2019).

The objective of this review is to discuss the importance of three evaluation models, namely Stufflebeam's CIPP model, outcome-based evaluation model, and Kirkpatrick's model, in the quality enhancement of different educational programs. The three models are critically analysed and their strengths and shortcomings are highlighted.

METHODOLOGY

For structuring and organizing this article, a literature survey was conducted to collect relevant information about the program evaluation models. Peer-reviewed and scholarly journals were searched for articles related to program evaluation models and their importance. Keywords included program evaluation', 'assessment', 'CIPP model', 'evaluation of educational programs, 'outcome-based model, and 'planning'. Major databases such as google scholar, PubMed, NIH Library, andElsevierwere searched for articles by entering the keywords. A total of 38 journal articles indexed in Scopus and Web of Science (WoS) were selected for this review. Articles on Stufflebeam's CIPP model, Kirkpatrick's model, and outcome-based evaluation models were particularly focused because the review aimed at analysing these three models. The strengths and inadequacies of the three models were weighed and presented.

REVIEW DISCUSSION

PROGRAM EVALUATION MODELS: THEORETICAL AND PRACTICAL IMPLICATIONS

Program evaluation is a systematic approach to assess, analyse and use the information about the progress, outcomes, goal achievements, and effectiveness of policies, projects, and programs (Usun, 2016). Every program and project have primary purposes, which are designed and executed through rigorous planning and policy inputs by stakeholders. Shadish et al. (1991) argued that program evaluation models could not be developed without the assistance of theories. They asserted that before jumping to practical manipulation of the program, understanding of knowledge, basic concepts, and rules are necessarily provided by theories. Theories provide a basic framework for designing and developing a program evaluation model. Chen (2016) stated that theories provide assumptions on how useful a program evaluation model be designed by understanding the basic components and contextual aspects of the program. From theoretical perspectives, logical conciseness, the systematic intervention of different program components, and some fundamental rules providing efficacy guidelines are the core components in developing a program. Without linking the program to a sound philosophical and theoretical background, the purpose of designing an ideal evaluation model cannot be met.

Historically, different theories have played influential roles in constructing and reforming different evaluation models for education and other programs. Widely acknowledged theories are reductionism, complexity theory, and general system theory (Figure 1). According to Chen (2016), reductionism theory suggests the breakup of a program into its core components which may be crucially analysed and understood. Frye & Hemmer, (2012) stated that reductionism theory focuses first on the understanding the integral program, and then analysing its constituent elements and their contribution to the outcome of the integral program. They further elaborate that a linear relationship between the elements of the program can create room for changes which can impart a predictable impact on the program's outcome depending on the magnitude of the changes, e.g., as evident in the Logic evaluation model. The general system theory initially proposed by Bertalanffy in the 1920s for medicine but later on adopted for several other disciplines is the reverse of reductionism theory, and it weighs the whole system (program) as more important than its components (Frye & Hemmer, 2012). Mizikaci, (2006) explained the general system theory by highlighting its assumption of "the wholeness and universal application of the principles of organization". As per this explanation, "whole" is the crucial component of the general system theory than components, and "whole" specifies the nature of the components, which are difficult to understand if they are isolated from the "whole". Complexity theory or theory of complex adaptive system (CAS) in general states that systems (programs, organizations, environments, etc.) consist of interacting components, where diversity prevails, certainty is rare, ambiguity is certain, equilibrium is rare, evolution is common, and changes frequently occurs due to interacting components of the system, and due to cause and effects phenomena (Frenken, 2006; Morrison, 2008; Norberg & Cumming, 2008; Walton, 2014). Cunningham (2003) stated that complex theory considers a system to be dynamic, not at equilibrium, indeterminate, open to sharing information with the surrounding, based on feedback, and where the whole is regarded more than parts.

How these and other theories are helpful in constructing the program evaluation models? To answer this question, evaluators and those involved in developing a program evaluation model are supposed to consider the framework of the theory and the nature of the program. Theories provide mechanisms and concepts on how to construct an evaluation model. Evaluators consider theoretical guidelines such as which methods are suitable for the evaluation of a specific

program, where and when those methods are applied. At this point, some methods may be skipped or modified to integrate the whole evaluative efforts (Boulay, & Han, 2008). The nature of the program determines which theoretical framework best suits its evaluation purposes. Rogers (2000) suggested that a particular program model should have four elements: (i) activities of the program, (ii) anticipated outcomes of the program, (iii) operative mechanisms, and (iv) the context. These elements are derived either from a specific theory or a combination of different theories. Educational programs differ in their components, intended outcomes, contextual framework, and operative mechanisms. Reductionism theory may be ideal for developing an evaluation model for a particular educational program but not suitable for another one. Similarly, other theories may have relevance for some programs while inappropriate for another program. Therefore, the program evaluator must consider simultaneously the basic concept of an evaluation theory (scientific approach, theoretical backgrounds, applicability, underlying mechanisms, etc.) and the nature of the program to be evaluated (program's objectives, intended outcomes, components, and contexts, etc.).
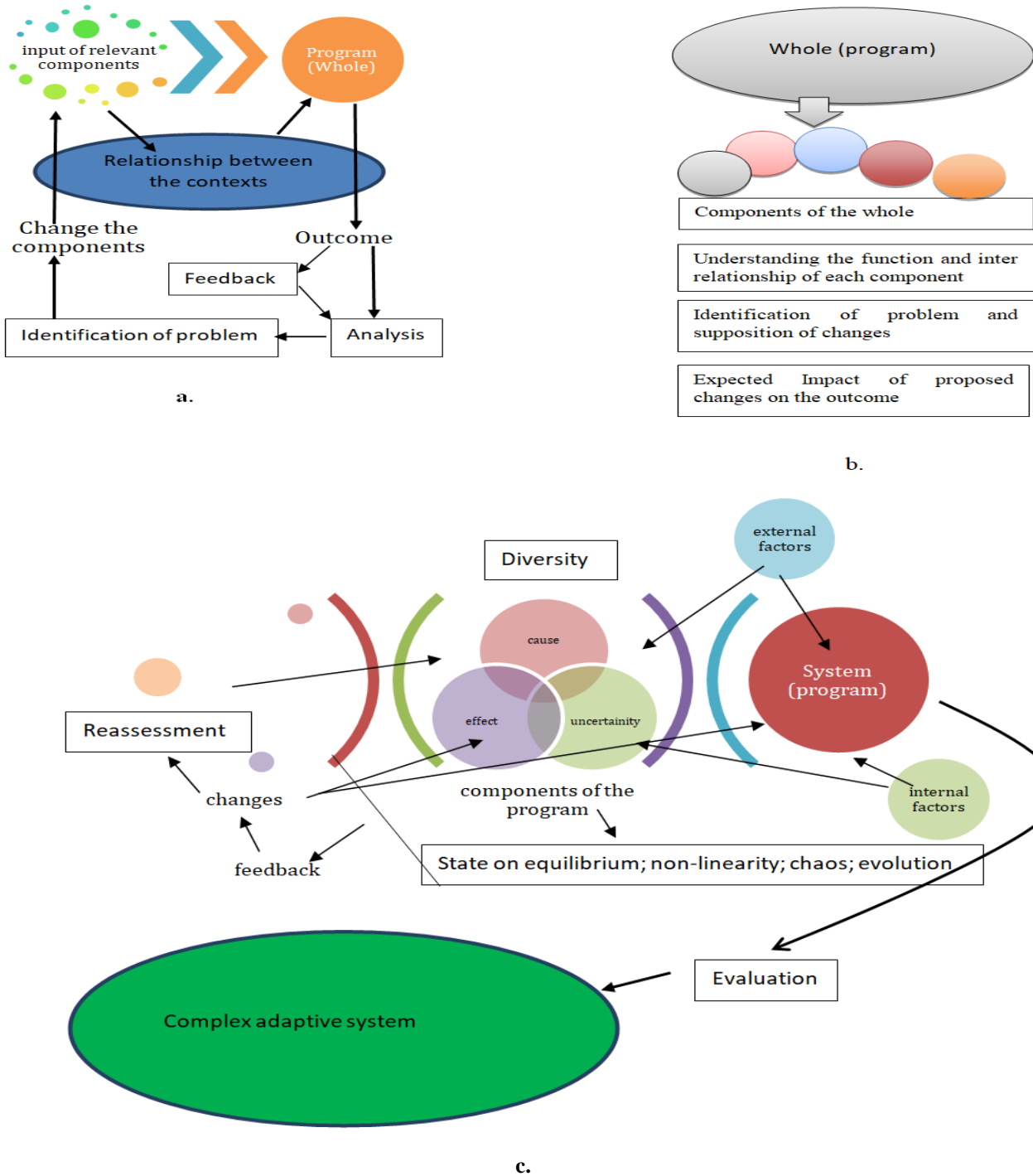

Figure 1: An illustration of the theoretical basis of program evaluation models; a. General system theory, b. Reductionism theory, c. Complexity theory

DIFFERENT EVALUATION MODELS AND THEIR APPLICATION IN EDUCATIONAL PROGRAMS

Educational institutions notably higher educational institutes, in most of the countries, have no proper mechanisms to provide information about the outcomes of their educational programs; instead, they emphasize only on course, activities and subject contributions, and resource and research output, which according to Nusche (2008) are not adequate indicators of the quality of educational programs. To grasp a clear picture of the quality, efficiency, and affectivity of the educational programs, a comprehensive evaluation mechanism is always needed, which can lead to the removal of flaws and improvement of educational programs. Over the last few decades, efforts have been made to improve the quality, integrity, and standards of different educational programs by employing different approaches. One of the approaches is the use of a systematically structured tool termed as "program evaluation model". The purpose of an evaluation model is to assess whether the program fulfills the required needs, provides anticipated services, delivers the desired outcomes, achieves its goals and objectives, and is functioning in the way it was planned (Posavac, 2015). Evaluation of an educational program intends to check the program's progress, identify merits and flaws, and acquire information. After analysing the quality indicators of educational programs by an evaluation model, previous decisions are either maintained or modified to achieve better program outcomes. In recent years, the evaluation of programs has become a dynamic profession with wide application in health, governments, organization, and education (Madaus & Kellaghan, 2000). In the succeeding paragraphs, three evaluation models –outcome-based evaluation model, The Kirkpatrick 4 step model, and Stufflebeam's CIPP models –are discussed with their implementation in educational programs.

OUTCOME-BASED EVALUATION MODEL

The Institute of Museum and Library Services defines outcome-based evaluation as a systematic approach towards evaluating the estimation of the program's intended outcomes, effectiveness, and the benefits that the program have provided to the participants (http://www.nysl.nysed.gov/libdev/obe/). According to Brewer (2011), the outcome-based evaluation model has its origin in the United States, and it focuses on the results and effectiveness of the program and the benefits clients draw from the program. The author noted that information obtained through an outcome-based evaluation model might lead to the summation or formation of the program. He suggested its usefulness in educational programs, health systems, and organizational evaluation. Schalock (2001a) demonstrated the primary purposes of the outcome-based model as to streamline:

a. The expected and actual outcomes of a program.

b. Accountability demands of the program.

c. Goals and objectives.

d. Efficiency and affectivity.

e. Functionality and policy on which the program runs.

f. The use of information and feedback to improve the program.

In another article, Schalock (2001b) described that the outcome-based evaluation model takes into account stakeholders, promoters, and program evaluators. He elaborated that the model employs four major approaches during the evaluation process of a given program, i.e. program evaluation, effectiveness evaluation, impact evaluation, and policy evaluation (Table 1). A review of the literature indicates that the outcome-based evaluation model has been used to assess the progress of several education programs. Celinska et al. (2013) applied an outcome-based evaluation model to analyse functional family therapy in youth who exhibited behavioural issues. They reported that evaluation revealed significantly improved results in youth's behavioural outcome. Asch et al. (2014) employed an outcome-based evaluation model in medical education based on the patient outcomes. They suggested that the OBE is an effective tool in measuring the progress of medical education and revealed that training, skills, and experience were the major quality indicators of medical education program. In other studies, the OBE model has been employed to evaluate nursing competency (skills, learning potential, caring, etc.) in medical education (Hsieh & Hsu, 2013), social skill therapy (Epp, 2008), teaching and learning at university (Morris, 2008), inter-professional learning in health care (Nisbet et al., 2008), faculty development program (Chen et al., 2013), and education and training (Alexandrov & Sancho, 2017).

Table 1: A summary of the outcome-based evaluation model

Primary agents of the model	Role of the agents	Components of the model	Indicators/feedback	Results
Stakeholders	Policymaking, program designing, decision implementation	Program evaluation	a. The overall progress of the program is satisfactory or not b. The	Summative or formative

			program is functioning adequately or not	
Evaluators	Evaluation, assessment, and analyzing the program	Effectiveness evaluation	a. The outcomes are effective or not b. Goals and objectives are met or not	Summative or formative
Promoters	Funding	Impact evaluation	a. Does the program has an impact on society b. Is the program better than another program	Summative or formative
		Policy evaluation	a. Were the policies adequate or not b. What were the flaws or worth of policies	Summative or formative

Limitation and strengths

The outcome evaluation model stresses on the outcomes of the program. It has been widely used in the evaluation of educational programs and has been found an effective tool; however, the model has some limitations as well. According to Schalock (2001b), when evaluators consider the appropriate measures and scientific methods while evaluating the program, its efficiency increases. However, he also highlighted that the influence of internal and external factors, validity, and reliability of tools might produce flawed assessments and decrease the efficiency of the outcome-based evaluation model. Ewell (2008) noted that the outcome-based model is flexible, transparent, comparable, and portable at the implementation phase. Based on the views of other critics, Tam (2014) highlighted some limitations of the outcome-based approaches in education. Those limitations are specificity, narrowness, quantifiability, and observability which lead to reductionism and negligence of integrative assessment of the educational program.

THE KIRKPATRICK EVALUATION MODEL

The Kirkpatrick evaluation model is a valuable tool to evaluate educational programs and professional training programs, which has been used by several organizations and institutions for evaluating their progress in the US and throughout the world (Smidt et al., 2009; Praslova, 2010; Gill & Sharma, 2013). The model was initially developed by Kirkpatrick in 1959 for evaluating training programs, later on, it was modified and made applicable for assessing the effectiveness of several programs ranging from health care, medical education, organizational performance to higher education (Praslova, 2010; Liao, & Hsu, 2019). Kirkpatrick & Kirkpatrick (2006), justified the need for program evaluation to produce better outcomes of the program and enhance its effectiveness but also emphasized for considering the needs, objectives, subject contents, participants, appropriate facilities, and several other factors before the planning, designing, and execution of that particular program. They explained the Kirkpatrick evaluation model on the basis of its four levels which are interdependent on each other:

1. Reaction –how the participants react i.e. whether they are satisfied or not, and how they feel about the evaluation program.

2. Learning –is the program effective to enhance the learning capabilities of the participants and increase their knowledge and skills.

3. Behaviour –to what extent the program is effective in changing the attitude and behaviour of the participants.

4. Results –the overall outcome of the program has improved or stagnant.

Basically, these levels were elucidated for the training programs to improve the reaction, behaviour, learning, and results of the participant. In the broader context of educational programs, a modified version of the Kirkpatrick evaluation model can be effectively implemented in educational institutions, not only focusing on the participants but also the curriculum, needs, and goals of the programs, and their productive outcomes. Dorri et al. (2016) evaluated a nursing program at Shahadaye Lenjan Hospital in Isfahan province by employing the Kirkpatrick model. They recorded positive outcomes of the four levels of the Kirkpatrick model on the in-service training and nurses. Paul et al. (2016) demonstrated that under certain circumstances, the Kirkpatrick evaluation model is a valuable tool in improving the curriculum of educational programs. Dewi & Kartowagiran (2018) applied the Kirkpatrick evaluation model in evaluating an internship program comprising students and instructors. They obtained promising results for improvement in teaching and learning categories, schedule, material, and level of satisfaction. In other studies, effectiveness of the model for evaluating teachers' performance (Naugle et al., 2000), hospitality industry (Chang, 2010), health information management courses and programs (Rouse, 2011), educational quality (Misut et al., 2013), professional development (Wu et al., 2016), high impact leadership program (Miller, 2018), vocational training programs (Ravicchio & Trentin, 2015), and e-learning (Galloway, 2005) has been well established.

Strengths and limitations

Bates (2004) wrote in his review that the model is popular and advantageous because it simplifies the complex process of training evaluation in a number of ways. The Kirkpatric model presents many advantages as it focuses on a comprehensive approach targeting major four components i.e. reaction, learning, the behaviour, and outcomes where potential improvement is possible and can lead to the successive progress of the program (Carpenter, 2011). Mann &Rajeev et al.(2009) identified the focus of the Kirkpatrick's evaluation on the behaviour of the trainees as the main strength of the model. According to Kaufman and Keller (1994), the Kirkpatric model was developed with the intention of evaluating training programs and which is still mainly in practice. Organizations and institutions do not only need to evaluate training but other components of their programs too, which need a more comprehensive model, which will cover all the elements of the programs. In their review, Reio et al. (2017) outlined some limitations of the model. They asserted that in the Kirkpatric model more weightage is given to upper levels (behaviour and outcomes) than lower levels (reaction and learning); therefore most of the organizations and professionals tend to neglect lower levels. Interdependency of the four levels is another drawback because it is not necessary that the execution of one level would lead to better outcomes for the next level. Similarly, difficulty in evaluating level 3 and four are considered by some researchers as drawbacks of the model (Moreau, 2017). Cahapay (2021) suggested that while applying the Kirkpatric model in higher education, evaluators must consider its limitations, such as consideration of lower levels as less important, rigidity, and causal linkage of the four levels.

STUFFLEBEAM'S CIPP EVALUATION MODEL

CIPP evaluation model was designed by Stufflebeam in 1971, comprising of four quality indicators i.e. Context, Input, Process, and Product (Aziz et al., 2018). The model has been successfully used to monitor and improve the quality of projects, evaluation systems, institutions, and educational programs throughout the world. According to Stufflebeam (2000), the CIPP model uses four key indicators to evaluate a program. These components are context, input, process, and product (Figure 2). He illustrated that the context includes evaluation of needs, complications, and prospects of the program. In educational programs, context may be the identification of resources, goals, policies, and potential problems. It also includes beneficiaries of the programs. Overall missions and goals, background information, and cultural context of the educational program are also covered in context evaluation (https://poorvucenter.yale.edu/). The context is all about the recent information about the prospective functioning of the program, and the evaluators use different techniques such as surveys, interviews, documents review during context evaluation (Brewer, 2011). Stufflebeam & Zhang (2017) suggested that before designing a program, stakeholders should address the needs, goals, priorities and anticipated problems, potential risks, and opportunities. Context evaluation of the program provides information to decision-makers in the form of evaluation reports which help them in planning and setting goals in an appropriate manner (Stufflebeam & Zhang, 2017). Input evaluation assesses the strategies, resources, both financial and services, mechanisms and designs of the program's functioning, action plans, cost-effectiveness, and arrangements (Stufflebeam & Zhang, 2017). In the case of educational programs, the input may be financial resources, human resources, infrastructural resources, and a documentary work plan. In this approach, availability of financial and service resources, infrastructure and environment, and feasibility are judged and reported whether they are appropriate for executing the program. Process evaluation is an essential phase of the CIPP model, which focuses on the progress of the program. It involves monitoring, assessing, and documentation of the implementation of the designed plan in light of set goals (Stufflebeam & Zhang, 2017). The final component of the CIPP model is the product evaluation which emphasizes that whether the outcomes of the program coheres with the objectives. Stufflebeam & Zhang (2017) illustrated that the product evaluation measures the outcomes and cost-effectiveness of the program. In an educational program, product evaluation focuses on the final results, appropriateness of objectives concerning to outcomes, and the overall cost-effectiveness of the program.
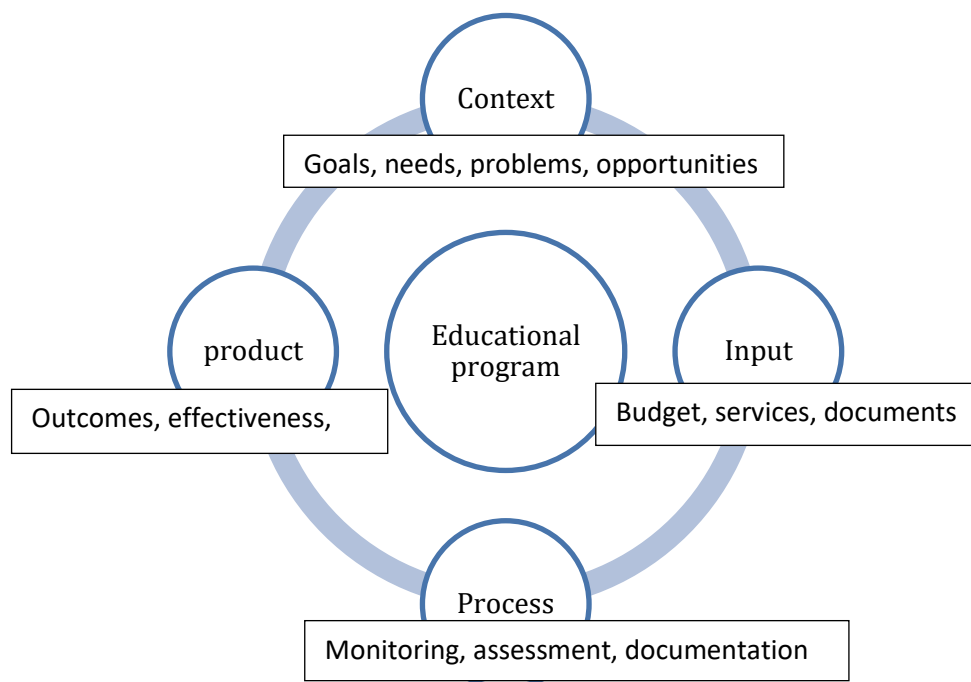
Figure 2: Graphical illustration of the CIPP evaluation model

Like other evaluation models, the CIPP model has wide application in a diverse range of education programs and has been used extensively throughout the world by educational institutes to assess the progress of their programs. AbdiShahshahani et al. (2015) used the CIPP model to assess the status of a Ph.D. program in reproductive health in different midwifery and nursing schools in Tehran. Their results revealed that the quality of Ph.D. in reproductive health assessed through context, input, process, and the product was appropriate. They further recommended consistent evaluation of the program for improvements and better outcomes. Mokhtarzadegan et al. (2015) conducted an analytical study to evaluate the strengths and weaknesses of in-service training of employees of Sheraz University applying the CIPP model. They recorded significant differences between the education level of employees and the product phase. They revealed that in-service training of the employees enhanced their efficiency; however, some weak areas were also identified which needed improvement. Akpur et al. (2016) outlined that the instruction program at Yildiz Technical University evaluated through the CIPP model had both positive and negative responses from students and teachers. The respondents revealed their grievances about the availability of audio-visual aids, skills, and English knowledge, they, however, showed positive ideas towards the curriculum. Afzal and Yousaf (2017) developed and applied the CIPP evaluation model for the assessment of the BS 4-year program in Education at four public universities in Punjab. Their study found that students and teachers showed satisfactory responses towards the context, input, process, and product of BS education. Martínez et al. (2018) employed the CIPP model to evaluate the stipulated goals set for skill input in students of a Bachelor's degree in a foreign language. They assessed the input of communicative skills of students in the first four semesters of the bachelor's degree and found a discrepancy in the proposed objectives and actual outcomes. Lippe & Carter (2018) highlighted the effectiveness of Stufflebeam's CIPP model in the assessment of curriculum. They evaluated the quality and merit of the nursing program and observed both missing contents in the curriculum and strengths of the program. Aziz et al. (2018) conducted a study to analyse the quality parameters of education at the Welfare School System in Rawalpindi using the CIPP model. Their findings revealed that theoretical and rote-learning were more emphasized by teachers than creating intellectual abilities in students. Gunung& Darma (2019) applied the CIPP model to measure the effectiveness of teaching implementation in polytechnic institutes in Bali. The researchers identified some obstacles in the product component of the model, whereas other components revealed fair scores.

Strengths and limitation

The CIPP model is a simple, flexible, and helpful tool to measure the effectiveness of the program because it focuses on all the strategies and elements of the evaluation (Hakan & Seval, 2011). Brewer (2011) noted that the model has a significant impact on improving the processes because it was developed to be applied at planning and implementation stages. The author further argues that the CIPP model helps in improving the effectiveness and accountability of programs considering the costs. According to Mohebbi et al. (2011), the CIPP model is comprehensive and aids in objectively evaluating the program. Furthermore, on a need basis, evaluators can choose a specific component of the CIPP model instead of its four components to evaluate the program. Finney (2020) identified the simplicity and broad applicability of the CIPP model as its significant strengths. Besides its multipurpose nature, comprehensiveness and simplicity, the model has some limitations. Anh (2018) outlined that the model is time-consuming when all of its components are applied to program evaluation. He highlighted that context evaluation holds similarity to need assessment.

WHICH PROGRAM EVALUATION MODEL SHOULD BE EMPLOYED IN EDUCATION?

Like other programs and projects, the evaluation of educational programs is necessary to achieve high standards, better outcomes, and meet the objectives. The evaluation can be done before designing a particular educational program or during the already executing program. The purpose of the evaluation is to make an educational program effective in all respects before and after the initiation of that program. The three models discussed in this review have some similarities, differences, strengths, and weaknesses (Table 2). Educational stakeholders should consider several aspects of the evaluation models and the educational programs to be evaluated. Both the evaluation models and educational programs have diverse characteristics. The prime characteristics for selecting a suitable evaluation model, the stakeholders should consider their feasibility, simple application, cost-effectiveness, objectivity, comprehensiveness, and time consumption. The outcome-based evaluation model focuses on the results and effectiveness of the program, and the benefits the clients draw from the program and it has been applied in specific educational programs, health systems, and organizational evaluation (Brewer, 2011). However, the model, in general, is specific, narrow, quantifiable, and observable in nature (Tam, 2014) and thus has limitations in broad applicability in a diverse range of educational programs. Although the Kirkpatric model is more comprehensive and broader than OBE because it focuses on reaction, learning, behaviour, and outcomes, and tend to simplify the complex process; however, some limitations like focus on training, weighing higher levels more than lower levels, and difficulty in evaluating level 3 and four make it less appropriate for evaluating the educational programs (Moreau, 2017). The CIPP model seems to possess a balanced approach towards evaluation despite its few limitations. First, the model offers a comprehensive, feasible, and straightforward framework, which makes it a more suitable tool for evaluating diverse educational programs. Second, it takes into account pre, during, and after execution approaches for evaluating a program. Moreover, based on the specific needs, either a single component or whole components may be employed while evaluating an educational program.

Table 2: Characteristics of OBE, the Kirkpatrick model, and the CIPP models (Modified from Stufflebeam, 2000; Schalock, 2001a; Bates, 2004)

Evaluation model	Components	Approaches	Strengths	Limitation	Areas for improvement
Outcome-based model	Program, effectiveness, impact, and policy evaluation	Learner-centred; consumer assessment, functional assessment, personal evaluation	Outcome-oriented, flexible, comparable	specificity, narrowness, quantifiability, and observability	Needs comprehensive modifications; should be generalized
the Kirkpatrick model	Reaction, Learning, Behaviour, and Results	Learner-centred; personal assessment; behavioural evaluation	Simplifying the complex process, focus on training and behaviour, reduces the measurement demands	Interdependency of components, non-comprehensive, suitable for training programs; Does not take into account the effect of variables; the significance of upper levels; causal effects of components	Equal weightage should be given to each component; comprehensive mechanisms should be used; should be modified to be applicable in other programs beyond training programs
The CIPP model	Context, input, process, and product	The program-centred; uses different methods (surveys, interviews, document review, etc.)	Multipurpose application, flexible, comprehensive, easy to employ	Time-consuming	Implementation of a specific component of the model may reduce time

CONCLUSION

The program evaluation is a necessary process to design, execute, and improve the progress of the program. During the last few decades, different evaluation models have been developed and extensively used for evaluating different projects and programs. Educational stakeholders have been applying many models to evaluate educational programs for improvement. The three models –the outcome-based evaluation model, the Kirkpatric model, and the CIPP model – discussed in this review have some strengths and weaknesses. Among the compared models, the CIPP model seems

more appropriate for its implantation in evaluating educational programs because it is broader, comprehensive, flexible, cost-effective, and feasible. The model can be implemented at different stages of educational programs. The review concludes that employment of the CIPP model for evaluating educational programs can achieve plausible results about the overall progress of the educational programs.

ACKNOWLEDGEMENT

AUTHORS CONTRIBUTION

This manuscript is a part of the Ph.D. thesis of Mr. Zafar Iqbal. Mr. Zafar Iqbal designed and executed the study. Mr. Muhammad Anees reviewed the literature. Mr. Rahim Khan supervised and organized this research. Mr. Abdul Wadood and Ms. Shakila Malik surveyed the literature sources.

REFERENCES

1. AbdiShahshahani, M., Ehsanpour, S., Yamani, N., Kohan, S., & Hamidfar, B. (2015). The evaluation of reproductive health PhD program in Iran: a CIPP model approach. *Procedia-Social and Behavioral Sciences*, 197, 88-97. https://doi.org/10.1016/j.sbspro.2015.07.059
2. Afzal, A., & Yousaf, M. (2017). Development of an Evaluation Model for BS 4 Years Program at University Level. *Pakistan Journal of Social Sciences*, 37(1), 114-124.
3. Akpur, U., Alci, B., & Karatas, H. (2016). Evaluation of the Curriculum of English Preparatory Classes at Yildiz Technical University Using CIPP Model. *Educational Research and Reviews*, 11(7), 466-473.
4. Alexandrov, N., & Sancho, M. R. (2017). Learning outcomes based evaluation of HPC professional training. *Procedia Computer Science*, 108, 2141-2150. https://doi.org/10.1016/j.procs.2017.05.135
5. Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434-450.
6. Anh, V. T. K. (2018). Evaluation models in educational program: Strengths and weaknesses. *VNU Journal of Foreign Studies,* 34(2), 140-150. https://doi.org/10.25073/2525-2445/vnufs.4252
7. Asch, D. A., Nicholson, S., Srinivas, S. K., Herrin, J., & Epstein, A. J. (2014). How do you deliver a good obstetrician? Outcome-based evaluation of medical education. *Academic Medicine*, 89(1), 24-26. https://doi.org/10.1097/ACM.0000000000000067
8. Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP model for quality evaluation at school level: A case study. *Journal of Education and Educational Development*, 5(1), 189-206. http://dx.doi.org/10.22555/joeed.v5i1.1553
9. Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27(3), 341-347. https://doi.org/10.1016/j.evalprogplan.2004.04.011
10. Boulay, D. A., & Han, H. (2008). Examining Values, Use, and Role in Evaluation: Prospects for a Broadened View. Online Submission.
11. Brewer, E. W. (2011). Evaluation models for evaluating educational programs. In Assessing and evaluating adult learning in career and technical education (pp. 106-126). IGI Global. https://doi.org/10.4018/978-1-61520-745-9.ch007
12. Butler, M. O. (2020). Evaluation. In Oxford Research Encyclopaedia of Anthropology.
13. Cahapay, M. B. (2021). Kirkpatrick Model: Its Limitations as Used in Higher Education Evaluation. International *Journal of Assessment Tools in Education*, 8(1), 135-144. https://doi.org/10.21449/ijate.856143
14. Carpenter, J. (2011). Evaluating social work education: A review of outcomes, measures, research designs and practicalities. *Social Work Education*, 30(02), 122-140. https://doi.org/10.1080/02615479.2011.540375
15. Celinska, K., Furrer, S., & Cheng, C. C. (2013). An outcome-based evaluation of functional family therapy for youth with behavioral problems. *Journal of Juvenile Justice*, 2(2), 23-36.
16. Chang, Y. H. E. (2010). An empirical study of Kirkpatrick's Evaluation Model in the hospitality industry. PhD dissertation, Florida International University. https://doi.org/10.25148/etd.FI10120807
17. Chen, W., Kelley, B., & Haggar, F. (2013). Assessing faculty development programs: Outcome-based evaluation. *Journal on Centers for Teaching and Learning*, 5, 107-19.
18. Chen, H. T. (2016). Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. *Evaluation and Program Planning*, 59, 109-118. https://doi.org/10.1016/j.evalprogplan.2016.05.012
19. Cunningham, R. (2003). Complexity theory and school improvement: some possible connections.Paper presented at the British Education Research Association conference, Edinburgh University, September, 2003. http://www.leeds.ac.uk/educol/documents/00003214.doc
20. Darma, I. K. (2019). The effectiveness of teaching program of CIPP evaluation model. *International ResearchJournal of Engineering, IT and Scientific Research*, 5(3), 1-13. https://doi.org/10.21744/irjeis.v5n3.619

21. Dewi, L. R., & Kartowagiran, B. (2018). An evaluation of internship program by using Kirkpatrick evaluation model. *Research and Evaluation in Education*, 4(2), 155-163. https://doi.org/10.21831/reid.v4i2.22495

22. Dorri, S., Akbari, M., & Sedeh, M. D. (2016). Kirkpatrick evaluation model for in-service training on cardiopulmonary resuscitation. *Iranian Journal of Nursing and Midwifery Research*, 21(5), 493-497. https://doi.org/10.4103/1735-9066.193396

23. Epp, K. M. (2008). Outcome-based evaluation of a social skills program using art therapy and group therapy for children on the autism spectrum. *Children and Schools*, 30(1), 27-36. https://doi.org/10.1093/cs/30.1.27

24. Ewell, P. (2008). Building academic cultures of evidence: a perspective on learning outcomes in higher education. In Symposium of the Hong Kong University Grants Committee on Quality Education, Quality Outcomes–the way forward for Hong Kong, Hong Kong (Vol. 15, p. 2011).

25. Finney, T. L. (2020). Confirmative Evaluation: New CIPP Evaluation Model. *Journal of Modern Applied Statistical Methods*, 18(2), 1-24. https://doi.org/10.22237/jmasm/1598889893

26. Frenken, K. (2006). Technological innovation and complexity theory. *Economics of Innovation and New Technology,* 15(2), 137-155. https://doi.org/10.1080/10438590500141453

27. Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE guide no. 67. *Medical Teacher*, 34(5), 288-299. https://doi.org/10.3109/0142159X.2012.668637

28. Galloway, D. L. (2005). Evaluating distance delivery and e-learning is Kirkpatrick's model relevant?. *Performance Improvement*, 44(4), 21-27.

29. Gill, M., & Sharma, G. (2013). Evaluation of vocational training program from the trainees' perspective: An empirical study. *Pacific Business Review International*, 6(5).

30. Gunung, I. N., & Darma, I. K. (2019). Implementing the context, input, process, product (CIPP) evaluation model to measure the effectiveness of the implementation of teaching at Politeknik Negeri Bali (PNB). *International Journal of Environmental and Science Education*, 14(1).

31. Kirkpatrick, D., & Kirkpatrick, J. (2006). Evaluating training programs: The four levels. Berrett-Koehler Publishers, Inc., USA

32. Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: development, reliability and validity. *Procedia-Social and Behavioral Sciences*, 15, 592-599. https://doi.org/10.1016/j.sbspro.2011.03.146

33. Hsieh, S. I., & Hsu, L. L. (2013). An outcome-based evaluation of nursing competency of baccalaureate senior nursing students in Taiwan. *Nurse Education Today*, 33(12), 1536-1545. https://doi.org/10.1016/j.nedt.2013.01.012

34. Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly*, 5(4), 371-80. https://doi.org/10.1002/hrdq.3920050408

35. Liao, S. C., & Hsu, S. Y. (2019). Evaluating a continuing medical education program: New world Kirkpatrick model approach. *International Journal of Management, Economics and Social Sciences*, 8(4), 266-279. https://doi.org/10.32327/IJMESS/8.4.2019.17

36. Lippe, M., & Carter, P. (2018). Using the CIPP model to assess nursing education program quality and merit. *Teaching and Learning in Nursing*, 13(1), 9-13. https://doi.org/10.1016/j.teln.2017.09.008

37. Madaus, D. L. S. G. F., & Kellaghan, T. (2000). Evaluation models: Viewpoints on educational and human services evaluation (Vol. 49). Springer Science & Business Media.

38. Martínez, G. R., Pinta, D. A., & Santacruz, L. Á. B. (2018). CIPP model to evaluate the principles applied to develop the Input Skills in the Bachelor Degree Program of EFL. *International Journal of Education and Learning Systems*, 3, 117-128.

39. Miller, B. J. (2018). Utilizing the Kirkpatrick model to evaluate a collegiate high-impact leadership development program (Doctoral dissertation). Texas A&M University, Texas ,College Station. https://oaktrust.library.tamu.edu

40. Misut, M., Pribilova, K., Orolinova, M., & Kotulakova, K. (2013). Quality of Education Improvement through Adapted Kirkpatrick Evaluation Model. *Advances in Education Research*, 44, 109-114.

41. Mizikaci, F. (2006). A systems approach to program evaluation model for quality in higher education. *Quality Assurance in Education.* 14 (1), 37-53.

42. Mohebbi, N., Akhlaghi, F., Yarmohammadian, M. H., & Khoshgam, M. (2011). Application of CIPP model for evaluating the medical records education course at Master of Science level at Iranian medical sciences universities. *Procedia-Social and Behavioral Sciences*, 15, 3286-3290. https://doi.org/10.1016/j.sbspro.2011.04.287

43. Mokhtarzadegan, M., Amini, M., Takmil, F., Adamiat, M., & Sarveravan, P. (2015). Inservice trainings for Shiraz University of Medical Sciences employees: effectiveness assessment by using the CIPP model. *Journal of Advances in Medical Education and Professionalism*, 3(2), 77-83.

44. Moreau, K. A. (2017). Has the new Kirkpatrick generation built a better hammer for our evaluation toolbox?. *Medical Teacher*, 39(9), 999-1001. https://doi.org/10.1080/0142159X.2017.1337874

45. Morris, M. M. (2008). Evaluating university teaching and learning in an outcome-based model: Replanting bloom. PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong, 2008. http://ro.uow.edu.au/theses/784

46. Morrison, K. (2008). Educational philosophy and the challenge of complexity theory. *Educational Philosophy and Theory*, 40(1), 19-34. https://doi.org/10.1111/j.1469-5812.2007.00394.x

47. Naugle, K. A., Naugle, L. B., & Naugle, R. J. (2000). Kirkpatrick's evaluation model as a means of evaluating teacher performance. *Education*, 121(1), 135-145.

48. Nisbet, G., Hendry, G. D., Rolls, G., & Field, M. J. (2008). Interprofessional learning for pre-qualification health care students: An outcomes-based evaluation. *Journal of inter-professional Care*, 22(1), 57-68. https://doi.org/10.1080/13561820701722386

49. Norberg, J., & Cumming, G. (2008). Complexity theory for a sustainable future. Columbia University Press.

50. Nusche, D. (2008). Assessment of learning outcomes in higher education: A comparative review of selected practices.OECD Education Working Paper No. 15. www.oecd.org/dataoecd/14/8/40257354.pdf

51. Posavac, E. J. (2015). Program evaluation: Methods and case studies. Routledge.

52. Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability*, 22(3), 215-225. https://doi.org/10.1007/s11092-010-9098-7

53. Rajeev, P., Madan, M. S., & Jayarajan, K. (2009). Revisiting Kirkpatrick's model–an evaluation of an academic training course. *Current Science*, 272-276.

54. Ravicchio, F., & Trentin, G. (2015). Evaluating Vocational Educators' Training Programs: A Kirkpatrick-lnspired Evaluation Model. *Educational Technology*, 22-28.

55. Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*, 29(2), 35-53. https://doi.org/10.1002/nha3.20178

56. Rogers, P. J. (2000). Program theory: Not whether programs work but how they work. In Evaluation models (pp. 209-232). Springer, Dordrecht.

57. Rooholamini, A., Amini, M., Bazrafkan, L., Dehghani, M. R., Esmaeilzadeh, Z., Nabeiei, P., & Kojuri, J. (2017). Program evaluation of an integrated basic science medical curriculum in Shiraz Medical School, using CIPP evaluation model. *Journal of Advances in Medical Education & Professionalism*, 5(3), 148-154.

58. Rouse, D. N. (2011). Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspectives in Health Information Management*, 8, 1-5.

59. Schalock, R. L. (2001a). Outcome-based evaluation. Springer Science & Business Media.

60. Schalock, R. L. (2001b). An Overview of Outcome-Based Evaluation and Its Application. *Outcome-Based Evaluation*, 5-16.

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

62. Smidt, A., Balandin, S., Sigafoos, J., & Reed, V. A. (2009). The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, 34(3), 266-274. https://doi.org/10.1080/13668250903093125

63. Stufflebeam, D. (2001). Evaluation models. New Directions for Evaluation, 89, 7-98. https://doi.org/10.1002/ev.3

64. Stufflebeam, D. L. (2000). The CIPP model for evaluation. In Evaluation models (pp. 279-317). Springer, Dordrecht.

65. Stufflebeam, D. L., & Zhang, G. (2017). The CIPP evaluation model: How to evaluate for improvement and accountability. Guilford Publications.

66. Tam, M. (2014). Outcomes-based approach to quality assessment and curriculum improvement in higher education. *Quality Assurance in Education*.22(2), 158-168. https://doi.org/10.1108/QAE-09-2011-0059

67. Usun, S. (2016). A review on the program evaluation strategies in distance education. *International Journal on New Trends in Education & Their Implications*, 7(3), 33-45.

68. Walton, M. (2014). Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning*, 45, 119-126. https://doi.org/10.1016/j.evalprogplan.2014.04.002

69. Wu, B., Hu, Y., Gu, X., & Lim, C. P. (2016). Professional development of new higher education teachers with information and communication technology in Shanghai: A Kirkpatrick's evaluation approach. *Journal of Educational Computing Research*, 54(4), 531-562.