Evaluation Model of Islamic Religious Education Learning Program at Madrasah 'Aliyah Level in East Java

Balya Ziaulhaq Achmadin[1*], Indah Aminatuz Zuhriyah[2], Ahmad Barizi[3]
Universitas Islam Negeri Maulana Malik Ibrahim Malang[1,2,3]
230101210004@student.uin-malang.ac.id [1,], zuhriyah@pgmi.uin-malang.ac.id [2] , abarizi_mdr@uin-malang.ac.id [3]
Koresponden*

Abstract : *Evaluation of Islamic Religious Education learning programs in madrasahs is an important component to ensure the quality and effectiveness of education implementation. However, the practice of evaluating Islamic Education programs in Madrasah 'Aliyah is often partial and limited to assessing students' learning outcomes. In fact, there are various program evaluation models that can be applied to produce a comprehensive evaluation. This study aims to identify PAI learning program evaluation models that can be applied in Madrasah 'Aliyah, analyze their suitability and implementation to improve the quality of PAI learning in Madrasah 'Aliyah. This research used a case study method by examining PAI learning programs in several Madrasah 'Aliyah in East Java. Data were collected through interviews, observations, and document reviews. Data analysis was conducted qualitatively. The results of this study identified several evaluation models of PAI learning programs that can be applied in Madrasah 'Aliyah, focusing on goal-based evaluation, stakeholder-based evaluation, and responsive evaluation. The models were evaluated based on their suitability to the context, scope, and availability of resources.*

Keywords : *Program Evaluation Model, Evaluation Principles, Islamic Education Learning, Madrasah 'Aliyah, East Java*

Abstrak : Evaluasi program pembelajaran Pendidikan Agama Islam di madrasah merupakan komponen penting untuk menjamin kualitas dan efektivitas pelaksanaan pendidikan. Namun, praktik evaluasi program PAI di

Madrasah 'Aliyah seringkali masih bersifat parsial dan terbatas pada penilaian hasil belajar peserta didik. Padahal, terdapat beragam model evaluasi program yang dapat diterapkan untuk menghasilkan evaluasi yang komprehensif. Penelitian ini bertujuan untuk mengidentifikasi model-model evaluasi program pembelajaran PAI yang dapat diterapkan di Madrasah 'Aliyah, menganalisis kesesuaian dan implementasi untuk meningkatkan kualitas pembelajaran PAI di Madrasah 'Aliyah. Penelitian ini menggunakan metode studi kasus dengan mengkaji program pembelajaran PAI di beberapa Madrasah 'Aliyah di Jawa Timur. Data dikumpulkan melalui wawancara, observasi, dan telaah dokumen. Analisis data dilakukan secara kualitatif. Hasil penelitian ini mengidentifikasi beberapa model evaluasi program pembelajaran PAI yang dapat diterapkan di Madrasah 'Aliyah, berfokus evaluasi berbasis-tujuan, evaluasi berbasis-pemangku kepentingan, dan evaluasi responsif. Model-model dievaluasi berdasarkan kesesuaian dengan konteks, cakupan, dan ketersediaan sumber daya.

Kata Kunci : Model Evaluasi Program, Prinsip Evaluasi, Pembelajaran PAI, Madrasah 'Aliyah, Jawa Timur

INTRODUCTION

In education, learning evaluation is an important part of the process of developing and improving the quality of education. Learning evaluation aims to measure and evaluate student achievement, the effectiveness of teaching methods, and the suitability of the curriculum used. By conducting a good learning evaluation, we can identify strengths and weaknesses in the learning process, so that needed improvements and enhancements can be made. Program evaluation is a very basic activity for micro curriculum development in this case program evaluation. Evaluation, which is often understood in education, is limited to assessment. This assessment is carried out formatively and summatively. When an assessment is carried out, it is considered to have conducted an evaluation. This understanding is not very precise. The implementation of assessment tends to only look at the achievement of learning objectives. In fact, in the educational process it is not only the value that is seen, but there are many factors that make a program successful or not. Assessment is only a small part of evaluation.[1]

The simple logic of thinking is that if a program only runs continuously without any evaluation, the program will become obsolete and is no longer in

---

[1] Nurul Hidayati Murtafiaf, *Evaluasi Pendidikan*, *Jakarta: Rineka Cipta*, 2018.

accordance with the needs of the community. Tyler explains that *"The process of evaluation is essential the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction"*.[2] What is meant is that the evaluation process is basically a process that determines the extent to which the educational objectives of the curriculum and learning program are achieved. Therefore, curriculum program evaluation is very important to do, so that all elements contained in the program have a major contribution in achieving educational goals and the results of the evaluation are worthy of being used as a foothold in making follow-up policies for a program.

In this research review, we will discuss learning evaluation models and principles, as well as some commonly used program evaluation models in educational contexts. An evaluation model is a framework used to direct the learning evaluation process, while evaluation principles are guidelines or basic rules that must be followed in carrying out learning evaluation.[3] First of all, we will discuss the principles of learning evaluation.[4] These principles include validity, reliability, objectivity and affordability. Validity refers to the extent to which the learning evaluation measures what it is supposed to measure, while reliability relates to the consistency of evaluation results obtained if the evaluation is repeated.[5]

Objectivity means that learning evaluation should be fair, independent and not influenced by subjective factors. Finally, affordability means that learning evaluations should be accessible and understandable to all interested parties, including students, teachers and parents.[6] Examples include the quantitative evaluation model, which includes: the Tyler model, the Taylor and Maguire theoretical model, the Alkin system approach model, the Countenance Stake model, the CIPP model, the microeconomic model. 2)

---

[2] T. A Putra, Andreas, "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)," *Jurnal IAIN*, 2018, 55–68.

[3] Balya Ziaulhaq Achmadin, Abdul Fattah, and Marno Marno, "Metode Dan Strategi Pengajaran Pendidikan Islam Terhadap Generasi Milenial," *Journal of Research and Thought on Islamic Education (JRTIE)* 5, no. 2 (2022): 102–29, https://doi.org/10.24260/jrtie.v5i2.2315.

[4] Arief Aulia Rahman and Cut Eva Nasryah, *Evaluasi Pembelajaran*, *Uwais Inspirasi Indonesia*, 2019.

[5] Putra, Andreas, "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)."

[6] Mardiah Mardiah and Syarifudin Syarifudin, "Model-Model Evaluasi Pendidikan," *MITRA ASH-SHIBYAN: Jurnal Pendidikan Dan Konseling* 2, no. 1 (2019): 38–50, https://doi.org/10.46963/mash.v2i1.24.

Qualitative evaluation models, which include: case study model, illuminative model, and responsive model There is also a grouped evaluation model that divides the evaluation model into four main models, namely *"measurement, congruence, educational system, dan illumination".[7]*

Furthermore, it reviews learning evaluation models according to the study of the discussion, where this model provides guidance on the steps that must be followed in conducting learning evaluations, such as data collection, data analysis, and preparation of evaluation reports.[8] Effective learning evaluation will provide valuable information for decision makers in their efforts to improve the quality of education and achieve the desired learning objectives. In conducting learning evaluation, certain evaluation models can be used to guide the evaluation process in a more structured manner.[9]

The author will review in more detail about each learning evaluation model, including the steps involved in each model. By using the right evaluation model, we can gather relevant and accurate information to measure the effectiveness of learning and improve the quality of education in a sustainable manner. Through understanding the learning evaluation models and principles, we can optimize the evaluation process to achieve the desired learning objectives. Systematic and comprehensive learning evaluation will provide valuable insights for educators and decision makers in improving learning quality and achieving optimal results.

The provisional hypotheses of this study focus on the effectiveness of various evaluation models of Islamic Religious Education (PAI) learning programs in Madrasah 'Aliyah in East Java. First, it is expected that the application of a stakeholder-based evaluation model will yield more comprehensive and relevant information on the effectiveness of PAI programs compared to an objective-based model. Furthermore, it is anticipated that a responsive evaluation model can increase stakeholder engagement, which in turn will have a positive impact on the quality of learning and learner outcomes. In addition, it is expected that madrasahs that apply the theory-based model will show significant differences in PAI program evaluation results when compared to madrasahs that use the traditional evaluation model. Finally,

---

[7] Junaidi Akhmad, "Evaluasi Program Measurement Model," *Pendidikan, Dan Ilmu Sosial* 2, no. 2 (2023): 81–90.

[8] D.dan Wahyudhiana, "Model Evaluasi Program Pendidikan," *Islamadina* 1, no. 1 (1993): 1–28.

[9] Afif Faizin and Hesti Kusumaningrum, "Review Model-Model Evaluasi Program Untuk Pendidikan Dan Pelatihan Online," *EduManajerial* 1, no. 1 (2023): 42–54, https://doi.org/10.15408/em.v1i1.32245.

it is expected that the stakeholder-based evaluation model will increase learner and parent satisfaction levels, creating a better learning environment. These hypotheses will be tested in a study at the Madrasah 'Aliyah level in East Java to gain a deeper understanding of the effectiveness of evaluation models in the context of PAI learning.

RESEARCH METHODS

This research uses a case study approach and literature to examine Islamic Religious Education learning programs in Madrasah 'Aliyah in East Java, researchers took samples of Madrasah 'Aliyah in Blitar and Malang districts. The case study method was chosen because this approach allows researchers to conduct in-depth investigations of contextual phenomena in real life.

In the data collection process, researchers used three main techniques, namely interviews, participatory observation, and document review. Interviews were conducted with various stakeholders, such as the head of madrasah, PAI teachers, and representatives of parents of students, to obtain information about the practice of evaluating PAI learning programs that have been implemented. Observations were conducted in the madrasah environment to directly observe the implementation of PAI learning activities and activities related to program evaluation. Meanwhile, document review was conducted on various administrative documents, curricula, and PAI learning tools applicable in the madrasah.[10]

The data collected through the three techniques were then analyzed qualitatively. The data analysis process includes data reduction, data presentation, and conclusion drawing. The data reduction stage was carried out by selecting, focusing, simplifying, and abstracting important information obtained. Then, the data that has been reduced is presented in the form of narratives, matrices, or diagrams to facilitate understanding and drawing conclusions.[11] Data analysis was conducted continuously to produce findings that could answer the research questions. Through a case study approach and comprehensive data collection techniques, this research seeks to gain an in-depth understanding of the PAI learning program evaluation models that can

[10] Sugiyono Sugiyono, *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D, Dan Penelitian Pendidikan)* (Bandung: Alfabeta, 2021).
[11] Balya Ziaulhaq Achmadin et al., "Synchronization the Merdeka Curriculum at Madrasah ' Aliyah Learning Akidah Akhlak : Facing The Challenges of Moral Degradation in the Era of Society 5 . 0," *JRTIE (Journal of Research and Thought on Islamic Education* 7, no. 1 (2024): 61–90.

be applied in Madrasah 'Aliyah, as well as to identify the model that best suits the context and needs of madrasahs in East Java.

RESULT AND DISCUSSION

Based on the research that has been carried out, it is found that the concept models of program evaluation are commonly applied in learning Islamic religious education in the context of Madrasah 'Aliyah in East Java. The research results can be seen in the following table:

TablE 1. Findings of program evaluation models and concepts

No.	Evaluastion Model	Function
1.	Concept evaluation model CIPP	Context, Input, Process, Product by identifying needs, problems, and opportunities within the program environment.
2.	Concept evaluation model Provus	Known as the Discrepancy Model, it focuses on comparing established standards or criteria with the actual performance of the program.
3.	Concept evaluation model Stake	Also known as the Responsive Model, it focuses on comprehensive program description and assessment. The evaluation begins by describing what actually happens in the program (antecedents, transactions, and outcomes).
4.	Concept evaluation model Kirkpatrick	One of the most widely used program evaluation models, especially in the context of training and development. It focuses on four levels of evaluation: reaction, learning, behavior and outcomes.
5.	Concept evaluation model Brinkerhoff	Also known as The Success Case Method, it focuses on identifying and studying the success stories of a program. The evaluation begins by defining the program's success criteria, then investigates cases where the program successfully achieved its goals.
6.	Concept evaluation model Measurement	It is an evaluation model that emphasizes quantitative measurement of program outcomes or impacts. The evaluation begins by defining the program objectives and developing measurable indicators to assess the level of achievement of these objectives. Next, the evaluator will collect data and measure the program's performance based on the predetermined indicators.
7.	Concept evaluation model Congruence	Known as the alignment model, it focuses on assessing the extent to which a program's goals, design, and implementation are congruent.
8.	Concept evaluation model Illuminative	Known as the illumination model, it is an evaluation approach that focuses on understanding and explaining the program implementation process in depth.
9.	Concept evaluation model Logik	Known as the theory of change model, it is an evaluation approach that focuses on describing the logical relationship between program components, from inputs, activities, outputs, to expected outcomes.

The detailed discussion regarding research findings, summary of data exposure, evaluation model concepts and hypotheses or research answers can be seen in the following discussion:

A. Concept Evaluation Model CIPP

Model CIPP *(Context, Input, Process, Product)* Stufflebeam is one of the most widely applied evaluation models in the context of learning evaluation. It was developed by Daniel Stufflebeam in the 1960s and emphasizes a comprehensive evaluation of all aspects of learning. As a result

of his efforts to evaluate the ESEA *(the Elementary and Secondary Education Act).*[12] According to Madaus, Scriven, Stufflebeam the important purpose of this evaluation model is to improve, according to the statement *"the CIPP approach is based on the view that the most important purpose of evaluation is not to prove but to improve".* Stufflebeam's evaluation model consists of four dimensions, namely: context, input, process, and product, so the evaluation model is named CIPP. The four words mentioned in the abbreviation CIPP are the targets of evaluation, namely the components and processes of an activity program.[13]

The main principle of the CIPP model is to provide useful information for decision-makers at every stage of the learning process. The model consists of four evaluation components, namely:

1. **Evaluasi Konteks** *(Context Evaluation)*

   In this context, evaluation focuses on identifying the needs, goals and environmental conditions that influence learning. Context evaluation helps determine whether the learning objectives set are in line with the real needs and conditions in the field. Referring to Sax's opinion, he explains *"Context evaluation is the delineation and speciafication and sample of individuals to be served and the project objectives. Context evaluation provides a rationale for justifying a particular type of program intervention".*[14] This shows that context evaluation is the activity of gathering information to determine objectives, defining the relevant environment.

   In addition, referring to the opinion, Stufflebeam & Shinkfield further explained that context evaluation: *"To assess the object overall status, to identify its deficiencies, to identify strengths at hand that could be used to remedy the deficiencies, to diagnose problems whose solution would improve the object's. A context evaluation also is aimed at examining whether existing goals and priorities are attuned to the needs of whoever is supposed to be served".* Based on this, it can be understood that

---

[12] Siti Robingah Pujianti, *Evaluasi CIPPO Program Pendampingan Kurikulum 2013 Bagi Guru Kelas X SMA Di Kota Semarang*, 2017, http://lib.unnes.ac.id/29547/1/1102413003.pdf.

[13] Siti Wahyuni Siregar, "Model Dan Rancangan Evaluasi Program Bimbingan Dan Konseling," *Hikmah* 11, no. 2 (2017): 271–90, http://jurnal.iain-padangsidimpuan.ac.id/index.php/Hik/article/view/747.

[14] A B D AMRI SIREGAR, "Evaluasi Model Cipp," *Evaluasi Program Dan Kelembagaan Pendidikan Islam* 9, no. 2 (2021): 163, http://repository.iainbengkulu.ac.id/5904/1/EVALUASI PROGRAM DAN KELEMBAGAAN PENDIDIKAN ISLAM.pdf#page=170.

context evaluation seeks to evaluate the status of the object as a whole, identify shortcomings, strengths, diagnose problems, and provide solutions, test whether goals and priorities are adjusted to the needs to be implemented.[15]

2. Evaluasi Masukan *(Input Evaluation)*

In this context, evaluation looks at the resources used in learning, such as curriculum, teaching methods, learning materials and teacher qualifications. Input evaluation helps to ensure that the inputs used are adequate to achieve the learning objectives. According to Stufflebeam & Shinkfield, the main orientation of input evaluation is to determine how program objectives are achieved.[16] Input evaluation can help organize decisions, determine what resources are available, what alternatives are taken, what are the plans and strategies to achieve goals, what are the work procedures to achieve them. The components of input evaluation include human resources, infrastructure, budget and various procedures and policies.[17]

3. Evaluasi Proses *(Process Evaluation)*

In this context, evaluation examines the implementation of learning, including the interaction between teachers and students, the effectiveness of the use of learning methods and media, and the suitability between the learning plan and its implementation in the classroom. the essence of process evaluation is: checking the implementation of a plan/program. The purpose is to provide feedback to managers and staff on how well program activities are running according to schedule, and using available resources efficiently, providing guidance to modify the plan to suit the needs, evaluating periodically how much those involved in program activities can accept and carry out their roles or duties.[18]

Process evaluation is used to detect or predict procedure design or implementation design during the implementation stage, provide information for program decisions, and as a record or archive of procedures

[15] K Wirahyuni, "Validity and Reliability of a Digital-Based Affective Evaluation Instrument on Indonesian MPK Learning at Ganesha University of Education," *International Journal of Early Childhood Special Education* 13, no. 2 (2021): 1138–45, https://doi.org/10.9756/INT-JECSE/V13I2.211159.

[16] I M S Sandhiyasa, "The Evaluation of the Academic Progress Information System SIsKA-NG Mobile Based on Heuristic and User Experience," *International Journal of Modern Education and Computer Science* 14, no. 2 (2022): 55–64, https://doi.org/10.5815/ijmecs.2022.02.05.

[17] Wahyudhiana, "Model Evaluasi Program Pendidikan."

[18] Siregar, "Model Dan Rancangan Evaluasi Program Bimbingan Dan Konseling."

that have occurred. Process evaluation involves the collection of assessment data that has been determined and applied in program implementation practices.[19]

4. Evaluasi Hasil *(Product Evaluation)*

In this context, evaluation assesses the end results of learning, such as the level of achievement of learning objectives, student understanding of the material, and skills acquired. Outcome evaluation helps determine whether learning objectives have been achieved. to measure, interpret, and determine the achievement of the results of a program, ensuring how much the program has met the needs of a program group served. Meanwhile, according to Sax, the function of outcome evaluation *"… to make decision regarding continuation, termination, of modification of program"* So, the function of outcome evaluation is to help make decisions regarding the continuation, end and modification of the program, what results have been achieved, and what is done after the program runs.

Based on some of the above opinions, it can be seen that product evaluation is an assessment carried out to measure success in achieving predetermined goals. The resulting data will determine whether the program is continued, modified or discontinued. The CIPP model is currently refined with one component O, which stands for outcome, thus becoming the CIPPO model. While the CIPP model stops at measuring outputs, CIPPO extends to the implementation of outputs.[20]

Through the CIPP model, evaluation is carried out thoroughly and systematically, starting from assessing the learning context, paying attention to the inputs used, monitoring the learning process, to measuring the final results achieved. The information obtained from each evaluation component can be used to make decisions in planning, implementing and improving learning programs. The application of the CIPP model in learning evaluation allows educators to obtain a comprehensive picture of the effectiveness of learning programs. The information generated can be utilized to develop a plan for continuous improvement and enhancement of learning quality.[21]

B. Evaluation Model Provus

The Provus model is also known as the Discrepancy Evaluation Model. This model was developed by Malcolm Provus in 1971 and focuses on

---

[19] Mardiah and Syarifuddin, "MODEL-MODEL EVALUASI PENDIDIKAN" 02, no. 01 (2007): 38–50.

[20] Siti Robingah Pujianti, *Evaluasi CIPPO Program Pendampingan Kurikulum 2013 Bagi Guru Kelas X SMA Di Kota Semarang*.

[21] SIREGAR, "Evaluasi Model Cipp."

identifying the discrepancy between the standards or criteria set and the performance or results achieved in learning. This model according to Madaus, Sriven & Stufflebeam departs from the assumption that to determine the feasibility of a program, the evaluator can compare what should be expected to happen (standard) with what actually happens *(performance).*[22] By comparing these two things, it can be seen whether there is a discrepancy, namely the standards set and the actual performance. This model, developed by Malcolm Provus, aims to analyze whether a program should be continued, improved, or discontinued.[23]

This model emphasizes the formulation of standards, performance, and discrepancy in a detailed and measurable manner. The program evaluation carried out by the evaluator measures the size of the gap in each program component. With the description of the gaps in each program component, the improvement steps can be done clearly. The main principle of the Provus model is that evaluation should be conducted continuously throughout the learning process. This model consists of five stages of evaluation, namely:

1. Penetapan Standar *(Standard Setting):* At this stage, the evaluator sets standards or criteria that will be used as benchmarks in assessing performance or learning outcomes.
2. Penentuan Kinerja *(Performance)* Current: Evaluators collect data and information about ongoing performance or learning outcomes.
3. Perbandingan Kinerja with Standards (Comparison): At this stage, the evaluator compares the performance or learning outcomes obtained with the standards that have been previously set.
4. Identifikasi Kesenjangan *(Discrepancy Identification):* Evaluators identify and determine the magnitude of the gap between performance or learning outcomes and established standards.
5. Perbaikan *(Installation):* After identifying the gaps, the evaluator looks for solutions and develops an improvement plan to reduce or eliminate the gaps.[24]

---

[22] Putra, Andreas, "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)."

[23] S B Malcolm, "Chester Barnard's Moral Persuasion, Authenticity, and Trust: Foundations for Leadership," *Journal of Management History* 16, no. 4 (2010): 454–67, https://doi.org/10.1108/17511341011073942.

[24] Alimni Alimni, Alfauzan Amin, and Dwi Agus Kurniawan, "The Role of Islamic Education Teachers in Fostering Students' Emotional Intelligence," *International Journal of Evaluation and Research in Education (IJERE)* 11, no. 4 (December 1, 2022): 1881, https://doi.org/10.11591/ijere.v11i4.22116.

Through the Provus model, evaluation is not only conducted at the end of learning, but also continuously throughout the process. This allows educators to immediately detect gaps and make improvements in the middle of the learning process, so that learning objectives can be achieved optimally. The application of the Provus model in learning evaluation can provide detailed and accurate information about performance and learning outcomes. By comparing actual performance with set standards, educators can identify areas for improvement and take appropriate corrective action to improve learning quality.

C. Evaluation Model Stake

This model was developed by Robert E. Stake of the University of Illinois who emphasized two basic activities in evaluation, namely description and judgment, and distinguished three stages, namely: antecedent (context), transaction/process, and outcomes. Description concerns two things that show the position of something that is the target of evaluation, namely: what the program aims to achieve, and what actually happens. The evaluator shows the steps of consideration that refer to the standard.[25]

Stufflebeam & Shinkfield explain the three stages of the Stake model program evaluation, namely: antecedents, transactions, and outcomes. Antecedents refer to basic information related to what conditions/events existed prior to program implementation. According to Stake, the information in this type is, for example, related to previous teaching and learning activities, and related to outcomes, such as: whether students have eaten breakfast before coming to school, whether students have completed their homework, whether students have a good night's sleep. To fully describe and define a program or learning at a point in time. Stake proposes that evaluators should identify and analyze conditions related to the antecendent. Stake model is also known as Countenance Model. This model was developed by Robert Stake in 1967 and focuses on two main matrices, the descriptive matrix and the judgment matrix.[26]

The central tenet of the Stake model is that evaluations should provide a clear picture of performance and learning outcomes, and provide a judgment or assessment of that performance and outcomes. The Stake model consists of three main components, namely:

---

[25] K Okoye, "Impact of Students Evaluation of Teaching: A Text Analysis of the Teachers Qualities by Gender," *International Journal of Educational Technology in Higher Education* 17, no. 1 (2020), https://doi.org/10.1186/s41239-020-00224-z.

[26] Mardiah and Syarifudin, "Model-Model Evaluasi Pendidikan."

1. *Antecedents* (Masukan): This component includes everything that comes before the learning takes place, such as student characteristics, curriculum, resources and teacher background.
2. *Transactions* (Proses): This component includes all activities and interactions that occur during the learning process, including teaching methods, learning strategies, and student behavior.
3. *Outcomes* (Hasil): It includes the expected and unexpected outcomes or impacts of learning, including changes in students' behavior, knowledge and skills.

In Stake's model, evaluation is done by comparing what is expected (intended) with what actually happens (observed) in each component. Evaluators also make judgments based on predetermined criteria or standards.[27] The application of the Stake model in learning evaluation can provide a comprehensive picture of learning performance and outcomes. The information obtained is not only limited to the achievement of objectives, but also includes the learning process and context that influences it. This allows educators to make more informed decisions in planning, implementing and improving learning programs.[28]

D. Evaluation Model Kirkpatrick

In addition to the learning evaluation models discussed earlier, there is also an evaluation model that focuses on the effectiveness of training or human resource development programs, namely the Kirkpatrick model. This model was developed by Donald Kirkpatrick in 1959 and consists of four levels of evaluation. The evaluation model developed by Kirkpatrick has undergone several refinements, the last of which was updated in 1998 and is known as Evaluating Training Programs: the Four Levels or Kirkpatrick's evaluation model. Evaluation of training programs includes four levels of evaluation, namely: (a) *reaction,* (b) *learning,* (c) *behavior,* and (d) *result.*[29]

A key principle of the Kirkpatrick model is that evaluation should be conducted systematically and comprehensively to assess the success of a training or development program. This model considers that the effectiveness

---

[27] Magdalena Ina et al., "Penyusunan Desain Dan Evaluasi Formatif," *Cendekia Pendidikan* 1, no. 1 (2023): 1–13, https://doi.org/10.9644/scp.v1i1.332.

[28] S D Wurdinger, *Using Experiential Learning in the Classroom: Practical Ideas for All Educators* (books.google.com, 2005), https://books.google.com/books?hl=en&lr=&id=nsxgovtz8Z4C&oi=fnd&pg=PP1&dq=experiential+learning&ots=I6SZEytC0a&sig=mSxy3I2cho6de1jyTZB8JMBSnNY.

[29] Wahyudhiana, "Model Evaluasi Program Pendidikan."

of a program cannot only be seen from the final results, but also the process and impact. The four levels of evaluation in the Kirkpatrick model are:

1. Reaksi *(Reaction):* At this level, the evaluation is conducted to determine participants' reactions or responses to the training or development program. This can include participant satisfaction, interest and motivation. Catalanello & Kirkpatrick explain that evaluating the reactions of trainees means measuring their satisfaction. A training program is considered effective if the participants find the training process enjoyable, so they are interested and motivated to learn and practice. Conversely, if participants are not satisfied with the training process, they will not be motivated to attend further training.[30]

   The success of the training activity process is inseparable from the interest, attention and motivation of the participants in following the course of this activity. People will learn better when they react positively to the learning environment. Participant satisfaction can be assessed from several aspects, namely the material provided; the facilities available; the material delivery strategy used by the learning media; the schedule of activities, to the menu and presentation of the consumption provided. Instruments to measure reactions include reaction sheets in the form of questionnaires. According to Kirkpatrick, in determining the instrument, the principle can be used to reveal as much information as possible, but the filling is as efficient as possible. Evaluation at this level focuses on the reactions of participants that occur during the activities, also known as training process evaluation.

2. Pembelajaran *(Learning):* This level evaluates the extent to which participants gain knowledge, skills or attitudinal changes as a result of the training or development program. According to Kirkpatrick & Kirkpatrick, the evaluation of learning outcomes can be seen in changes in attitude, improvement in knowledge, and or improvement in participants' skills after completing the program. Program participants are said to have learned if they have experienced a change in attitude, improvement in knowledge or improvement in skills.[31]

   To measure the effectiveness of the program, these three aspects need to be measured. Without a change in attitude, an increase in

---

[30] N I Fadlilah, "Evaluation of the Application of the Integration Model of D&M and UTAUT in the Online Learning System for Routing and Switching Essentials Subject," *AIP Conference Proceedings*, 2023, https://doi.org/10.1063/5.0128601.

[31] Mardiah and Syarifudin, "Model-Model Evaluasi Pendidikan."

knowledge or an improvement in skills in the trainees, the program can be said to have failed. Some call this assessment the assessment of learning outcomes (outputs). Therefore, the measurement of learning outcomes must determine: (a) what knowledge has been learned; b) what attitude changes have been made; c) what skills have been developed or improved. Measuring learning outcomes takes a long time when compared to measuring reactions. Measuring reactions can be done with a reaction sheet in the form of a questionnaire. According to Kirkpatrick & Kirkpatrick, assessment of learning outcomes can be done with a comparison group. The group that participated in the training and the group that did not participate in the development training are compared within a certain period of time. In addition, assessment of learning outcomes can also be done by comparing pre and post test results, written tests and performance tests.

3. Perilaku *(Behavior):* Evaluation at this level focuses on changes in participants' behavior after attending the training or development program, particularly in workplace application. The assessment focuses on changes in behavior after participants return to the workplace, also known as evaluation of outcomes and training activities. What changes occur in the workplace after participants have participated in the program, both in terms of knowledge, attitudes and skills. According to Kirkpatrick & Kirkpatrick, behavioral 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 attending the program and, (3) surveys/interviews with trainers, supervisors and subordinates of program participants after returning to the workplace.[32]

4. Hasil *(Results):* This level of evaluation measures the end result or impact that the organization gains from the training or development program, such as increased productivity, profitability or overall organizational performance. Evaluation at this stage focuses on the outcomes that occur because participants have attended a program. Some examples of outcomes in a corporate context include: increased production, improved quality, reduced costs, reduced accidents, increased profits. The way to evaluate final outcomes according to Kirkpatrick & Kirkpatrick is by: (1) comparing a control group with a group of program participants, (2) measuring

---

[32] Ahmad Dahlan, "Integrating Local Wisdom Values in Elementary School To Strengthen Students ' Caring Character Amaliyah Ulfah Universitas," 2018, https://www.semanticscholar.org/paper/a0e0b29c845a398e7d90e6c174c1879e625be30c.

performance before and after the training, (3) comparing the costs used with the benefits obtained after the training, and how the improvement is achieved.

Kirkpatrick's program evaluation model can be applied to learning programs in schools, because: (1) the focus is the same, which is directed at the process and results of learning by following a program, (b) learning changes at four levels are equally directed at aspects of knowledge, attitudes, and skills. However, the application of this evaluation model in learning programs needs to be modified to the school setting. First, evaluation of the outcome and impact of learning activities in the classroom is difficult to do, because schools find it difficult to monitor the extent to which students are able to apply the knowledge and skills they acquire in learning activities at school, as well as in the community within a certain time. Because to reach this level requires a long time, energy and a lot of money, moreover continued on the impact evaluation. Second, the focus of the learning program in a school setting can be directed at the competencies that have been determined.[33]

According to Holton, the strength of this model is its simplicity, its ability to help clarify criteria, and create assessment indicators. With the clarity of criteria and indicators that have been set, the achievements of a program will be measured properly. This model can be applied to evaluate learning programs in schools, even at smaller levels, such as classes and specific programs.[34] This model also has some disadvantages when applied in a school setting. Therefore, there should be adjustments and modifications, so that the purpose of evaluating a school program can be achieved by using this model. According to Bates Alliger & Janak, this model oversimplifies training effectiveness, as it does not consider individual or contextual influences in program evaluation.[35] In fact, the characteristics of the organization, the work/school environment and the characteristics of individual trainees are important inputs that influence the effectiveness of training processes and outcomes. Kirkpatrick's model implicitly assumes that examining these factors is not essential for effective program evaluation.

The application of the Kirkpatrick model in the evaluation of training or development programs can provide comprehensive information on program effectiveness. The evaluation is not only limited to participant reactions and

---

[33] Rahman and Nasryah, *Evaluasi Pembelajaran*.

[34] Faizin and Kusumaningrum, "Review Model-Model Evaluasi Program Untuk Pendidikan Dan Pelatihan Online."

[35] Mardiah and Syarifudin, "Model-Model Evaluasi Pendidikan."

learning, but also looks at behavioral changes and the impact on organizational performance. This allows decision-makers to assess the added value provided by the program and make necessary improvements.

E. Evaluation Model Brinkerhoff

This model of program evaluation related to training or development was developed by Robert Brinkerhoff in 1983, the Brinkerhoff model. This model focuses on analyzing the impact of training programs on individual and organizational performance. The main principle of the Brinkerhoff model is that evaluation should be conducted systematically and continuously to determine the effectiveness of training programs, both at the individual and organizational levels. The model considers that evaluation should be conducted not only at the end of the program, but also during the program implementation process. The Brinkerhoff model consists of six evaluation stages, namely:[36]

1. Evaluation Focus: At this stage, the evaluator determines the purpose, scope, and focus of the evaluation to be conducted.
2. Evaluation Design: At this stage, the evaluator designs the evaluation methods and instruments to be used.
3. Data Collection: This stage involves collecting data from various sources, both quantitative and qualitative.
4. Data Analysis: At this stage, the evaluator analyzes the data that has been collected to determine the effectiveness of the program.
5. Impact Assessment: This stage evaluates the impact of the program on individual and organizational performance.
6. Communication and Follow-up: In the last stage, the evaluator communicates the results of the evaluation and provides recommendations for future program improvements.

Based on the Brinkerhoff model literature, there are three evaluation approaches based on the elements in this model, including :

a. *Fixed vs Emergent Evaluation Design :* A good evaluation design is determined and planned systematically before implementation. The design is developed based on the program objectives along with a set of questions to be answered with information to be obtained from specific sources. An analysis plan is created in advance that will allow the user to receive the

---

[36] Putra, Andreas, "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)."

information as specified in the objectives. This design can be adjusted as needs change.[37]

b. *Formative vs Sumative Evaluation :* Formative evaluation is used to obtain information that can help improve the program, carried out during the implementation of the program. The focus of the evaluation revolves around the needs that have been formulated by the evaluator. Summative evaluation is carried out to assess the benefits of a program, from the results of this evaluation it can be determined whether a particular program will be continued or discontinued. Summative evaluation focuses on variables that are considered important for decision makers. Summative evaluation is conducted at the end of program implementation.

c. *Experimental & Quasi-Experimental Designs vs. Unobtrusive Inquiry :* Some evaluations use classical research methodology. In such cases, research subjects are randomized, treatments are administered and impact measurements are made. The purpose of the study is to assess the benefits of a program that was tried out. If students or programs are randomly selected, then generalizations are made to a somewhat wider population. In some cases intervention is not possible or desirable. If the process has been improved, the evaluator should look at documents, such as studying test scores or analyzing research conducted and so on. Data collection strategies mainly use formal instruments such as tests, surveys, questionnaires and use standardized research methods.[38]

The application of the Brinkerhoff model in the evaluation of training or development programs can provide comprehensive information about the effectiveness of the program, both at the individual and organizational levels. The evaluation is conducted on an ongoing basis, so that it can be used for program improvement during the implementation process. This allows decision-makers to optimize the impact of the program and improve overall organizational performance.

F. Evaluation Model Measurement

In addition to the evaluation models previously discussed, there is also an evaluation model that focuses on measuring the impact of training or human resource development programs, namely the Measurement model. This model

---

[37] C A Ramezan, "Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification," *Remote Sensing* 11, no. 2 (2019), https://doi.org/10.3390/rs11020185.

[38] Balya Ziaulhaq Achmadin, "Exploring The Interactive Media : Enhancing Understanding of Quran and Hadith Learning Through Quizizz Media in Madrasah Aliyah Negeri 2 Blitar," 2023, 239–48.

was developed by Jack Phillips in 1997 and is a further development of the Kirkpatrick model. This model can be seen as the oldest in the history of assessment and is more widely recognized in the educational assessment process. Assessment figures who are seen as developers of this model are R. Thorndike and R.I. Ebel.[39] As the name implies, this model emphasizes the role of measurement activities in carrying out the evaluation process. Measurement is seen as a scientific activity and can be applied in various fields of problems including the field of education. Measurement, according to this model, cannot be separated from the notion of quantity or amount. This amount will show the magnitude of objects, people or events so that the measurement results are always expressed in the form of numbers.

Measurement is thus seen as the activity of determining the magnitude of a certain attribute owned by objects, people, and events in the form of certain units of measure. In the field of education, this model has been applied in the assessment process to see and reveal individual differences and group differences in terms of abilities and interests and attitudes. The results of measurements regarding the above behavioral aspects are used for the purposes of student selection, guidance, and educational planning for the students themselves.[40]

The main principle of the Measurement model is that evaluation should be conducted comprehensively and emphasize on measuring the impact of the program on the organization. This model considers that the effectiveness of the program can not only be seen from the reactions, learning, and behavior of participants, but also must pay attention to the financial and non-financial impacts obtained by the organization. The Measurement Model consists of five levels of evaluation, namely:[41]

1. Reaksi *(Reaction):* At this level, evaluations are conducted to determine participants' reactions or responses to the training or development program, such as satisfaction and perceptions.
2. Pembelajaran *(Learning):* This level evaluates the extent to which participants acquire knowledge, skills or attitudinal changes as a result of the program.

---

[39] M R Yusof, "The Measurement Model of Geo-Education among Trainee Teachers in Malaysia," *International Journal of Evaluation and Research in Education* 10, no. 2 (2021): 714–19, https://doi.org/10.11591/ijere.v10i2.20706.
[40] A Belloni, "Program Evaluation and Causal Inference With High-Dimensional Data," *Econometrica* 85, no. 1 (2017): 233–98, https://doi.org/10.3982/ECTA12723.
[41] Akhmad, "Evaluasi Program Measurement Model."

3. Aplikasi *(Application):* Evaluation at this level focuses on the extent to which participants apply the knowledge and skills acquired on the job.

4. Dampak Bisnis *(Business Impact):* This level of evaluation measures the impact of the program on the organization's business performance and outcomes, such as increased productivity, profitability, or efficiency.

5. *Return on Investment* (ROI): At the last level, the evaluation is conducted to measure the added value or benefits gained by the organization in monetary terms compared to the costs incurred for the program.

The application of the Measurement model in the evaluation of training or development programs can provide comprehensive information about the effectiveness of the program, both at the individual and organizational levels. The evaluation is not only limited to participants' reactions, learning and behavior, but also pays attention to the financial and non-financial impacts obtained by the organization. This enables decision-makers to optimize resource investment in training or development programs and improve overall organizational performance.

G. Evaluation Model Congruence

In addition to the previously discussed evaluation models, there is also an evaluation model that focuses on the suitability or alignment between training or development programs and organizational strategies and goals, namely the Congruence model. This model was developed by David Nadler in the 1970s. Other leaders in the development of this model include W. Tyler, John B. Carrol, and Lee J. Cronbach. Tyler describes education as a process in which there are three things:[42] educational goals, learning experiences, and assessment of learning outcomes. Evaluation activities are intended as activities to see the extent to which educational goals have been achieved by students in the form of learning outcomes that they show at the end of educational activities. Given that educational objectives reflect the desired changes in behavior in students, what is important in the evaluation process is to check the extent to which the desired changes in behavior have occurred in students.[43]

By obtaining information about the achievement of educational goals that have been achieved by students, both individually and in groups, decisions

---

[42] I.M.A.O. Gunawan, "User Experience Evaluation of Academic Progress Information Systems Using Retrospective Think Aloud and User Experience Questionnaire," *Journal of Physics: Conference Series*, 2021, https://doi.org/10.1088/1742-6596/1810/1/012015.

[43] Putra, Andreas, "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)."

can be made about what actions need to be taken. Follow-up evaluation results that concern the interests of these students, for example: providing services or guidance to improve the results that have been achieved, providing material enrichment, and planning other programs for each student. In terms of the education system, the evaluation results are intended as feedback for the need to improve the weak parts of the system.

The main principle of the Congruence model is that evaluation should be conducted to determine the extent to which the training or development program is aligned with the organization's strategy and goals. This model considers that the effectiveness of the program can not only be seen from the impact on individuals, but also must pay attention to the suitability of the program to the needs and conditions of the organization. The Congruence model consists of three main components, namely:[44]

1. Input: In this component, an evaluation is conducted to determine the initial conditions of the organization, such as strategy, structure, culture, resources, and environmental characteristics.
2. Proses: Evaluation in this component focuses on the implementation of the training or development program, including the methods, materials, and learning processes used.
3. Output: In the last component, evaluations are conducted to measure the impact of the program on organizational performance, such as productivity, efficiency, or innovation.

The application of the Congruence model in the evaluation of training or development programs can provide comprehensive information about the extent to which the program is aligned with organizational strategies and goals. The evaluation is not only limited to the impact of the program on individuals, but also pays attention to the suitability of the program to the conditions and needs of the organization.[45] This enables decision makers to optimize resource investment in training or development programs and ensure that they make a significant contribution to the achievement of organizational goals.

H. Evaluation Model Illuminative

This evaluation model focuses on an in-depth understanding of the context and process of implementing a training or development program, the Illuminative model. This model was developed by Parlett and Hamilton in

---

[44] A Goncalves, "Generation and Evaluation of Synthetic Patient Data," *BMC Medical Research Methodology* 20, no. 1 (2020), https://doi.org/10.1186/s12874-020-00977-1.

[45] Murtafiaf, *Evaluasi Pendidikan*.

1972. The main principle of the Illuminative model is that evaluation should be holistic and oriented towards contextual understanding, not merely outcome measurement. This model considers that program effectiveness cannot only be seen from the final results, but must also pay attention to factors that affect program implementation in the field.[46]

This illuminative model emphasizes qualitative assessment. The purpose of this model of evaluation is to conduct a careful study of the system or program in question, which includes: (1) how the program is implemented in the field, (2) how the implementation is affected by the school situation where the program is developed, (3) what are the strengths and weaknesses and how the program affects students' learning experiences. Reported evaluation results are descriptive and interpretive, rather than measurement and prediction. Therefore, the fourth model of evaluation emphasizes the use of judgment.[47]

The difference between research and evaluation research is the existence of criteria in evaluation research. Based on the criteria, researchers in evaluation research give value to the object they are researching. Assessing the effectiveness criteria of a program evaluation model cannot be separated from the purpose/function of program evaluation. Program evaluation has the function of providing information that is used to help make decisions/policy making and further program preparation. In order for the resulting decision to be a good decision, it requires information that is complete, accurate, and reliable (valid and reliable) and timely.[48] Complete information means that the information generated from the evaluation covers the complete program components. Accurate information means that the information generated from the evaluation is information that accurately describes the actual situation of the evaluation object and can be trusted. To obtain accurate information, valid and reliable data collection instruments are needed.[49]

Timely information means that the information obtained from the evaluation results can be delivered to those who need it, to make decisions, develop policies and develop further programs. This timeliness requirement is

---

[46] Mardiah and Syarifudin, "Model-Model Evaluasi Pendidikan."

[47] Tanwir Syah Putra, "Model Evaluasi Pembelajaran Pendidikan Diniyah Di SMPN 2 Banda Aceh," 2018, 94, https://repository.ar-raniry.ac.id/id/eprint/6665/2/Tanwir Syah Putra.pdf.

[48] Indah Aminatuz Zuhriyah, "Landasan Pengembangan Substansi Manajemen Pendidikan Dasar Dalam Perspektif Islam," *MADRASAH* 1, no. 1 (April 21, 2012), https://doi.org/10.18860/jt.v1i1.1850.

[49] Minsih Minsih and Aninda Galih D, "Peran Guru Dalam Pengelolaan Kelas," *Profesi Pendidikan Dasar* 1, no. 1 (2018): 20, https://doi.org/10.23917/ppd.v1i1.6144.

related to the practicality of collecting, processing, and presenting/reporting information. This requires evaluation guidelines that are simple and easily understood by the users of the evaluation model, so that the evaluation process can take place more quickly without neglecting the completeness and accuracy of the information. So a good program evaluation model is: (1) comprehensive, involving all program components/subcomponents, including inputs, processes, outputs, and outcomes, (2) practical, i.e. easy to use and manage, (3) economical, i.e. requiring relatively little cost, as well as time and energy, (4) valid and reliable data collection instruments. The Illuminative model consists of three main components, namely:

1. Observasi Pengamatan *(Observation):* In this component, the evaluator makes in-depth observations of the program implementation process, the interaction between participants and instructors, and the dynamics that occur during the program.

2. Investasi Instruksional *(Instructional Transaction):* Evaluation in this component focuses on understanding the learning context, such as the role and function of the instructor, the learning methods used, and participants' responses.

3. Analisis Deskriptif *(Descriptive-Interpretative):* In the last component, the evaluator conducts descriptive and interpretative analysis to understand the meaning and implications of the findings obtained during the observation and instructional investigation.[50]

The application of the Illuminative model in the evaluation of training or development programs can provide in-depth information about the context and process of program implementation. The evaluation is not only limited to measuring the final results, but also takes into account the factors that influence the implementation of the program in the field. This allows decision-makers to comprehensively understand how the program works, identify strengths and weaknesses, and formulate recommendations for future program improvement and development.

I. Evaluation Model Logik

This evaluation model focuses on understanding the logic and theory of change underlying the training or development program, the Logic model. This model was developed by Carol Weiss in the 1990s. A logical model is a logical and precise depiction of a program according to certain conditions in order to solve problems. In general, the form of depiction uses a flowchart that

---

[50] Sri Warsono, "Pengelolaan Kelas Dalam Meningkatkan Belajar Siswa," *Manajer Pendidikan* 10, no. 5 (2016): 469–76.

explains the planned activities and expected outcomes of this evaluation model, in line with the above opinion, the W.K. Kellogg Foundation explains that: *"basically, a logic model is a systematic and visual way to present and share your under standing of the relationships among the resources you have to operate program, the activities you plan, and the changes or results you hope to achieve"*.[51] Unique to logical models is the use of tables and flow charts that contain inputs, activities, and outcomes. Most use text and arrows or graphics to describe the sequence of activities to produce change, and how those activities connect to the expected program outcomes achieved.

It is important to make an informed decision before using a logical model, as logical modeling is complex and involves one of three approaches, namely: conceptual, outcomes, and activities (applied) approaches or a mixture of these. In simple terms, a logical model can be described as follows:
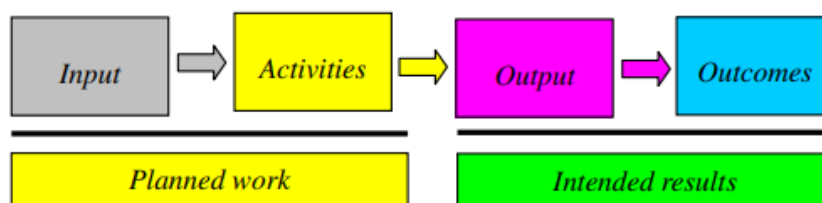
Logic models have been widely used in various fields since the 1980s and early 1990s. For example, logical models have been used to describe programs in education, health, international development, social work, social services, and other fields. The essential elements of a logical model according to United Way of America (1996) consist of three parts:[52] inputs, outputs (activities and participants or methodology), and outcomes. Inputs relate to the important resources that will be invested in the program (what we invest), outputs relate to what activities are carried out (what we do, and who we reach) and outcomes relate to the desired effect or change with the program being run. In order to make the logical model more focused, it is necessary to make its scope, for example, regarding the period of time the program is implemented,

---

[51] L Pereira, "Performance Evaluation in Non-Intrusive Load Monitoring: Datasets, Metrics, and Tools—A Review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018, https://doi.org/10.1002/widm.1265.

[52] Yuyun Affandi, Agus Darmuki, and Ahmad Hariyadi, "The Evaluation of JIDI (Jigsaw Discovery) Learning Model in the Course of Qur'an Tafsir," *International Journal of Instruction* 15, no. 1 (January 1, 2022): 799–820, https://doi.org/10.29333/iji.2022.15146a.

the period of the desired outputs and outcomes, and the form of change expected.[53]

A key principle of the Logic model is that evaluations should be conducted with an in-depth understanding of how the program is expected to achieve its intended goals. This model considers that the effectiveness of a program should not only be judged by the end results, but also by the logic of the intervention and the theory of change underlying the program. The Logic Model consists of several main components, namely:[54]

1. Input: In this component, the evaluation is conducted to identify the resources needed to run the program, such as budget, facilities, and manpower.
2. Activities: Evaluation in this component focuses on the activities or actions carried out in the program, such as training, guidance, or development.
3. Output: In this component, the evaluation is conducted to measure the direct results produced by the program activities, such as the number of participants who attended the training or the number of modules developed.
4. Outcome: Evaluation in this component focuses on the impact or changes produced by the program, either in the short, medium, or long term.
5. Impact: In the last component, the evaluation is conducted to measure the broader effects of the program on the wider community or environment.

Applying the Logic model to the evaluation of a training or development program can provide a deeper understanding of the logic of the intervention and the theory of change underlying the program. Evaluation is not limited to measuring outcomes, but also considers the series of processes that link inputs, activities, outputs, outcomes and impacts. This allows decision-makers to identify critical assumptions, test the validity of the theory of change, and formulate recommendations for future program improvement and development.

CONCLUSION

The research shows that there are various evaluation models of Islamic Religious Education (PAI) learning programs that can be applied in Madrasah 'Aliyah, including: goal-based evaluation, contextual-based evaluation, and responsive evaluation. Each evaluation model has advantages and limitations in the context of evaluating Islamic Education learning programs in Madrasah

---

[53] Murtafiaf, *Evaluasi Pendidikan*.
[54] Suyanto and Asep Jihad, *MENJADI GURU PROFESIONAL: Strategi Meningkatkan Kualifikasi Dan Kualitas Guru Di Era Global*, *Penerbit Erlangga*, 2013.

'Aliyah. Evaluation concepts and models ranging from CIPP, Provus, Stake, Kirkpatrick, Brinkerhoff, Measurement, Congruence, Illuminative, Logic, have their respective advantages where the selection of the model should consider factors such as scope, availability of resources, and stakeholder involvement. The contextual-based evaluation model is considered the most appropriate and effective for evaluating PAI learning programs in Madrasah 'Aliyah in East Java. This approach can produce a comprehensive evaluation, involve various perspectives, and improve the quality of PAI learning programs.

The application of contextual-based evaluation model can be done through the stages of: (a) identification of key stakeholders, (b) data collection from various sources, (c) holistic data analysis, and (d) formulation of recommendations for improving the PAI learning program in accordance with the needs and context of Madrasah 'Aliyah. The implementation of the context-based evaluation model requires commitment and collaboration from all stakeholders, including the madrasah head, PAI teachers, parents, and other related parties. This is necessary to ensure the quality and sustainability of PAI learning program evaluation in Madrasah 'Aliyah. Overall, this study confirms the importance of adopting a comprehensive and contextual program evaluation model to improve the quality of PAI learning in Madrasah 'Aliyah. This approach can produce an evaluation that is objective, participatory, and oriented towards continuous program improvement. The weaknesses in this research that can be continued for other researchers are related to the focus of using the evaluation model on one field of study of Islamic religious education subjects, so that the hope can be more comprehensive in the study that is the focus of research.

BIBLIOGRAPHY

Achmadin, Balya Ziaulhaq. "Exploring The Interactive Media : Enhancing Understanding of Quran and Hadith Learning Through Quizizz Media in Madrasah Aliyah Negeri 2 Blitar," 2023, 239–48.

Achmadin, Balya Ziaulhaq, Putri Bayu, Naflah Rifqi, and Benny Afwadzi. "Synchronization the Merdeka Curriculum at Madrasah ' Aliyah Learning Akidah Akhlak : Facing The Challenges of Moral Degradation in the Era of Society 5 . 0." *JRTIE (Journal of Research and Thought on Islamic Education* 7, no. 1 (2024): 61–90.

Achmadin, Balya Ziaulhaq, Abdul Fattah, and Marno Marno. "Metode Dan Strategi Pengajaran Pendidikan Islam Terhadap Generasi Milenial." *Journal of Research and Thought on Islamic Education (JRTIE)* 5, no. 2 (2022): 102–29. https://doi.org/10.24260/jrtie.v5i2.2315.

Affandi, Yuyun, Agus Darmuki, and Ahmad Hariyadi. "The Evaluation of JIDI (Jigsaw Discovery) Learning Model in the Course of Qur'an Tafsir." *International Journal of Instruction* 15, no. 1 (January 1, 2022): 799–820. https://doi.org/10.29333/iji.2022.15146a.

Akhmad, Junaidi. "Evaluasi Program Measurement Model." *Pendidikan, Dan Ilmu Sosial* 2, no. 2 (2023): 81–90.

Alimni, Alimni, Alfauzan Amin, and Dwi Agus Kurniawan. "The Role of Islamic Education Teachers in Fostering Students' Emotional Intelligence." *International Journal of Evaluation and Research in Education (IJERE)* 11, no. 4 (December 1, 2022): 1881. https://doi.org/10.11591/ijere.v11i4.22116.

Belloni, A. "Program Evaluation and Causal Inference With High-Dimensional Data." *Econometrica* 85, no. 1 (2017): 233–98. https://doi.org/10.3982/ECTA12723.

Dahlan, Ahmad. "Integrating Local Wisdom Values in Elementary School To Strengthen Students ' Caring Character Amaliyah Ulfah Universitas," 2018. https://www.semanticscholar.org/paper/a0e0b29c845a398e7d90e6c174c1879e625be30c.

Fadlilah, N I. "Evaluation of the Application of the Integration Model of D&M and UTAUT in the Online Learning System for Routing and Switching Essentials Subject." *AIP Conference Proceedings*, 2023. https://doi.org/10.1063/5.0128601.

Faizin, Afif, and Hesti Kusumaningrum. "Review Model-Model Evaluasi Program Untuk Pendidikan Dan Pelatihan Online." *EduManajerial* 1, no. 1 (2023): 42–54. https://doi.org/10.15408/em.v1i1.32245.

Goncalves, A. "Generation and Evaluation of Synthetic Patient Data." *BMC Medical Research Methodology* 20, no. 1 (2020).

https://doi.org/10.1186/s12874-020-00977-1.

Gunawan, I.M.A.O. "User Experience Evaluation of Academic Progress Information Systems Using Retrospective Think Aloud and User Experience Questionnaire." *Journal of Physics: Conference Series*, 2021. https://doi.org/10.1088/1742-6596/1810/1/012015.

Ina, Magdalena, Septiyani Alina, Bahri Tauriska Aprilis Bunga, and Komsiah Ani Fitri. "Penyusunan Desain Dan Evaluasi Formatif." *Cendekia Pendidikan* 1, no. 1 (2023): 1–13. https://doi.org/10.9644/scp.v1i1.332.

Malcolm, S B. "Chester Barnard's Moral Persuasion, Authenticity, and Trust: Foundations for Leadership." *Journal of Management History* 16, no. 4 (2010): 454–67. https://doi.org/10.1108/17511341011073942.

Mardiah, Mardiah, and Syarifudin Syarifudin. "Model-Model Evaluasi Pendidikan." *MITRA ASH-SHIBYAN: Jurnal Pendidikan Dan Konseling* 2, no. 1 (2019): 38–50. https://doi.org/10.46963/mash.v2i1.24.

Mardiah, and Syarifuddin. "MODEL-MODEL EVALUASI PENDIDIKAN" 02, no. 01 (2007): 38–50.

Minsih, Minsih, and Aninda Galih D. "Peran Guru Dalam Pengelolaan Kelas." *Profesi Pendidikan Dasar* 1, no. 1 (2018): 20. https://doi.org/10.23917/ppd.v1i1.6144.

Murtafiaf, Nurul Hidayati. *Evaluasi Pendidikan*. *Jakarta: Rineka Cipta*, 2018.

Okoye, K. "Impact of Students Evaluation of Teaching: A Text Analysis of the Teachers Qualities by Gender." *International Journal of Educational Technology in Higher Education* 17, no. 1 (2020). https://doi.org/10.1186/s41239-020-00224-z.

Pereira, L. "Performance Evaluation in Non-Intrusive Load Monitoring: Datasets, Metrics, and Tools—A Review." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018. https://doi.org/10.1002/widm.1265.

Putra, Andreas, T. A. "Evaluasi Program Pendidikan: Pendekatan Evaluasi Program Berorientasi Tujuan (Goal-Oriented Evaluation Approach)." *Jurnal IAIN*, 2018, 55–68.

Putra, Tanwir Syah. "Model Evaluasi Pembelajaran Pendidikan Diniyah Di SMPN 2 Banda Aceh," 2018, 94. https://repository.ar-raniry.ac.id/id/eprint/6665/2/Tanwir Syah Putra.pdf.

Rahman, Arief Aulia, and Cut Eva Nasryah. *Evaluasi Pembelajaran*. *Uwais Inspirasi Indonesia*, 2019.

Ramezan, C A. "Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification." *Remote Sensing* 11, no. 2 (2019). https://doi.org/10.3390/rs11020185.

Sandhiyasa, I M S. "The Evaluation of the Academic Progress Information System SIsKA-NG Mobile Based on Heuristic and User Experience." *International Journal of Modern Education and Computer Science* 14, no. 2 (2022): 55–64. https://doi.org/10.5815/ijmecs.2022.02.05.

SIREGAR, A B D AMRI. "Evaluasi Model Cipp." *Evaluasi Program Dan*

Kelembagaan Pendidikan Islam 9, no. 2 (2021): 163. http://repository.iainbengkulu.ac.id/5904/1/EVALUASI PROGRAM DAN KELEMBAGAAN PENDIDIKAN ISLAM.pdf#page=170.

Siregar, Siti Wahyuni. "Model Dan Rancangan Evaluasi Program Bimbingan Dan Konseling." *Hikmah* 11, no. 2 (2017): 271–90. http://jurnal.iain-padangsidimpuan.ac.id/index.php/Hik/article/view/747.

Siti Robingah Pujianti. *Evaluasi CIPPO Program Pendampingan Kurikulum 2013 Bagi Guru Kelas X SMA Di Kota Semarang*, 2017. http://lib.unnes.ac.id/29547/1/1102413003.pdf.

Sugiyono, Sugiyono. *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R&D, Dan Penelitian Pendidikan)*. Bandung: Alfabeta, 2021.

Suyanto, and Asep Jihad. *MENJADI GURU PROFESIONAL: Strategi Meningkatkan Kualifikasi Dan Kualitas Guru Di Era Global*. Penerbit Erlangga, 2013.

Wahyudhiana, D.dan. "Model Evaluasi Program Pendidikan." *Islamadina* 1, no. 1 (1993): 1–28.

Warsono, Sri. "Pengelolaan Kelas Dalam Meningkatkan Belajar Siswa." *Manajer Pendidikan* 10, no. 5 (2016): 469–76.

Wirahyuni, K. "Validity and Reliability of a Digital-Based Affective Evaluation Instrument on Indonesian MPK Learning at Ganesha University of Education." *International Journal of Early Childhood Special Education* 13, no. 2 (2021): 1138–45. https://doi.org/10.9756/INT-JECSE/V13I2.211159.

Wurdinger, S D. *Using Experiential Learning in the Classroom: Practical Ideas for All Educators*. books.google.com, 2005. https://books.google.com/books?hl=en&lr=&id=nsxgovtz8Z4C&oi=fnd&pg=PP1&dq=experiential+learning&ots=I6SZEytC0a&sig=mSxy3I2cho6de1jyTZB8JMBSnNY.

Yusof, M R. "The Measurement Model of Geo-Education among Trainee Teachers in Malaysia." *International Journal of Evaluation and Research in Education* 10, no. 2 (2021): 714–19. https://doi.org/10.11591/ijere.v10i2.20706.

Zuhriyah, Indah Aminatuz. "Landasan Pengembangan Substansi Manajemen Pendidikan Dasar Dalam Perspektif Islam." *MADRASAH* 1, no. 1 (April 21, 2012). https://doi.org/10.18860/jt.v1i1.1850.