



Evaluation Model and Its Urgency on Elementary Education Programs

Suklani✉

Pendidikan Agama Islam, Institut Agama Islam Negeri Syekh Nurjati Cirebon, Indonesia

DOI: [10.31004/obsesi.v7i2.4201](https://doi.org/10.31004/obsesi.v7i2.4201)

Abstrak

Evaluation of an education program is a process of systematic activity to see how successful the planned education program is and to provide information to stakeholders in order to make a decision. The problem that often occurs is that the object of evaluation in education is still focused on the learning process, even though the evaluation includes all educational components and programs in a comprehensive manner. The purpose of this research is to describe program evaluation models and their urgency in evaluating educational programs. The novelty of this study is the implementation of various evaluation models in elementary education. The results of this study indicate that there are more than 50 program evaluation models, of which there are five program evaluation models that can be applied in evaluating educational programs, namely the Context, Input, Process, Product (CIPP), outcomes (CIPPO), Provus, Countenance (Stake), and Kirkpatrick evaluation models. The contribution is that with this evaluation information will be obtained about the achievements of the educational program, including its quality, effectiveness, efficiency and productivity, so that a decision can be made whether the program is worth continuing, revising, stopping or reformulating so that goals, objectives, processes can be found from the previous one.

Kata Kunci: *models; program-evaluation; elementary-education*

Copyright (c) 2023 Suklani

✉ Corresponding author : Suklani

Email Address: suklani@syekhnurjati.ac.id (Cirebon, Indonesia)

Received 7 December 2022, Accepted 1 March 2023, Published 19 March 2023

Introduction

Program evaluation is a systematic and objective process that collects, analyzes, and interprets information. Program evaluation is more specifically related to gathering and documenting information about specific aspects of the program to enable good decision-making (Böhm & Brun, 2008). The goal of high-quality program evaluation is to bring about significant improvements in program practice, implementation, and reproduction. The ultimate goal is to arrive at objective, definitive, valid, and intelligent conclusions regarding the achievement of objectives and answers to specific questions related to the program's overall effectiveness. Program evaluations are typically used to measure progress towards achieving goals, provide account information to stakeholders, improve program implementation, convince funding agencies about improving community support initiatives, and inform policy decisions (Lee Smith & Kelloway, 2016).

Conceptually, program evaluation is also concerned with gathering the methods, skills, and sensitivities required to determine the need for student or community services and the feasibility of using them, whether the services are intensive enough to meet unmet needs, and

whether the services offered fit into the plans and help people who need them at a reasonable cost (Dubey et al., 2020). Moreover, the benefits of program evaluation can be felt in various contexts. It can assist in developing a concrete understanding of the program's desired outcomes and personnel needs or promote an analysis of program effectiveness and cost efficiency. In addition, the scope of program evaluation has been broadened to cover more complex issues, no longer focusing solely on building causal relationships between expectations and results achieved. Although program evaluation varies in its implementation style, it should provide a basis for valid comparisons between similar programs (Wang, 2009).

Specifically, the evaluation of elementary education programs is a process of systematically extracting, discovering, and determining information about the planning, goals, values, uses, and effectiveness of education according to predetermined criteria and objectives (Parhan et al., 2020). The purpose of evaluating elementary education programs is to provide recommendations to interested parties as material for consideration in making decisions about the programs implemented (Casta et al., 2021). According to the program evaluation results, its use is to decide whether the program is worth continuing, revising, or terminating because it is no longer useful. Supposedly, the evaluation of elementary education programs is familiar and used to be routinely carried out in schools or other educational institutions (Kultsum et al., 2022). The evaluation is not only an assessment of one aspect but also comprehensively covers various aspects of the program. In addition, evaluation can also measure the achievements of each implemented program. Program evaluation can be applied to classroom learning activities, policy evaluation, planning evaluation, process evaluation, outcome evaluation, development evaluation, and others (Lee Smith & Kelloway, 2016).

Nevertheless, the problem that often occurs is that evaluation is still understood in education, only limited to assessment, which is carried out in a formative and summary manner. In fact, the assessment is only one part of the evaluation and supervision of education, is more of an individual student character, and does not comprehensively cover programs and institutions (Maesyaroh et al., 2020). Furthermore, if an assessment has been carried out, it is usually considered to have carried out an evaluation. Understanding this definition is incorrect. If it is the case, the assessment implementation tends to be seen only from learning outcomes. In reality, the success of educational processes and outcomes is not only due to one or two aspects but many factors influencing it, including one of them, in the elementary education program. Thus, the evaluation of elementary education programs not only deals with measurement and assessment based on the completion of questions but also thoroughly examines many factors (Nirmala et al., 2018).

Therefore, evaluation of elementary education programs really needs to be understood and disseminated to all educators and stakeholders because program evaluation is crucial in developing the quality of education (Munthe, 2015). Apart from the emergence of these problems, it is not uncommon for both stakeholders and evaluators to be confused about the selection of program evaluation models that are more appropriate and in accordance with the type of program, objectives, intricacies of the problem, and existing conditions in an educational institution or other institutions (Gadsden et al., 2016). Hence, the main study in this paper is about the meaning, urgency, principles, goals, classification, and evaluation models of elementary education programs. The general objective of writing this paper is to find out and examine what models are appropriate to choose and implement in evaluating elementary education programs so that they are in accordance with the objectives to be achieved.

Several recent studies such as those conducted by Effendi et al. (2022) regarding the implementation of the evaluation model Kirkpatrick concluded that at the reaction level, students showed a level of satisfaction with the suitability of the material with the module. At the learning level, students are satisfied with the relevance of the material in everyday life. At the behavior level, students are satisfied with the tutorial process and at the result level, the majority of students feel that their knowledge and experience have increased after following

the tutorial. (Stewart et al., 2022) strengthened that program evaluation is important for assessing the implementation and outcomes of local, state, and federal programs. The Program Evaluation Toolkit provides tools and resources to support individuals responsible for evaluating and monitoring local, state, or federal programs. The

Other researchers argue that in congruency, evaluation at the stage of antecedents that education planning at school in the category is in accordance with the applicable education planning standards, at the stage of transactions (Harjanti et al., 2019). Anh (2018) provides readers with comprehensive discussions on the four well-known evaluation models in education: Tyler's objective model, Stake's responsive model, Scriven's goal free model and Stufflebeam's CIPP model. These models have a long history and have been thoroughly developed over time. The application of these four models is found in many fields of evaluation, but mostly in educational program evaluation.

Since the number of program evaluation models is very large, with more than 50 models, evaluators sometimes feel confused about what model is most appropriate to apply in evaluating programs in an educational institution, where these evaluation models have characteristics that differ from one model to another (Şeker, 2020). On this basis, it is expected that the study of this paper can become one of the scientific solutions to overcome the lack of attention of the stakeholders of an educational institution to the importance of evaluating elementary education programs, selecting the appropriate model, and implementing it properly, continuously, and sustainably so that the quality of education in these institutions are always maintained and even continue to be better in the future (Mansir et al., 2020).

Methodology

This paper used a library research method with a qualitative approach. This library research was carried out by examining or reviewing the literature of books, scientific journals, articles, dictionaries, documents, or other literature, related to the research topic (Cohen et al., 2005). It is called library research because the materials and data needed in this research come from the library in dictionaries, encyclopedias, books, journals, documents, magazines, newspapers, and others (Cohen et al., 2005).

The source of this research data consisted of the primary data needed by researchers. The primary data was literature about program evaluation models and their urgency in evaluating elementary education programs and their analysis according to studies, theories, and opinions of competent education and learning experts or scientists (Gardner et al., 2011). Other sources were supporting data, such as books, journals, articles, or literature related to the research title. In collecting the data, library techniques were employed, i.e., data collection techniques from several trusted libraries by collecting, analyzing, and reviewing concepts, data, theories, discussions, or information sourced from books, Accredited journals, articles, notes, or reports related to the problem under study (Flick, 2013).

Furthermore, the data analysis technique utilized was the Miles and Huberman model analysis technique, including 1) data collection from several researched literature according to the research title, 2) data reduction by summarizing data, choosing what is important, new, and unique, and categorizing not important data to be ignored, 3) displaying data in the form of brief descriptions, making relationships between categories, and others; graphs, matrices can also support data display, and others, 4) drawing conclusions, where initial conclusions are temporary and will change if no strong evidence is found (Sugiyono, 2019). However, if the initial conclusion is supported by valid and consistent evidence, the conclusion is credible conclusion (Qu & Dumay, 2011).

Results and Discussion

Program Evaluation Models

CIPP model program evaluation

Evaluation of elementary education programs has a long history. Madaus and Stufflebeam mentioned that program evaluation efforts began at the end of the 18th century. The practice grew rapidly in the 1940s, 1950s, and 1960s. Moreover, an evaluation model should follow evaluation principles that are theoretically sound, easy to use, standards-based, and practical. Each program evaluation model also has its characteristics, each emphasizing certain aspects of the program being evaluated. Stufflebeam and Zhang further stated that the Context, Input, Process, Product (CIPP) model is one of many existing approaches to program evaluation.

Other program evaluation models have particular limitations that hinder their applicability in some educational contexts, for example, experimental designs. Thus, the CIPP model was created explicitly to overcome this limitation. The CIPP model is also a comprehensive framework for guiding the formative and summative evaluation of programs, projects, personnel, products, institutions, and systems. In addition, the CIPP model is designed to provide timely information in a systematic way for decision-making and establishes the applicability of the CIPP model for summative assessment and accountability. Unlike many other evaluation models, the CIPP model was created for formative assessments and extended only to summative assessments. This evaluation model is flexible and not strictly defined based on a prepared design. For instance, although input evaluations often include staffing plans, the inclusion of these plans is not an absolute requirement (Olasina, 2017). The timing of this model is also flexible and can be applied either during or after specific processes (Sopha & Nanni, 2019).

The CIPP model developed by Stufflebeam is an acronym for context, input, process, and product. These four words are the objectives and evaluation components of the process of an activity program. The advantages and uniqueness of the CIPP evaluation model are that it is more comprehensive than other evaluation models since the object of evaluation is not only results but includes context, input, process, and results. In addition to having advantages, this evaluation model also has limitations. Among other things, the application of this model in classroom learning programs has a less elevated level of implementation if there are no modifications (Ayertei, 2018). Besides, the starting point of the CIPP evaluation model rests on the view that the success of an elementary education program is influenced by many factors, including the characteristics of the students, the environment, program objectives, the equipment used, and the procedures and mechanisms for implementing the program itself (Karim et al., 2020).

CIPPO Model Program Evaluation

The context, input, process, product, and outcomes (CIPPO) evaluation model is widely known as a development of Stufflebeam's CIPP model. The CIPP model views the program as a system. Thus, if evaluators have decided to use this model, they will have to analyze the program based on its components. Since the CIPP model only stops at measuring output (product), this CIPP model has been developed a lot so that this model is now refined with one component, 'O,' which stands for the outcome, to become CIPPO. In other words, CIPPO is a modification of the CIPP model. According to Jaedun (2010), to evaluate a program, it is also necessary to evaluate the impact, benefits, expectations of change, or outcome, i.e., how successful a graduate is in his workplace after graduating from an educational institution. For example, to evaluate the education and training program, in addition to the four components of context, input, process, and product, an evaluation of the impact or outcome (O) is also needed, namely how successful graduates are in society and the workplace (Moore, 2002).

CIPPO components are described in Jaedun (2010) as follows: 1) Context evaluation attempts to describe and detail the environment, unmet needs, population and sample served, and program objectives, 2) Input evaluation is the initial ability of students and schools to support the program, including the ability of schools to provide the right officers and others (Suzana et al., 2021). Questions related to input lead to solutions to problems that drive the program implementation in question. As Jaedun (2010) mentioned, input evaluation aims to provide information, determining how available resources can support program implementation, 3) Process evaluation refers to how far the activities carried out in the program are in accordance with the plan, who is the person in charge of the program, and when will the activities' implementation end. Process evaluation, according to Stufflebeam, is an ongoing check on the plan implementation, 4) Product evaluation aims to measure, interpret, and assess program achievements. According to Stufflebeam, product evaluation is directed at things that show changes that occur in raw inputs. Finally, 5) outcome evaluation refers to the output results from program implementation, and how much the program benefits teachers in carrying out the learning and assessment process in class.

Provus model program evaluation (discrepancy model)

According to Darodjat & M (2015), the model developed by Malcolm Provus departs from the assumption that to determine the feasibility of a program, evaluators can compare standards with performance, meaning what should happen with what actually happened. By comparing these two things, a gap can be seen, namely, the standard set with the actual performance. This model is intended to analyze a program, and whether the program is worth continuing, improving, or stopping altogether. This model also refers to standard, performance, discrepancy in detail, and measurability. In this case, the evaluator measures how big the gap is in each component of the program evaluation. Recognizing the gaps in each program component allows corrective steps to be carried out clearly (Lee Smith & Kelloway, 2016).

Additionally, Ananda & Rafida (2017) asserted that the program discrepancy evaluation model initiated by Provus is carried out to know the degree of conformity between the standards set in the program and the performance that occurs from the program. The standards here are benchmarks formulated and set in the hope that there will be effective results, whereas performance is the real procedures and results that appear when the program is implemented (Heimann, 2008). Meanwhile, the steps that can be taken in the Provus evaluation model are: 1) Conduct the preparation of the evaluation design. At this stage, activities are performed, including formulating the objectives of the evaluation program, preparing clients, staff, and other equipment, and formulating standards with measurable formulations. Usually, in this step, the evaluator consults with the program developer. 2) Equipment installation. At this stage, whether the existing equipment is in accordance with the requirements must be known. Activities that must be carried out in this stage include reviewing standard setting, reviewing ongoing programs, and examining gaps between plans and results achieved, 3) Process. At this stage, an assessment is made of which goals have been achieved. In this case, the evaluator collects data from the program implementation process, 4) Product measurement. At this evaluation stage, data analysis is conducted, and the output level achieved is determined. The question that can be posed at this stage is: Has the program reached its final goal or not?, 5) Comparison. The results obtained are compared with the goals set at this evaluation stage. In addition, at this stage, the evaluator records all findings about the gap to be submitted to the decision maker to decide on the follow-up of the program (Lee, 2009).

Stake Model Program Evaluation (Countenance Model)

This model was developed by Robert E. Stake of the University of Illinois. Stake emphasizes two basic activities in evaluation: description and judgment, i.e., what the

program aims to achieve and what actually happened. Komaneet et al. (2013) also describe the stages of evaluating the Stake model program: antecedents, transactions, and outcomes. Antecedents refer to information related to situations and events before program implementation that may be closely related to outcomes. This information is related to the teacher's background, the appropriate curriculum, the availability of resources, and previous teaching and learning activities that are also related to outcomes, such as whether students have had breakfast before going to school, whether students have done their homework, and whether students sleep at night enough.

At the transaction stage, questions are asked: what actually happened during the program implementation; is the program being implemented in accordance with the program plan? This evaluation stage includes information on processes or activities experienced by students in their interactions with teachers, parents, counselors, tutors, and other students (Afnan et al., 2020; Syabibi et al., 2021). In addition, Stake recommends that evaluators act wisely in the evaluation process to see the actualization of the program evaluation model. Meanwhile, outcomes are related to what has been obtained with the program, and whether the program implemented is in accordance with what is expected, including abilities, achievements, attitudes, and goals (Wahyudhiana, 2015).

The Stake evaluation model is known as the overall model and is also called the consideration evaluation model. It is said to be a consideration evaluation model since the evaluator considers comparing the results of the program evaluation and the results of other programs and comparing the results of program implementation and predetermined standards. The Stake's countenance model aims to develop a curriculum assessment plan. Stake also emphasizes the role of the evaluator in developing curriculum goals into measurable, specific goals ([DBE], 2011). The countenance model further consists of two matrices: first, description is a matrix comprising planning and observation categories, and second, a judgment matrix covering standard and consideration categories. Each category has three focuses: pre-program situation, program process, and program results.

The steps to be taken in using the Stake evaluation model are as follows: a) Data collection. Evaluators collect data relating to initial conditions, transactions, and outcomes through document studies or interviews. Before data collection, evaluators create terms of reference relating to antecedents, transactions, and outcomes. b) Data analysis. Data analysis is carried out by logical and empirical analysis. Logical analysis is conducted to provide considerations regarding the relationship between antecedents, transactions, and results. Furthermore, empirical analysis works the same principle as logical analysis, but empirical data is used. c) Congruence analysis. Congruence analysis compares what is formulated in the objectives with what happens in the field (observation). d) Judgment of results. Furthermore, the evaluator's task is to provide consideration regarding the program being studied (Ananda & Rafida, 2017; Huther & Shah, 2000).

Kirkpatrick model program evaluation

Kirkpatrick's model is an evaluation model developed by Kirkpatrick (1998) in Komaneet et al. (2013) and is also called the "four levels" or Kirkpatrick's evaluation. Evaluations carried out with this system include four evaluation levels: evaluation of reactions, learning, behavior, and results. The first stage: the reaction is evaluating the participant's response to learning in the form of feelings, thoughts, desires, assistance, and implementation of the learning environment. The second stage: learning, measures the learning transfer process. At this stage, learning refers to how participants change attitudes, increase knowledge, or improve skills due to following the program (Karim & Hartati, 2020).

The third stage: behavior, evaluates the degree of change in behavior caused by the participant's participation in the program. Assessment at this stage will also explain the differences between what they have learned before and after this program (Karim et al., 2022). The fourth stage: the level of performance/result, aims to measure the extent to which the

curriculum has helped achieve the previously set goals and what has been achieved as the final result after the learning process (Effendi et al., 2022).

Moreover, Kirkpatrick (1998)'s evaluation model has undergone several improvements, most recently updated in 1998, which is known as Evaluating Training Programs: The Four Levels or Kirkpatrick's Evaluation Model. Evaluation of a training program includes four levels of evaluation: a) reaction, b) learning, c) behavior, and d) results. In Wahyudhiana (2015), it is explained about the four levels of evaluation, which are the sequence of ways to evaluate this Komanee et al. (2013) model as follows: 1) Reaction Evaluation. Komanee et al. (2013) explained that evaluating the training participants' reaction means measuring the participants' response and satisfaction. The training program is considered effective if the training process is enjoyable and satisfies the participants according to their needs so that they are interested and motivated to continue learning and practicing because people will learn better when they respond positively to the learning environment (Suzana et al., 2020). Conversely, if the participants are unsatisfied and do not feel the need, they will not be motivated to participate in the next training. Participant satisfaction can be assessed from several aspects, including the material provided, the facilities available, the strategies and methods of delivery of the material used, learning media, activity schedules, and the menu and consumption presentation provided (Jansen et al., 2019), 2) Learning evaluation. According to Komanee et al. (2013), evaluation of learning outcomes can be seen in changes in attitude, knowledge improvement, and/or skill improvement of participants after completing the program. Program participants are said to have learned when they have experienced a change in attitude, improved knowledge, and increased skills. If there is no change in the three domains, the program can be said to have failed. Evaluation of this model is also called the evaluation of learning outcomes, 3) Behavior Evaluation. This assessment is also called the evaluation of outcomes and training activities. This evaluation emphasizes changes in behavior when students return to the workplace after participating in the program concerning aspects of knowledge, attitudes, and skills. According to Komanee et al. (2013), behavior evaluation can be done by (1) comparing the behavior of the control group with the behavior of program participants, (2) comparing behavior before and after participating in the program, and (3) surveys or interviews with superiors, subordinates, trainers, or with those who others after returning to work, 4) Evaluation of Results. The focus of evaluation at this stage is the final result that occurs after the participant has joined a program. The way to evaluate the final results, according to Komanee et al. (2013), is to (1) compare the control group with the program participant group, (2) measure performance before and after attending the training, and (3) compare the costs incurred with the benefits obtained and how to improve or improve going forward.

This model's strengths and advantages are its simplicity and comprehensive nature, and it can be applied in various learning and training situations. This model can be applied to evaluate learning programs in schools and even at a smaller level, such as the class level and a particular program. This model can be applied in schools because 1) the pressure point is the same, i.e., in the learning process, and 2) the desired changes in learning outcomes are equally directed at aspects of knowledge, attitudes, and skills or prowess. However, if this evaluation model is applied in learning programs, it must be modified and adapted to school settings (Gurr, 2015).

The Urgency of Elementary Education Program Evaluation

Program evaluation is a crucial stage as part of the policy and program implementation process, i.e., as a stage in providing information about policies and programs that are currently or have been implemented. In practice, evaluation requires evaluation theories that can be identified with a prescriptive approach or theoretical model. The importance of evaluation, including evaluation of educational programs, is because information can be obtained about programs that have been achieved and those that have not been achieved. With evaluation,

the quality, effectiveness, and performance efficiency or productivity of an educational institution can also be seen in implementing programs and improving them.

Furthermore, the importance of evaluation activities is to review the extent to which educational programs and policies can be realized effectively in accordance with the planning and expected results and to what extent the things that have been achieved can meet predetermined criteria (Cera et al., 2014). Hence, based on the results of this evaluation, a decision can be made on whether these programs and policies will be continued, postponed, revised, discontinued, or reformulated so that new goals, objectives, and alternatives can be found that is different from those before (Hajaroh, 2018).

Why is an evaluation a vital process in education and learning activities? Some answers include: 1) Evaluation for teachers can determine the effectiveness of their performance during the teaching process. 2) Evaluation can provide information to improve the ongoing curriculum or determine the clarity of specific goals to be achieved. 3) Evaluation activities can encourage students to improve the learning process further. 4) Evaluation is an integral part of the learning process. 5) Appropriate evaluation can determine the program's effectiveness and students' success in learning activities. 6) For program designers and owners, evaluation can be used to make decisions about whether the program needs to be improved or continued and which parts are considered to have deficiencies and weaknesses that need to be improved. 7) Evaluation is an essential tool that provides feedback for students, teachers, parents, curriculum developers, graduate users, educational policymakers, and the community. 8) Evaluation is a crucial tool to determine students' success in mastering the goals set. Last, 9) individual students can utilize information on evaluation results in determining the future related to their field of work or later career (Maor & Mcconney, n.d.; Sanjaya, 2008).

From some explanations above, the CIPP model program evaluation objectives can be illustrated in the following chart:

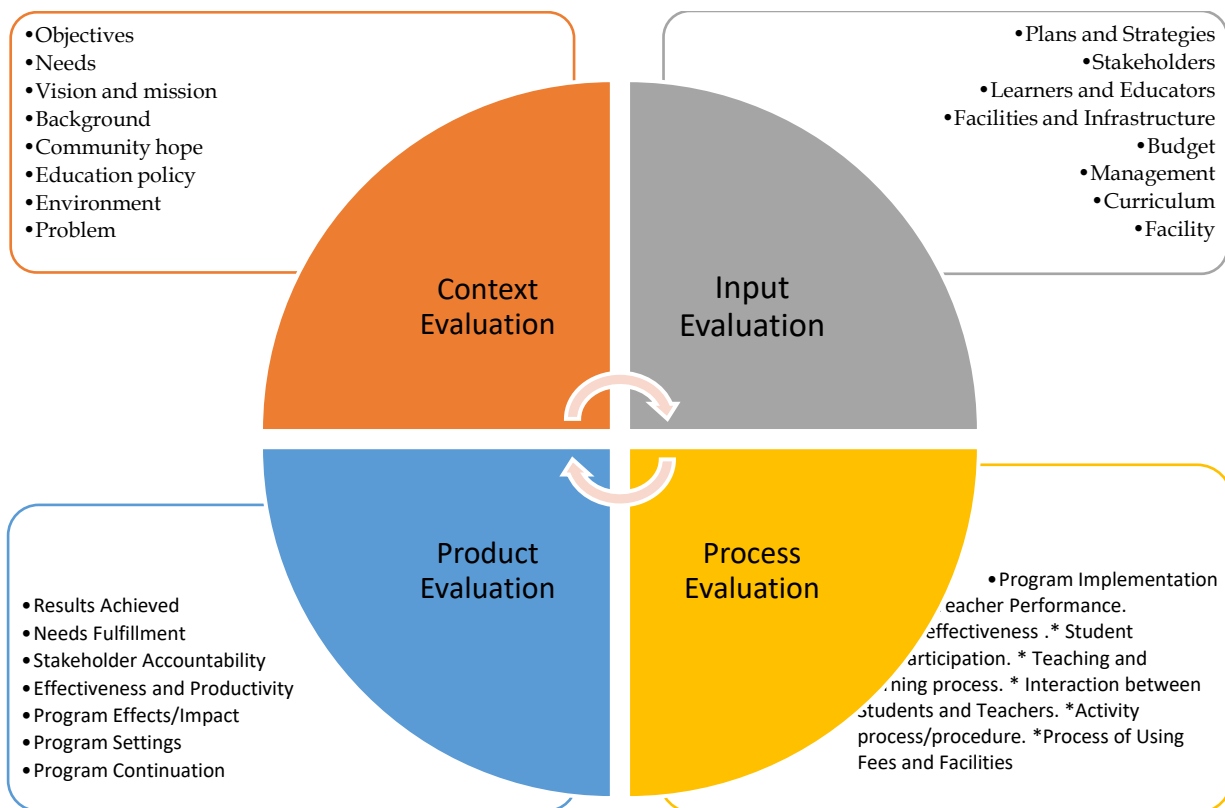


Figure 1. CIPP Model Evaluation Process

Source: Data is proceed in 2023

One of the activities in evaluating educational programs is evaluating the learning process. Evaluation of teaching and learning programs is essential from elementary to higher education levels to ensure the quality of student learning, teachers' professionalism, and teaching quality. Activities that are also the object of evaluation are teaching activities and the overall learning process, which is carried out both inside and outside the classroom. The object of evaluation of teaching activities includes teaching coordination and management, dissemination of teaching methods, learning and assessment activities, and revision and improvement of procedures.

These procedures are conducted in response to the training objectives and competencies students wish to develop. Evaluation of the teaching activity must also pay attention to all the procedures carried out and evaluate them quantitatively and qualitatively (Luria et al., 2014). Moreover, the teacher's evaluation of teaching activities is internally carried out to ensure that the teaching objectives have been achieved. This evaluation is based on the desire to create an organizational culture that is principally based on the dedication and achievements of the participants within it (Cano-Hurtado, 2009).

Conclusion

The research results indicate more than 50 program evaluation models. Each of these evaluation models has its characteristics that differ from one model to another so that one model will be more appropriate to be applied to one evaluation activity with a specific purpose than other evaluation models. In this regard, evaluation models of elementary education programs, in terms of their aims and objectives, are classified into six groups: 1) Goal-oriented evaluation models, with the characters being Bloom and Provus. 2) Decision-oriented evaluation model, with the main character being Stufflebeam. 3) Transactional evaluation model, with the characters being Stake and Rippey. 4) Evaluation research model, with the characters being Campbell and Colley. 5) The goal-free evaluation model, with Sriven as the main character. Lastly, 6) the adversary evaluation model, with the characters being Levine and Owens. Meanwhile, models that evaluators can use as a reference in evaluating elementary education programs include the CIPP Model (Stufflebeam), the CIPPO model, the Provus model (Discrepancy model), the Stake model (Countenance model), the Kirkpatrick model (Four-level evaluation model), and others. The selection and use of an evaluation model from these models fundamentally depend on the program's purpose, type, and characteristics, the program owner's policy, the evaluator's ability, and for whom the evaluation is carried out.

Acknowledgment

We thank the Chancellor, the Dean, and all the lecturers under the department of Pendidikan Agama Islam at IAIN Shekh Nurjati Cirebon, as well as the family and colleagues who are willing to provide access and support this research so that it become the output/publication.

Reference

- Afnan, D., Karim, A., Irfan, A., Rianto, A., Wildanu, E., Hidayat, R., Aziza, D. K., Kardiyanti, E. N., Rengganis, A., Kamaluddin, M., & Mar'ati, A. (2020). Phubbing and social interaction: An analysis of smartphone usage in higher education. *Journal of Advance Research in Dynamical & Control Systems*, 12(6), 2325–2334.
- Ananda, R., & Rafida, T. (2017). *Pengantar Evaluasi Program Pendidikan*. Perdana Publishing.
- 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>
- Ayertei, T. G. (2018). Effects of Classroom Attendance and Learning Strategies on the Learning Outcome. *Journal of International Education in Business*, 11(2), 195–219.

<https://doi.org/https://doi.org/10.1108/JIEB-01-2017-0004>

- Böhm, G., & Brun, W. (2008). Intuition and affect in risk perception and decision making. *Judgment and Decision Making*, 3(1), 1–4. <https://doi.org/10.1017/S1930297500000115>
- Cano-Hurtado, J. . (2009). *An Evaluation Model Of The Teaching Activity Of Academic Staff*. Valencia University of Technology.
- Casta, C., Rohidi, T. R., Triyanto, T., & Karim, A. (2021). Production of aesthetic tastes and creativity education of Indonesian glass painting artists. *Harmonia: Journal of Arts Research and Education*, 21(2), 266–277. <https://doi.org/10.15294/harmonia.v21i2.30348>
- Cera, R., Mancini, M., & Antonietti, A. (2014). Relationships between metacognition, self-efficacy and self-regulation in learning. *ECPS-Educational, Cultural and Psychological Studies*, 7, 115–141. <https://doi.org/10.7358/ecps-2013-007-cera>
- Cohen, L., Manion, L., & Morrison, K. (2005). *Research Methods in Education* (5th Editio). : Taylor & Francis e-Library.
- DBE. (2011). *Curriculum and Assessment Policy Statement 10–12 (General): Mathematical Literacy*. Department of Education.
- Dubey, S., Biswas, P., Ghosh, R., Chatterjee, S., Dubey, M. J., Chatterjee, S., Lahiri, D., & Lavie, C. J. (2020). Psychosocial impact of COVID-19. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 14(5), 779–788. <https://doi.org/10.1016/j.dsx.2020.05.035>
- Effendi, Zainuddin, & Ahmad. (2022). Implementasi evaluasi model kirkpatrick terhadap kualitas sistem pembelajaran. *Jurnal Educatio: Jurnal Pendidikan Indonesia*, 8(1), 1–8. <https://doi.org/10.29210/1202221160>
- Flick, U. (2013). *Qualitative Data Analysis* (the SAGE h). SAGE Publication Ltd.
- Gadsden, V. L., Ford, M., & Breiner, H. (2016). Parenting matters: Supporting parents of children ages 0-8. In *Parenting Matters: Supporting Parents of Children Ages 0-8*. <https://doi.org/10.17226/21868>
- Gardner, W. L., Coglisier, C. C., Davis, K. M., & Dickens, M. P. (2011). Authentic leadership: A review of the literature and research agenda. *The Leadership Quarterly*, 22(6), 1120–1145. <https://doi.org/10.1016/j.leaqua.2011.09.007>
- Gurr, D. (2015). *A Model of Successful School Leadership from the International Successful School Principalship Project*. 136–150. <https://doi.org/10.3390/soc5010136>
- Hajaroh, M. (2018). *Fondasia: Kajian Ilmiah Fondasi Pendidikan*. Forum Kajian Fondasi Pendidikan (FOKSiP) UNY Yogyakarta.
- Harjanti, R., Supriyati, Y., & Rahayu, W. (2019). Evaluation of learning at elementary school level of “Sekolah Alam Indonesia (SAI)”: Evaluative research using countenance Stake’s model. *American Journal of Educational Research*, 7(2), 125–132. <https://doi.org/10.12691/education-7-2-2>
- Heimann, G. (2008). *Corporate Social Responsibility Global Standards & Policies in Practice*. The Liberian International Ship.
- Huther, J., & Shah, A. (2000). *Anti-Corruption Policies and Programs: A Framework for Evaluation* (No. 2501; Policy Research Working Paper). <https://documents1.worldbank.org/curated/en/578241468767095005/pdf/multi-page.pdf>
- Jaedun, A. (2010). *Metode Penelitian Evaluasi Program*.
- Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review*, 28, 100292. <https://doi.org/10.1016/j.edurev.2019.100292>
- Karim, A., Bakhtiar, A., Sahrodi, J., & Chang, P. H. (2022). Spiritual leadership behaviors in religious workplace: The case of pesantren. *International Journal of Leadership in Education*, 00(00), 1–29. <https://doi.org/10.1080/13603124.2022.2076285>
- Karim, A., Faiz, A., Parhan, M., Gumelar, A., Kurniawaty, I., Gunawan, I., Wahyudi, A. V., &

- Suanah, A. (2020). Managerial leadership in green living pharmacy activities for the development of students' environmental care in elementary schools. *Journal of Critical Reviews*, 7(13), 714–719. <https://doi.org/10.31838/jcr.07.13.125>
- Karim, A., & Hartati, W. (2020). Spiritual tasks of teachers in higher order thinking skills-oriented learning. *International Journal of Psychosocial Rehabilitation*, 24(8), 4568–4580. <https://doi.org/10.37200/IJPR/V24I8/PR280474>
- Kirkpatrick, D. L. (1998). *Evaluating Training Programs: The Four Levels*. Berrett-Koehler Publisher, Inc.
- Komanee, S., Hoxsuwan, S., & Phusee-on, S. (2013). The contemplative education learning environment management model for undergraduate students. *The Journal of Global Business Management*, 9(3), 16–24. <http://www.jgbm.org/page/3>
- Kultsum, U., Parinduri, M. A., & Karim, A. (2022). Comparative studies between public and private Islamic schools in the era of globalization. *International Journal of Evaluation and Research in Education (IJERE)*, 11(1), 421–430. <https://doi.org/10.11591/ijere.v11i1.22182>
- Lee, H. J. (2009). Peer evaluation in blended team project-based learning; What do students find important? *Educational Technology & Society*, 15, 214–224. <https://www.jstor.org/stable/jeductechsoci.15.4.214>
- Lee Smith, S., & Kelloway, E. K. (2016). Respect in the workplace: an evaluation of a short online intervention program. *Journal of Organizational Effectiveness*, 3(4), 395–410. <https://doi.org/10.1108/JOEPP-02-2016-0008>
- Luria, G., Kalish, Y., & Weinstein, M. (2014). Learning disability and leadership: Becoming an effective leader. *Journal of Organizational Behavior*, 35(6), 747–761. <https://doi.org/10.1002/job.1896>
- Maesyaroh, S., Jannah, W. N., & Karim, A. (2020). Penerapan program Lisasbil (lihat sampah ambil) dalam usaha penguatan karakter siswa sekolah dasar di lingkungan SDN 1 Kamarang. *Seminar Nasional Pendidikan Dasar*, 61–69. <http://e-proceedings.umpwr.ac.id/index.php/semnaspgsd/article/view/1371>
- Mansir, F., Karim, A., Tumin, & Purnomo, H. (2020). Transforming Islamic Boarding School as Indonesian of Islamic Educational Institution in the Digital Era. *International of Psychosocial and Rehabilitation*, 24(8), 8777–8788. <https://doi.org/10.37200/IJPR/V24I8/PR280874>
- Maor, D., & Mcconney, A. (n.d.). Wisdom of the elders: mentors' perspectives on mentoring learning environments for beginning science and mathematics teachers. *Learning Environments Research*, 18. <https://doi.org/10.1007/s10984-015-9187-0>
- Moore, A. B. (2002). Community Development Practice: Theory in Action. *Journal of the Community Development Society*, 33(1), 20–32. <https://doi.org/10.1080/15575330209490140>
- Munthe, A. P. (2015). Pentingnya Evaluasi Program Di Institusi Pendidikan. *Jurnal Scholaria*, 5(2), 1–14. <https://ejournal.uksw.edu/scholaria/article/view/13>
- Nirmala, S. D., Rahman, R., & Musthafa, B. (2018). Students' elementary literacy skill and critical thinking skill supported by school literacy program (SLP). *Journal of Teaching and Learning in Elementary Education (JTLEE)*, 1(2), 92. <https://doi.org/10.33578/jtlee.v1i2.5880>
- Olasina, G. (2017). An evaluation of educational values of YouTube videos for academic writing An Evaluation of Educational Values of YouTube Videos for Academic Writing. *The African Journal of Information System*, 9(4). <https://digitalcommons.kennesaw.edu/ajis/vol9/iss4/2>
- Parhan, M., Faiz, A., Karim, A., Nugraha, R. H., Subakti, G. E., Rindu, M., Islamy, F., Budiyantri, N., Fuadin, A., & Tantowi, Y. A. (2020). Internalization Values of Islamic Education at. *International Journal of Psychosocial Rehabilitation*, 24(08), 14778–14791.
- Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. In *Qualitative Research in*

- Accounting and Management*. <https://doi.org/10.1108/11766091111162070>
- Sanjaya, W. (2008). *Perencanaan dan Desain Sistem Pembelajaran*. Kencana.
- Şeker, H. (2020). Elementary and middle school students' school attitudes and extracurricular activities. *Journal of Elementary Education*, 13(3), 347–364. <https://journals.um.si/index.php/education/article/view/579>
- Sopha, S., & Nanni, A. (2019). The CIPP Model Applications in Language Program Evaluation. *The Journal of Asia TEFL*, 16(4), 1360–1367. <https://doi.org/10.18823/asiatefl.2019.16.4.19.1360>
- Stewart, J., Joyce, J., Haines, M., Yanoski, D., Gagnon, D., Luke, K., Rhoads, C., & Germeroth, C. (2022). *Program Evaluation Toolkit: Quick Start Guide*. National Center for Education Evaluation and Regional Assistance at IES.
- Sugiyono. (2019). *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, R&D dan Penelitian Pendidikan)*. Alfabeta.
- Suzana, S., Karim, A., Amanah, A., & Munajim, A. (2021). Bermain Kognitif Matematika Anak Melalui Pemanfaatan Barang Bekas Pada Masa Pandemi Covid-19. *Jurnal Pendidikan Anak Usia Dini Undiksha*, 9(2), 1–15. <https://doi.org/http://dx.doi.org/10.23887/paud.v9i2.34259>
- Suzana, S., Munajim, A., Casta, C., Pratama, G., Sulaeman, E., Sukarnoto, T., Ridwan, M., & Karim, A. (2020). Gadget and the internet for early childhood distance learning. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 8019–8028. <https://archives.palarch.nl/index.php/jae/article/view/3518>
- Syabibi, M. R., Karim, A., Kulkarni, S., & Sahil, A. (2021). Communicative cultural dakwah of Abdurrahman Wahid in pluralistic society. *Karsa: Journal of Social and Islamic Culture*, 29(2), 1–33. <https://doi.org/10.19105/karsa.v29i2.5220>
- Darodjat., & M, W. (2015). Model Evaluasi Program Pendidikan. *Jurnal Islamadina*, 14(1), 1–28. <https://jurnalnasional.ump.ac.id/index.php/ISLAMADINA/article/view/1665>
- Wang, V. C. . (2009). *Assessing and Evaluating Adult Learning in Career and Technical Education*. Zhejiang University Press.