



**TRAINING EVALUATION OF ELEMENTARY SCHOOL TEACHERS
OF 3T REGIONS OF MAHAKAM ULU REGENCY
BY USING KIRKPATRICK**

Sebastianus Widanarto Prijowuntato and Bonifasius Widharyanto

Sanata Dharma University, Indonesia

widanarto@usd.ac.id; bwidharyanto@gmail.com

*correspondence: widanarto@usd.ac.id

<https://doi.org/10.24071/ijiet.v5i2.3540>

received 28 February 2021; accepted 15 March 2021

Abstract

This study aims to determine the success of training for elementary school teachers in Mahakam Ulu Regency, North Kalimantan in 2016 and 2017. The evaluation method used is the evaluation method proposed by Kirkpatrick. In this study, two of the four levels were investigated. The study method used is the Kirkpatrick evaluation model level one and two. At level one, participants conduct evaluations related to satisfaction and level two participants evaluate learning. At level one, participants are asked to fill in satisfaction instruments. The results of the satisfaction instrument entry were analyzed using weighting on each of the satisfaction dimensions. At level two, participants are asked to work on the pre-test and post-test instruments. Data were analyzed using the Wilcoxon Test. The results showed that the level of participant satisfaction for first level was 83.848% for 2015 and 83,178% for 2016. Both scores showed positive reactions to the training. For second level, learning, the average increase in knowledge was 10.984 for 2015 and 9.4 for 2016. The increase in the mean score was significantly based on the Wilcoxon test.

Keywords: evaluation, Kirkpatrick, satisfaction, training

Introduction

Human resource development requires to be conducted continuously. One of the goals of human resource development is to make teachers to have competences and to contributes to the state development. Law Number 5 of 2014 on State Civil Apparatus (ASN) article 3 states that ASN is a profession based on the principle of competence in accordance with the field of duty. One of the ways conducted to develop human resource is training. Training is conducted by many institutions to develop employees' knowledge, skill and competence. In addition, training aims to increase enthusiasm and services oriented to the interest of community, nation, state and homeland.

Education plays a crucial role to develop the nation. Moreover, teachers contribute to prepare students to participate in the development. Teachers teach and guide students by providing knowledge and skills in accordance with the

development of science and technology. Therefore, teachers should get involved in the development of both science and technology.

Trainings conducted aim to sharpen and recall knowledge that has been previously learned. Nurjanah (2018) states that training is a process of transferring knowledge through education and coaching success. It is like soccer players. Even though they can kick, pass, and dribble, practicing with the team is a must. They keep practicing to kick, pas, and dribble. In addition, Aminah (2015) states that the success of training program is determined by the formulation of training process consisting of identification of training requirements, training plans, development, implementation, and evaluation. Those process aims to ensure the training program is in line with organizations or institutions' necessity.

On the other hand, teachers should quickly adapt to the advancement in information technology. Technological development leads to wide interaction and at the same time, disrupts various areas of human life (Pahlevi, 2019). Technological development massively changes the world. Transportation, economic, telecommunication, cultural, as well as educational sectors also experience significant changes. Teachers should prepare students to enter the modern world that is increasingly giving rise to uncertainty.

Teachers in Department of Education of Mahakam Ulu Regency of North Kalimantan conducted education and training (*diklat*) in Sanata Dharma University of Yogyakarta. This training is conducted for elementary school teachers in Mahakam Regency. Training materials cover four competencies namely pedagogic, personality, professional and social competencies (*Undang-Undang Guru dan Dosen*, 2005).

Trainings aim to broaden teachers' knowledge, duties and roles in increasing their ability to master five fields of study in Elementary School, 2013 Curriculum, learning instrument arrangement and increase teachers' spiritual and social attitudes. The success of a training needs to be measured in terms of the level of success by determining success indicators. Besides, training implementation needs to be evaluated so that relevant parties can improve weaknesses in training, decide further training, and consider the benefit for organizations (Kirkpatrick & Kirkpatrick, 2006).

This study aims to evaluate the implementation of education and trainings for elementary school teachers in Mahakam Ulu Regency of North Kalimantan in 2016b and 2017. Evaluation model used is Kirkpatrick's Evaluation Model. This study only examined level 1 and level 2 out of four levels of Kirkpatrick's Evaluation Model. Due to study limitation, the evaluation of level 3 and level 4 cannot be conducted because it can be conducted in teachers' place of origin.

Program evaluation is used to obtain accurate and objective information on program implementation (Ramadhon, 2019). Program evaluation is conducted to identify the achievement of implemented programs. Relevant parties can make decisions to improve program implementation, determine further programs, replicate the program, and determine the impact of the program for institutions and society.

Stufflebeam (1971) define evaluation as a process of describing, obtaining, and providing information that is useful to assess alternative decisions. Therefore, there are three things to consider. First, assessing process is known as systematical and continuous activities. Second, evaluation process covers three steps that are

making questions that must be answered, obtaining relevant information, and providing information in decision making. Third, evaluation is known as a process to serve decision making.

There are many training participants fail to apply knowledge and skills obtained during the training. One of the causes is the absence of assistance to participants after they finished the training. For example, after the training, teachers do not implement their learning outcomes. There are several possible causes. First, teachers still implement their old habits. Second, their school environment does not support the implementation of new knowledge. Third, teachers still have difficulties in implementing knowledge and skills they have obtained. Fourth, facilities and infrastructures are not adequate for the implementation of knowledge and skills. Fifth, participants attend trainings for the sake of formality, to meet the conditions required.

Therefore, the evaluation of education and training becomes a crucial part in the program itself (Topna, 2012). Moreover, Topna (2012) states that training evaluation ensure participants' ability to implement the training in their working environment. The successful participants implementing the training are expected to give good impact for their organization.

Program Evaluation Models

There are many evaluation models widely used to evaluate programs. Some of them are Tyler's Objectives Model, Scriven's Goal-free model, Stake's Responsive Model, Stufflebeam's CIPP Model, Discrepancy Evaluation Model (Anh, 2018), Kirkpatrick model (Kirkpatrick & Kirkpatrick, 2006).

The development of evaluation model began in 1949 (Muryadi, 2017; Anh, 2018). This development was started by Tyler from 1930 to 1945. This model is pioneered by Tyler known as Tyler's Objectives Model. The characteristics of Tyler's model is that the model evaluates the level or degree of instructional goals or objectives being achieved. The model involves careful formulation in accordance with educational goals (students, society), learning materials, learning psychology, and educational philosophy. If the goals are not achieved, instructional programs may fail. Tyler's Objectives Model can only be used to evaluate clearly defined goals.

In 1959, Kirkpatrick proposed program evaluation model known as Kirkpatrick's Model. This model became known in 1994 when Kirkpatrick published a book titled "Evaluating Training Program". The model consists of four levels namely reaction, learning, behavior, and result. This model will be discussed further.

In 1960s, Daniel Stufflebeam proposed an evaluation model known as Context, Input, Process, and Product. This model is created to increase and achieve accountability of school programs in United States (Anh, 2018). CIPP model by Stufflebeam is defined as a comprehensive framework to guide the evaluation of programs, projects, personnel, products and systems (Stufflebeam, 2003). Evaluation process using CIPP used to monitor and assess the activity of a program implementation. This model is based on learning by doing and good moral (objective).

Discrepancy Evaluation Model was proposed by Provus in 1969 (Provus, 1969; Buttram & Covert, 1969). Discrepancy Evaluation Model produces

information for program assessment and improvement. Provus defines evaluation as a comparison between actual performance and desired standard. In the discrepancy model (gap), there are five stages namely program design, installation, process, product, and cost-benefit analysis.

Robert E. Stake proposed a system to evaluate education in 1972 (Anh, 2018). This model is known as Stake's Responsive Model. The evaluation is responsive if it is oriented to program activities instead of program goals. This model emphasizes the stakeholder's main interests obtained from conversations with the stakeholder continually during the evaluation.

Goal-free evaluation model by Michael Scriven was introduced in 1972 (Anh, 2018). This model is driven by educational investment happened at that time. The evaluation happened at that time was influenced by project goals. Therefore, Scriven proposed goal-free model. Goal is defined as statements of wide program goals in which the outcome is expected. The characteristics of this model are outcome-focused, intentional, unanticipated, assessor-free, and unrelated to the rhetoric of instructional makers.

Kirkpatrick's Evaluation Model

There are four levels of training evaluation model proposed by Kirkpatrick, namely reaction, learning, behaviour, and result pelatihan (Kirkpatrick & Kirkpatrick, 2009). The first stage is reaction. This stage measures participants' reaction to the training. This similar to measure participants' satisfaction to the training conducted. The training is considered successful if participants feel interested and motivated to participate in the training.

The interest in and motivation for the training are measured from training materials including modules provided, instructors, training venues, accomodations, food, and services for participants. If participants give positive responses to the service provided, the training is considered successful. In contrast, if participants give negative responses to the service provided, the training is considered unsuccessful. The training at the first level is useful to provide input for training organizers (Kirkpatrick & Kirkpatrick, 2006).

The second stage of Kirkpatrick's evaluation is learning. Learning can be defined as change in knowledge, attitudes, and skills of training participants (Kirkpatrick & Kirkpatrick, 2006). In the second stage, the successful training can be measured by using pre-test and post-test. If there is an increase in the score of the post-test, the education and training can be considered successful.

In the third stage, behavior is defined as change in participants's behavior. In the first and the second stage, participants is possible to be successful, however, if there is no change in behavior, it can be said that the education and training have failed. This behavior change means participants have a desire ti change, know what and how to do, work in the right situation, and give reward over behavior change.

The fourth stage is result. Result can be defined as final outcomes as a result of participants get involved in education and training. The final outcomes can be in the form of students' score, improvement of school discipline, increases in enthusiasm and motivation of students and teachers, and so on. Final coutcomes of a training sometimes cannot be seen instantly, however, the final outcomes can be seen several years after students graduate from the school.

The advantages of the evaluation by Kirkpatrick are easier to be implemented, does not only include final tests, and is more comprehensive due to softskill and hardskill measurement (Kholik, 2020). In addition, the evaluation is simple and can be implemented in various training situations (Nuraini, 2017). Furthermore, Kholik (2020) states that disadvantages of the evaluation by Kirkpatrick are the inputs are not considered in the training, the outcomes are difficult to measure because the evaluation is out of training implementation (Kholik, 2020). Those disadvantages can be anticipated with a commitment between relevant parties to achieve the success of education and training.

Method

This study was quantitative study. The data was obtained during the training of teachers of Mahakam Ulu Regency in Yogyakarta in 2016 and 2017. The data includes training satisfaction and pre-test and post-test scores. The data of satisfaction was obtained by using instruments of satisfaction, while the data of pre-test and post-test scores were obtained by using instruments developed by Intitute for Study and Community Services of Sanata Dharma University. The study method used was evaluation method proposed by Kirkpatrick. This study used two levels of Kirkpatrick's evaluation method, namely reaction level and learning level. The first level is reaction. Training participants' reaction measures participants' satisfaction to the training conducted. The data of participants' satisfaction was obtained by developing satisfaction instruments using likert scale. Dimensions developed in satisfaction instruments were training materials, mood, instructors, facilities include modules, training venues, accomodations, and food. Measurement of the degree of training participants' satisfaction was conducted by using the following aspects:

The second level is learning. The success level of learning was measure before and after the training was conducted. The success level was examined by using Wilcoxon test to find out the level of significance.

Finding and Discussion

Findings

The training of elementary school teachers of Mahakam Ulu Regency of North Kalimantan was held in *Batik Hotel*, at *jalan Dr. Sutomo*, Yogyakarta. The training was conducted from July 27, 2015 to September 27, 2015 and from August 21, 2016 to October 17, 2016.

The first level is reaction. The average results of the satisfaction instrument compared to the satisfaction criteria are as follows:

Table 1. Satisfaction Criteria

Range	Interpretation
< 50%	Participants show bad reactions to the training
50.1% - 60%	Participants show better reactions to the training
60.1% - 80%	Participants show positive reactions because they realize that they receive useful input during the training
80,1% - 100%	Participants show highly positive reactions that is impressive, useful, and highly applicable

The instrument result distributed to training participants is showed in Table 2 below.

Table 2. Training Participants' Satisfaction

No	Indicator	Score Average 2015	Score Average 2016	Description
1.	Training materials	81.67	80.42	Participant's positive reaction to the training is high, memorable, useful, and very applicable
2.	Training participants' mood	89.38	91.25	Participant's positive reaction to the training is high, memorable, useful, and very applicable
3.	Instructor	87.00	84.64	Participant's positive reaction to the training is high, memorable, useful, and very applicable
4.	Training Venue	86.81	82.08	Participant's positive reaction to the training is high, memorable, useful, and very applicable
5.	Training Schedule	74.38	77.5	Participants show positive reaction because they realize that they get useful input during the training
Average		83.848	83.178	Participant's positive reaction to the training is high, memorable, useful, and very applicable

The second stage is learning. In this stage, teachers are given pre-test and post-test. Questions given are related to training materials provided including Natural Science, Social Science, Mathematics, Indonesian, Cultural Arts. The result of pre-test and post-test is shown in Table 3 below.

Table 3. Learning Outcomes

Participants	2015		2016	
	Pre-test	Post-test	Pre-test	Post-test
1.	48	58.33	37.00	50.33
2.	30	50.00	39.00	50.33
3.	37	36.67	33.00	40.33
4.	35	46.67	42.00	47.67
5.	58	65.00	41.00	48.33
6.	65	73.33	44.00	44.33
7.	43	53.33	31.00	39.67
8.	32	56.67	31.00	37.67
9.	32	58.33	39.00	49.00
10.	32	41.67	30.00	33.67
11.	30	43.33	38.00	44.33
12.	32	35.00	37.00	45.67
13.	43	48.33	37.00	49.00
14.	45	61.67	30.00	39.00
15.	32	48.33	45.00	61.67
16.	50	56.67	30.00	43.00
17.	28	51.67	36.00	55.67

18.	35	45.00	32.00	48.33
19.	40	40.00	41.00	43.00
20.	50	46.67	43.00	53.00
Average	39.85	50.8335	36.8	46.2
Std. Dev	10.205	9.740	5.001	6.580

In order to determine types of the test used, normality test and homogeneity test are necessary. The normality test used is Kolmogorov-Smirnov test. Meanwhile, the homogeneity test used is one-way ANOVA. The normality test result is showed in Table 4 while the homogeneity test result is showed in Table 5.

According to the normality test result in Table 4, Asymp.Sig (2-tailed) value is 0.865 for the year of 2015 and 0.998 for the year of 2016. Since to both values are greater than $\alpha = 0.05$, data distribution follows normal distribution. In Table 5, the significance value for the year of 2015 is 0.001 and the year of 2016 is 0.000. Since both values are smaller than 0.05, the data is not homogeneous. Since parametric test requirements are not met, non-parametric test, Wilcoxon Test is used to determine significance of the differences before and after the training for the year of 2015 and 2016.

Table 4. One-Sample Kolmogorov-Smirnov Test

		Year of 2016	Year of 2015
N		40	40
Normal Parameters ^{a,b}	Mean	41.5000	45.3418
	Std. Deviation	7.47907	11.30856
	Absolute	.062	.095
Most Extreme Differences	Positive	.056	.095
	Negative	-.062	-.067
Kolmogorov-Smirnov Z		.393	.599
Asymp. Sig. (2-tailed)		.998	.865

a. Test distribution is Normal.

b. Calculated from data.

Table 5. ANOVA Homogeneity Test

		Sum of Squares	df	Mean Square	F	Sig.
2015	Between Groups	1206.373	1	1206.373	12.124	.001
	Within Groups	3781.089	38	99.502		
	Total	4987.462	39			
2016	Between Groups	883.600	1	883.600	25.870	.000
	Within Groups	1297.925	38	34.156		
	Total	2181.525	39			

The result of pre-test and post-test of each year are examined for the significance level using Wilcoxon Test. The result of Wilcoxon test is showed in Table 6.

Table 6. Statistics^a Test

	Post-test2015 - Pre-test2015	Post-test2016 - Pre- test2016
Z	-3.662 ^b	-3.921 ^b
Asymp. Sig. (2-tailed)	.000	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Based on Table 6 above, Asymp. Sig (2-tailed) value is 0.000 for the year of 2015 and 0.000 for the year of 2016. Since both values are smaller than $\alpha = 0.05$, the hypothesis stating the difference before and after the training is accepted (H_a is accepted).

Discussion

Reaction

In the first level, reaction, training participants' average score is above 80 for the year of 2015 and 2016. The score shows highly positive reaction. In addition, they are impressed with the training they joined, because it is useful for those who have duty and responsibility as teachers. Furthermore, the training is highly applicable for participants. Based on the average score of participants' satisfaction, the education and training of teachers of Mahakam Ulu Regency is considered successful.

If each indicator is observed, the lowest score is training schedule. This indicator has the average score of 74.38 for the year of 2015 and 77.5 for the year of 2016. If it is compared to the indicator of satisfaction criteria, training schedule is categorized in positive reaction because participants realize that they have useful input during the training.

Based on the training schedule, participants started the training at 07:30 – 21:00 on Monday – Saturday. On Sunday, they conducted cultural study and faith building activities at 06:00 – 16:00. In this case, participants have busy schedule because they had almost no free time during two months.

Meanwhile, other indicators can reach above 80. Improvements for the education and training are always done by organizers. For example, difference in the food taste. The taste of food in Yogyakarta tends to be sweet, while in Mahakam Ulu Regency tends to be common (not salty and not sweet). When participants gave suggestions, organizers immediately serve appropriate food.

The same goes for instructors. Participants were asked to directly give suggestion to instructors and organizers if in explaining the materials instructors speak too fast, too slow, unclear and so on. Instructors always tried to give explanation based on the context owned by participants. They naturally explained it by using language that was easily understood by participants.

Principles of openness and honesty instilled during the training provide positive impacts to the training atmosphere. A sense of belonging was built between participants, organizers, instructors and hotel staff. They care for, remind,

and help each other. On the other hand, since the training took a long time, organizers tried to make participants feel comfort. This is one of the success keys in the training.

Learning

The evaluation in the second stage is used for the learning occurred. Measurement was conducted by using pre-test and post-test that provide a difference. The average score of pre-test for the year of 2015 is 39.5. The higher score of pre-test is 6.5 while the lowest score of post-test 28. The average score of post-test is 50.8335 with the highest score of 73.33 and the lowest score is 35.00.

The test score for the year of 2016 also increased although not as many as the increase in the score for the year of 2015. The average score of pre-test for the year of 2016 is 36.8 with the highest score of 45.00 and the lowest score of 30.00. The average score of post-test for the year of 2016 is 46.2 with the highest score of 61.67 and the lowest score 33.67.

In the second stage, the education and training participants' knowledge increases. The average score of pre-test and post-test shows good increasement for the year of 2015 and 2016. Materials examined are Natural Science, Social Science, Mathematics, Civics, Indonesian Language, Science.

Based on the date above, the understanding of teachers' knowledge for the year of 2015 is relatively uneven compared to the year of 2016. This is shown by the score obtained in the pre-test and post-test before and after the training. Besides, teachers' knowledge can be observed from standard deviation of each pre-test and post-test for the year of 2015 and 2016. Based on Table 3, standard deviation for the year of 2016 is smaller than 2015.

According to Wilcoxon test, there is a difference in teachers' knowledge before after attending the training. The difference is the increase in teachers' knowledge. In 2015, the increase in average score of knowledge is 10.9833 and in 2016 the increase in average score of knowledge is 9.4. Therefore, the education and training of Mahakam Ulu Regency is considered successful.

The difference between the average score of for the year of 2015 and the year of 2016 requires to be studied. There are several causes due to the difference in knowledge level. The differences include training instructor, the monitoring of teachers' knowledge understanding, teachers' knowledge and skills before the training are different, each teacher's learning style is different. Several instructors of the education and training in 2015 are different from the education and training in 2016. First, each instructor has different teaching style, different teaching method, different ability to adapt, different teaching approach.

Second, the education and training in 2015, the the monitoring of teachers' knowledge understanding was conducted every week on Saturday. Organizers provided test related to materials given. In 2016, the monitoring of teachers' knowledge understanding was not as intensives as the previous training. Therefore, teachers did not used their time to review the material given.

Third, teachers who attended the education and training in 2015 have different initial knowledge compared to teachers who attended the training in 2016. This is shown in the average score and standard deviation of pre-test and post-test in Table 3. Teachers' skills are influenced by knowledge they acquire (Nirmala, Nurparidah, & Nopiantin, 2015). Teachers should recall their

knowledge and skills continuously so that their knowledge and skills can be internalized properly.

Brain performances can be improved in various ways (Ahmad, 2021). One of the ways is learning new thing. Knowledge can be acquired from various sources such as internet, books, journals, newspapers, and so on. Having new knowledge means that individuals who have it like to read, learn new things, search for something new.

Fourth, each teacher's learning style is different. Widharyanato (2017) states that learning styles are related to individual and a process of acquiring knowledge. In the education and training, instructors should be aware and learn each teacher's learning style. Material explanation should be adapted to each teacher's learning style (Khongpit, Sintanakul, & Nompnonkrang, 2018). The appropriate adjustment teaching style and learning style between instructors and participants will create better learning outcomes.

Conclusion

The education and trainings of elementary school teachers of Mahakam Ulu Regency in 2015 and 2016 are considered successful if the evaluation is conducted by using Kirkpatrick's Evaluation Model. In the first stage, the average positive participant's reaction to the training is high, memorable, useful, and very applicable. Training organizers should pay attention to training schedule arrangement so that participants still have a chance to enjoy their free time. In the second stage, elementary school teachers of Mahakam Ulu Regency can improve their knowledge. The average score of teachers' knowledge about elementary school materials improves after they have significant training. This can be seen in the test result showing $p.value < 0.05$.

References

- Ahmad. (2021, April 5). *yahoo.com*. Retrieved from *gamedia.com*: <https://www.gamedia.com/best-seller/meningkatkan-kecerdasan-otak/>
- Aminah, H. (2015). Model evaluasi kirkpatrick dan aplikasinya dalam pelaksanaan pelatihan (level reaksi dan pembelajaran) di pusat pendidikan dan pelatihan perum Jakarta. *Jurnal Riset Manajemen Sains Indonesia*, Vol. 6(No. 1), 376 - 384.
- Anh, V. T. (2018). Evaluation models in educational program: Strengths and weaknesses. *Journal of Foreign Studies*, 34(2), 140 - 150. doi:10.25073/2525-2445/vnufs.4252
- Buttram, J. L., & Covert, R. W. (1969). *The discrepancy evaluation model: A systematic approach for the evaluation of career planning and placement programs*. ERIC.
- Dosen, U.-U. G. (2005). Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 Tentang Guru dan Dosen. Lembaran Negara Republik Indonesia Tahun 2005 Nomor 157.
- Kholik, A. (2020). Model Kirkpatrick dalam evaluasi program pendidikan dan pelatihan penguatan kepala sekolah. *Jurnal Sipatokkong BPSDM Sulawesi Selatan*, 1(3), 219 - 226. Retrieved from <https://ojs.bpsdmsulsel.id>

- Khongpit, V., Sintanakul, K., & Nompnonkrang, T. (2018, June). The VARK learning style of the university student in computer course. *International Journal of Learning and Teaching*, 4(2), 102 - 106.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler Publisher.
- Kirkpatrick, J., & Kirkpatrick, W. K. (2009). *The Kirkpatrick four level TM: A fresh look after 50 years 1959 - 2009*. Kirkpatrick Partners, LLC.
- Muryadi, A. D. (2017, Januari). Model evaluasi program dalam penelitian evaluasi. *Jurnal Ilmiah Penjas*, 3(1), 1 - 16.
- Nirmala, S. A., Nurparidah, R., & Nopiantin, R. (2015, Juni). Perbedaan pengetahuan dan keterampilan mahasiswa dalam penanganan kegawatdaruratan maternal antara kurikulum konvensional dan kurikulum berbasis kompetensi. *IJEMIC*, 2(2), 19 - 25.
- Nuraini. (2017, Januari - Juni). Implementasi evaluasi model Kirkpatrick (Level 1 dan 2) diklat teknis substatif pemina kegiatan ekstrakurikuler keagamaan pada SMP di pusdiklat tenaga teknis pendidikan dan keagamaan. *Andragogi Jurnal Diklat Teknis*, 5(1), 34 - 53.
- Nurjanah, A. (2018, Januari - April). pengukuran keberhasilan diklat melalui model evaluasi Kirkpatrick. *Jurnal Diklat Keagamaan*, 12(32), 71 - 82.
- Pahlevi, I. (2019). Revolusi industri 4.0 dan pembangunan ekonomi berkelanjutan. *Prosiding Seminar Nasional Bagian II Pusat Penelitian Badan Keahlian DPR RI* (p. iii). Jakarta: Pusat Penelitian Badan Keahlian DPR RI.
- Provus, M. M. (1969). *The discrepancy evaluation model: An approach to local program improvement and development*. Pittsburgh, Pennsylvania: The Board of Public Education.
- Ramadhan, S. (2019). Penerapan model empat level Kirkpatrick dalam evaluasi program pendidikan dan pelatihan aparatur di pusdiklat migas. *Forum Diklat*, 6(1), 43 - 54.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Annual Meeting of the American Association of School Administrators*, (pp. 1 - 30). Atlantic City, New Jersey.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate, *International handbook of educational evaluation* (pp. 31 - 62). Dordrecht, The Netherlands: Kluwer Academic Publications.
- Topna, H. (2012, Sep - Oct). Evaluation of training and development: An analysis of various models. *Journal of Business and Management (IOSR-JBM)*, 5(2), 16 - 22. Retrieved from www.iosrjournal.org
- Utomo, A. P., & Tehupeiory, K. P. (2014). Evaluasi pelatihan dengan metode Kirkpatrick snalysis. *Jurnal Telematika*, 9(2), 37 - 41.
- Widharyanto, B. (2017). Gaya belajar model VARK dan implementasinya di dalam pembelajaran keterampilan berbahasa. *1st International Conference on Education, Language, and Arts* (pp. 1 - 16). Jakarta: Universitas Negeri Jakarta. Retrieved from <http://u.lipi.go.id/1493924774>