

## An evaluation of internship program by using Kirkpatrick evaluation model

<sup>\*1</sup>Lathifa Rosiana Dewi; <sup>2</sup>Badrun Kartowagiran

<sup>1,2</sup>Department of Educational Research and Evaluation, Graduate School of Universitas Negeri Yogyakarta

Jl. Colombo No. 1, Depok, Sleman, Yogyakarta 55281, Indonesia

<sup>\*</sup>Corresponding Author. E-mail: [lathifarosianadewi@gmail.com](mailto:lathifarosianadewi@gmail.com)

*Submitted: 20 December 2018 | Revised: 21 December 2018 | Accepted: 22 December 2018*

### Abstract

This study was aimed at evaluating an internship program using Kirkpatrick's evaluation program. The subjects of the study were students of batch 2015 and instructors. Slovin formula was used to calculate the sample. A questionnaire and teaching assessment sheet were used as instruments for collecting data. This study used content validity and exploratory factor analysis as the validity of the test. Reliability was estimated by Cronbach's Alpha. The results of this study showed that (1) in facility, the level of satisfaction was in the 'very satisfactory' category (77.01%); (2) in instructor, the level of satisfaction was in the 'very satisfactory' category (82.76%); (3) in schedule, the level of satisfaction was in the 'satisfactory' category (50.57%); (4) in material, the level of satisfaction was in the 'very satisfactory' category (89.66%); and (5) in students' teaching abilities. The improvement was in the 'very satisfactory' category.

**Keywords:** *program evaluation, internship, Kirkpatrick model*

### Introduction

Teachers' quality determines the quality of education. Teachers are said to be qualified when they have competencies to plan, teach, evaluate, guide, train, research, and conduct community service (article 39 of Law of Republic of Indonesia No. 20 of 2003). According to Jailani (2014), there are some teachers who are not qualified to teach; in public elementary schools 78.93%, private elementary schools 71.06%, public secondary schools 45.88%, private secondary schools 39.01%; public high schools 34.71%, and private high school 35.27%. This may give unfavorable effects to the educational practices in Indonesia. Meanwhile, teachers in Indonesia still have an important role in the national education. Teachers, therefore, are expected to have good competencies.

Teachers who have good competencies are believed to have good abilities in teaching. This statement is supported by Ardiansyah

(2013) who states that teachers who have good competencies can teach well. There are four competencies which should be owned by the teacher namely pedagogic competence, personal competence, professional competence, and social competence. Pedagogic competence is the teacher's competence in managing teaching and learning processes. Their ability in managing the class, arrange the students' seats, and others are examples of pedagogic competences. Personal competence is competence to influence students to have good attitudes. Professional competence is a teacher's competence in mastering the material. The last is social competence, where teachers should be able to have good interaction with the students, other teachers, and parents. These competencies can lead to the success of the teaching and learning process.

Hallo and Munadi (2014) mention the same thing that teachers have important roles in the success of teaching and learning process. The success of the teaching and learn-

ing process cannot be realized if they do not have good competencies. This can be the reason why teachers should have good competencies since they still are prospective teachers. This is aimed at making them ready when they should become teachers in the field. If they do not have good competencies, they will be just teachers who only transfer knowledge.

Nowadays, some teachers only transfer knowledge to the students. They just deliver the class material without knowing whether their students understand the material or not. Teachers should play their role to teach, evaluate the teaching and learning process, and improve anything that needs to be improved there. This should be realized by teachers from the very first time they are in the teaching and learning process. This can happen when they are trained to be a teacher while they are in the university. Each university has a program called Teaching Training Internship (TTI). This is a program where prospective teachers train to be a real teacher while they are in university.

TTI is a program which is held in the last semester of the curriculum. This program trains the prospective teachers to teach and do anything real teachers do in the classroom. This program aims to build the prospective teachers' characters so that they are ready to be teachers. Mardiyono (2006) argues that this program focusses on the prospective teachers' abilities in teaching in the classroom and doing school administration. This means that prospective teachers learn about not only how to manage the classroom and deliver the material, but also how to do school administration. It is in line with Kiggundu and Nayimuli (2009) who insist that teaching training is the activity to integrate the theory obtained from the class with practice. Some teacher training institutions implement this program, however, some do not. They have another program called internship programs.

Both internship and TTI have the same characteristics in that they train the prospective teachers to be real teachers. The aim of this internship program is to give students experience in teaching. This program lets the students in each batch teach in an addressed

school. In the initial phase, they will be trained to create lesson plans and develop class material. They will then apply what they have learned in the teaching practice in the classroom. This program is a mandatory program which means that each student teacher should take it in a year. There is an instructor who comes from the addressed school. The instructor is an English teacher at that school. The instructor should guide the student teacher how to plan a lesson, create instructional material, manage the class, and do many other things that teachers do in the classroom.

This program is divided into two. In the first semester, students should attend the debriefing. Debriefing means that students are guided to create a lesson plan, develop instructional material, and complete classroom activities. At the end of the semester, students are expected to submit the lesson plan class material. In the second semester, students do the teaching practice at the assigned school.

This program has been running for some years, but it has not been evaluated in an appropriate way. This means that the evaluation process in the program just merely gives how many students come and the strengths and weaknesses of the students. However, it has not been reported. Considering its importance, the program should be evaluated.

There are many approaches that can be applied to conduct a program evaluation. Fitzpatrick, Sanders, and Worthen (2011, p. 114) explain that the differences in evaluation approaches come from the background, experience, and worldview of the authors. This means that each approach is affected by the author. This means that an author can choose the approach which is appropriate for the evaluation process.

One of the evaluation models that can be used is Kirkpatrick's evaluation model. This model aims to evaluate the training program. There are four levels in this evaluation model namely reaction, learning, behavior, and result. Kirkpatrick and Kirkpatrick (2006, p. 21) mention that reaction assesses the satisfactory level of the program; learning assesses what knowledge has been obtained and improved; behavior assesses the changes of the

trainees' behavior after the program; and result assesses the final result, focusing on the benefit for the institution.

### Evaluation Principles

An evaluation is a systematic process which gives out information about program achievement. It means that evaluation gives information whether the objective has been achieved or not. Evaluation is a systematic process to gather data, information, and interpretation so that this can be used as the basis for policy making, decision making, or creating another program as the results of the evaluation. This can be information that can be used to revise, stop, or continue the program (Abrory & Kartowagiran, 2014).

Evaluation is different from research in terms of objectives. While research is aimed at obtaining new theories, evaluation is not. People cannot get new theories from evaluation. What people obtain from evaluation is merely information about the success of a program. Besides, evaluation can give information on the impact, or effectiveness, of a program (Stufflebeam, Madaus, & Kellaghan, 2002). It indicates that evaluation has the same method with research, but the result is really different. Research does not create a new theory but information. The information is really useful for policymaking.

In doing an evaluation process, the evaluator should follow the standards that need to be done. This is in line with Yarbrough, Shulha, Hopson, and Caruthers (2011) who believe that there are four standards that should be followed namely utility, accuracy, feasibility, and propriety. The explanation of these standards is as follows. (1) Utility means that the information which is obtained from evaluation should be useful and practical. In other words, the information can be used as a basis for decision making and for the success of the program. (2) Accuracy means that the information which is gathered should fulfill the requirements for rules of data gathering. In this case, the process of information gathering should be conducted in the right way of research in terms of instrumentation, validity, reliability, measurement, and generality. (3) Feasibility means that an evaluation study

should be proper both in the politic or cost-effectiveness. This means that, when doing an evaluation, everything should be considered. Politics means that there is no interest while doing the evaluation. For example, policy-making requires evaluation and, thus, evaluation is developed. Besides, cost-effectiveness should be considered so that there is no wasted cost. (4) Propriety means that evaluation should be done legally. This means that evaluation cannot be done in secret. The code of ethics of evaluation should be obeyed.

Evaluation is a process to measure a program, make a decision, and know the usefulness of a program. Evaluation is done when the decision maker or stakeholders are curious about the success of the program (Irambona & Kumaidi, 2015). Evaluation has an important role in the running of a program. Without evaluation, people do not know whether the program is successful or not so that follow-ups can be taken.

### Kirkpatrick's Evaluation Model

Kirkpatrick's evaluation model was employed to evaluate a training program. There are four stages in this evaluation model, including: reaction, learning, behavior, and result. These four stages can be described as follows (Kirkpatrick & Kirkpatrick, 2006, p. 21).

#### *Reaction*

In this stage, the researchers measure the level of participants' satisfaction with the program. Training programs are considered successful if the trainees are happy with the program so that they are motivated to learn. Interest, attention, and motivation of participants in following the course of training are indicators of the success of the program. In this first stage, trainees will be given a questionnaire of satisfaction on matters relating to training such as materials, instructors, training environment, and consumption in the training.

#### *Learning*

Learning can be defined as a change of attitude, improvement of knowledge, and or enhancement of the skills of the participants

after the program. There are three components to be measured in this evaluation: what knowledge has been learned, what attitude has changed, and what skills have been developed or improved. To measure all three components, then, it takes a test.

#### *Behavior*

In this evaluation, what is assessed is the attitude change of the trainees after returning from the program. The focus in this level is whether or not the trainee applies what has been obtained from the program.

#### *Result*

Evaluation at this stage is at the final stage. It is focused on the final results after the participants follow the program.

#### *Internship*

An internship is a program which is implemented in order to prepare prospective teachers to become teachers who have good skills. Inside is a professional preparation stage where a student has gained knowledge to be applied in the field with the supervision of several interested parties and within a certain period of time (Hamalik, 1990). Thus, an internship program is a program in which a student does science applications that have been obtained. In education, internship can be interpreted as the application of competences which are possessed by a teacher in school.

There are several objectives of holding an internship program of education as expressed by Hamalik (1990). These include developing a more comprehensive view to the intern about education, equipping the intern with experience about the implementation and responsibility of education as a teacher, enabling the intern to get knowledge from supervisors in school, and providing an overview to the intern about the professional code of ethics of a teacher.

In recent literature, internship is defined as an experiential learning that integrates both the theory and knowledge which are acquired in the classroom with practice (Kiser, 2016). The purpose of holding an internship is to gain valuable experience about the application

of science that has been obtained previously and make connections between the science and the field of profession based on the future career goals. Kiser (2016) mentions several important things in the internship, that is, the time spent during the internship, how time is used, the quality of the internship, and the application of the previous learning.

Based on the importance of internship evaluation, the research objective is to find out five levels of satisfaction towards the components of the internship program. These are levels of satisfaction towards (1) facilities, (2) instructors, (3) scheduling, (4) content material, and (5) students' improvement.

#### **Method**

The study was conducted in the vicinity of Muhammadiyah University of Yogyakarta (or *Universitas Muhammadiyah Yogyakarta* - UMY). Of the four Kirkpatrick's model, only two are conducted: reaction and learning. For the first level, the study is intended to find out the satisfaction level towards the program seen from facilities, instructors, scheduling, and material. For the second, the study is intended to find out the students' improvement of teaching abilities.

The subjects of the study were students of English education department batch 2015 and some instructors. The sample for this study consisted of 87 of 103 students. The number of respondents was calculated by using the Slovin formula.

A questionnaire was used to gather data about the satisfaction level towards the internship program. There were four aspects namely facilities, instructors, material, and schedule. Meanwhile, improvement of teaching abilities was obtained by using performance sheets. In addition, students and teachers in each school were interviewed to gather additional information.

The validity measures implemented in the study were of content and construct. Content validity is one which confirms what the instrument is supposed to measure (Azwar, 2015, p. 111). The questionnaire and interview guidelines were judged by three experts, and the data were subjected to the Aiken formula. All instruments were valid because the

Aiken value was higher than 0.7. It is in line with Azwar (2015, p. 149) who mentions that coefficient value can be said to be valid when the value is higher than 0.35. For the construct validity, factor analysis was used. There were four aspects in the questionnaire: facilities, instructors, schedule, and material. From the results of the construct validity measures, one item in the facilities and schedule aspect which should be dropped. The questionnaire reliability was estimated using Cronbach's Alpha. There were 36 items. The reliability value was 0.844. This can be said to be reliable.

For the quantitative data of the students' survey, the descriptive statistics proposed by Azwar (2017, p. 148) as presented in Table 1 was employed. After analyzing the quantitative data, the results were interpreted qualitatively. The results from the quantitative analyses were then cross-checked with the students and teachers before a conclusion was made.

Table 1. Normal curve statistics for students' satisfaction

Score X	Categories
$X > M + 1.5 SD$	Very satisfactory
$M + 0.5 SD < X \leq M + 1.5 SD$	Satisfactory
$M - 0.5 SD < X \leq M + 0.5 SD$	Fairly satisfactory
$M - 1.5 SD < X \leq M - 0.5 SD$	Less satisfactory
$X \leq M - 1.5 SD$	Not satisfactory

Notes:

M: Ideal mean of the concerned component in this research.

$[ M = \frac{1}{2} (\text{highest ideal score} + \text{lowest ideal score}) ]$

X: the total point scored by each respondent regarding to each item/component to evaluate.

SD: Ideal standard deviation of each component.

$[ SD = \frac{1}{6} (\text{highest ideal score} - \text{lowest ideal score}) ]$

#### Students' Satisfaction toward Facilities

In this section, each student scored five points as an ideal minimum score and the maximum ideal score was 25. Thus, the ideal mean was 15, and the standard deviation became 3.33. The facilities were judged satisfactory if the mean score belongs to the first category (Very satisfactory). The criteria are defined in Table 2.

Table 2. Evaluation criteria of facilities, instructor, schedule, and material

Score X	Categories
$X > M + 1.5 SD$	Very satisfactory
$M + 0.5 SD < X \leq M + 1.5 SD$	Satisfactory
$M - 0.5 SD < X \leq M + 0.5 SD$	Fairly satisfactory
$M - 1.5 SD < X \leq M - 0.5 SD$	Less satisfactory
$X \leq M - 1.5 SD$	Not satisfactory

#### Students' Satisfaction Level toward Instructor

There are 20 questions used in this instructor aspect. Based on the criteria, the ideal minimum score was 20 and the ideal maximum score was 100. Thus, the ideal mean was 60 and the ideal standard deviation was 13.3. The instructor was considered to be satisfied if the mean score belongs to the first category (Very satisfactory). Then, the very satisfactory category was converted to a percentage.

#### Students' Satisfaction Level toward Schedule

The schedule aspect included two questions. The ideal minimum score of this aspect was 2 and the ideal maximum score was 10. Thus, the mean ideal of this aspect was 6 and the standard deviation was 1.33. The schedule was judged to be satisfactory if the mean score belongs to the first category (Very satisfactory).

#### Students' Satisfaction Level toward Material

There were seven questions in the material aspect. Based on the criteria, the ideal minimum score was 7 and the ideal maximum score was 35. Thus, the ideal mean was 21 and the standard deviation was 4.67. The material was considered to be satisfactory if the mean score belongs to the first category (Very satisfactory).

#### Students' Teaching Ability Improvement

To measure students teaching ability improvement, the instructors were asked to fill the performance sheet. There were 'increase' and 'not increase' category. The instructor should fill the sheet by putting check marks. For 'increase' category, there were five improvement categories as mentioned in Table 3. Then, each category was converted to a percentage.

Table 3. Evaluation criteria of students' teaching ability improvement

Score X	Categories
$X > M + 1.5 SD$	Very high
$M + 0.5 SD < X \leq M + 1.5 SD$	High
$M - 0.5 SD < X \leq M + 0.5 SD$	Fairly high
$M - 1.5 SD < X \leq M - 0.5 SD$	Less high
$X \leq M - 1.5 SD$	Not high

### Findings and Discussion

Students' satisfaction becomes the most important aspect of any program. In this internship program, students' satisfaction will affect student's motivation and this can lead to the program success. Badu (2013) asserts that program effectiveness is where the training program is fun and enjoyable so that students can get a high motivation to learn.

Evaluation of reaction for the internship program was measured based on the students' satisfaction toward the program. There were 34 statements in the questionnaire, grouped into four aspects namely facilities, instructor, schedule, and material. Each aspect has a different number of statements. The facility aspect has five statements, instructor aspect has 20 statements, schedule aspect has two statements, and the material aspect has seven statements.

The indicator that represents the level of satisfaction toward the program is comfort and suitability. Comfort means that the rooms were well equipped. This can be known from the using of media, air conditioner, and air freshener. Suitability means the readiness of the room. Two statements for this suitability factor is the readiness of room before it was used and room capacity was suitable for students' number. The result showed that 77.01% of the students reported that facilities were in the very satisfactory category; 21.84% satisfactory category; and 1.15% fairly satisfactory category. Each item in the facility aspect then was categorized 'very satisfactory' and 'satisfactory'. Four items (the using of air conditioner, media, room readiness, and also room suitability to the student's number were in the very satisfactory of fresheners was only in the satisfactory category. Based on the interview, students said that the room which was used for the coaching was well equipped.

However, the using of fresheners was less. Vonny (2016) states that facilities can give satisfaction. This means that when students were asked about satisfaction, they will mention facilities aspect as one of the indicators. The implication from this study is that the better the facilities, then, the higher the increase. From the study, it can be concluded that a program can be said as satisfying where the facilities are good. The internship program can be regarded as successful because more than 50% of the students stated that the room for coaching has been equipped by the good facilities.

Instructor becomes one of the most important roles in a coaching program. The instructor should be selected carefully because they can give either good or bad effects for trainees. Instructors of internship programs need to be evaluated because they give important material before students do the teaching practice. There were 20 statements to measure the students' level of satisfaction toward the instructor. These statements include the instructor's readiness before the coaching, the delivery strategy, the delivery of materials, the ability to communicate orally, the ability to communicate in writing, and the use of media.

A total of 82.76% of the students stated that the instructor aspect was in the very satisfactory category. They mentioned that instructors' abilities in delivering the material were good so that they could understand the material well. Besides, they deliver the material in detail and a fun way. Students enjoyed joining the coaching and they could understand the material well.

This study found that students felt satisfied with the use of media and teaching video. This means that the instructor did not give them the teaching video as only an example. Some students reported that their instructor did not use the media often. This can lead to the conclusion that they just talked in the class without doing anything. Putri and Kartika (2016) report the same thing that the highest level of satisfaction was attached to instructors who have good abilities in delivering material and who can be fun too. For example, the instructor used jokes while delivering the material.

The internship program was scheduled for eight sessions in the semester. Each group had a different schedule based on the agreement between students and instructors. This was revealed by the interview with students and instructor. They said that the internship schedule was flexible so that each group had a different schedule. This means that a group may complete the internship program in only two months but the others may not.

This aspect actually included three items, but one item should be deleted due to the factor analyses. These items were the time to start the coaching and the time to end the coaching. These items can represent students' satisfaction levels because the schedule is one of the crucial things. When the coaching was not based on the schedule, this can affect the students' responses.

Students' level of satisfaction toward the schedule was only categorized by 'satisfactory'. A number of 50.57% of the students mentioned that the schedule was in the satisfactory category. Students reported that instructors used time in each coaching. Students felt useless because the coaching time did not give them any information. Zahro and Wu (2016) state that time allocation in a program should be evaluated so that there would be an improvement of the schedule for the next coaching. To anticipate the instructor who has not kept the right schedule, there should be a team for monitoring the internship program. It is in accordance with Rohani (2015) who mention that the needs of a quality control team will give a good supervisory function. Supervisors should check the coaching time in a week, for example. They cannot just come then go, but they should be there along the coaching time. This aimed to decrease the bias. It means that when students do the best to teach, then, there is no supervisor who does not come to supervise, giving students disadvantages. It is in line with Sahraini and Madya (2015) who report that teachers who have good abilities in teaching will not be appreciated because the evaluation has no regular schedule.

Material becomes one of the most important aspects of evaluation. The better the material, the better the impacts it gives to the

trainees. There are seven statements which are divided into two factors, namely material suitability with learning and material conformity to students' needs. These items are material conformity with the lesson plan, the systematics of material delivery, the interrelationship within the material, the suitability of the material with the curriculum used in the partner school, the way the selection of teaching materials, the way of choosing learning strategies, and how to manage the class.

In this study, evaluation toward material was in the very satisfactory category, as high as 89.66%. This was confirmed by the students' interviews. They mentioned that the instructor gave the lesson plan before coaching so that they knew what will be done in the coaching. Besides, the instructor gave the suitable material for them like curriculum, syllabus, and lesson plan which was used in each school. It leads to the students' understanding of what should be written in the lesson plan and what should be done in the teaching practice. In other words, the material was really useful for their needs in the teaching practice. The material in the internship program has been fitted to the students' need in both coaching and teaching practice. Utomo and Tehupeioro (2014) mention the same thing about the importance of aligning the material delivered to the students with the program objective. Program coordinators should keep this right. This means that the material which was suitable for the students' needs should be kept, while material which was not used for the internship program could be considered to be deleted.

Evaluation in learning is conducted to assess what has been learned by students, what kind of ability which improved, and what has changed (Kirkpatrick & Kirkpatrick, 2006, p. 21). This evaluation only focused on the improvement of ability in teaching. After coaching, students should do the teaching practice three times. They did the teaching practice in a class for the instructor to give them a grade.

The evaluation result was that the students' abilities in the practice teaching improved by a high-level category. This means that their teaching has changed in each time

of the teaching practice. There were five schools, coded School 1, School 2, School 3, School 4, and School 5.

From the descriptive data, it can be interpreted that 79% of students who did their teaching practice in School 1 showed improvement in their teaching abilities. Students at School 2 gave a higher score of 92%. At School 3, improvement was marked by 72%. Students at School 4 improved their teaching ability as much as 92%. Students at School 5 showed the lowest percentage of 55%. This was supported by the qualitative data from the interviews with instructors. They stated that some students have learned well but the others have not. Students who have not improved were those who did not change their way of teaching. On the average, however, it is indicated that students' ability in teaching improved by the high level of category. More than 50% of the students improved their ability in teaching by the high-level category in each school. The internship program can be said to be successful because there was an improvement in the students' teaching abilities. It is in line with Al Yahya and Norsiah (2013) who stated that ability improvement is an indication of success in a program.

## Conclusion and Recommendations

### Conclusion

Based on the findings of the reaction aspect, it can be concluded that three aspects have occupied the 'very satisfactory' category. These were facilities, instructor, and material. On the other hand, the schedule aspect did not obtain the 'very satisfactory' category. This was mostly caused by the fact that the instructor was over-timed in each coaching.

For the learning aspect, there were more than 50% of students in each school who had the 'high level' category of improvement. It can be concluded that students understood the material well so that they could apply the material learned from the coaching in the teaching practice.

### Recommendations

Some recommendations are proposed for program coordinators. The program co-

ordinators should monitor the internship program from the beginning until the end. This means that they should know what the strengths and what weaknesses of the program are. Coordinators can come to the coaching session in each school or they can just interview students about what has missed in the program.

Besides, coordinators should evaluate the internship program periodically. It has been known that evaluation can be done before the program, whilst program, and at the end of the program. It is highly recommended that the coordinators have a team for such periodical evaluations. This can prevent the program from various difficulties and weaknesses.

Coordinators, lecturers, and instructors can create the criteria of the success of the internship program. It means that there should be specific criteria to measure the success of the program. This will help them in giving a quality evaluation to the program.

## References

- Abroy, M., & Kartowagiran, B. (2014). Evaluasi implementasi Kurikulum 2013 pada pembelajaran matematika SMP negeri kelas VII di Kabupaten Sleman. *Jurnal Evaluasi Pendidikan*, 2(1), 50–59.
- Al Yahya, M. S., & Norsiah, B. M. (2013). Evaluation of effectiveness of training and development: The Kirkpatrick model. *Asian Journal of Business and Management Sciences* (Vol. 2).
- Ardiansyah, J. (2013). Peningkatan kompetensi guru bidang pendidikan di Kabupaten Tana Tidung. *EJournal Pemerintahan Integratif*, 1(1), 38–50.
- Azwar, S. (2015). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2017). *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar.
- Badu, S. Q. (2013). The implementation of Kirkpatrick's evaluation model in the learning of initial value and boundary condition problems. *International Journal of Learning and Development*, 3(5), 74–88. <https://doi.org/10.5296/ijld.v3i5.4386>



- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Boston, MA: Pearson Education.
- Hallo, D., & Munadi, S. (2014). Evaluasi kinerja guru SMA yang bersertifikat profesional di Kabupaten Halmahera Barat. *Jurnal Evaluasi Pendidikan*, 2(2), 111–122.
- Hamalik, O. (1990). *Sistem internship kependidikan teori dan praktek*. Bandung: Mandar Maju.
- Irambona, A., & Kumaidi, K. (2015). The effectiveness of English teaching program in senior high school: A case study. *REiD (Research and Evaluation in Education)*, 1(2), 114–128. <https://doi.org/10.21831/reid.v1i2.6666>
- Jailani, M. S. (2014). Guru profesional dan tantangan dunia pendidikan. *Al-Ta'lim Journal*, 21(1), 1–9. <https://doi.org/10.15548/jt.v21i1.66>
- Kiggundu, E., & Nayimuli, S. (2009). Teaching practice: A make or break phase for student teachers. *South African Journal of Education*, 29, 345–358.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs* (3rd ed.). San Francisco, CA: Berrett-Koehler.
- Kiser, P. M. (2016). *The human services internship: Getting the most from your experience*. Boston, MA: Cengage Learning.
- Law of Republic of Indonesia No. 20 of 2003 on National Education System (2003).
- Mardiyono, S. (2006). Praktik pengalaman lapangan terpadu dalam peningkatan kualitas calon guru. *Jurnal Cakrawala Pendidikan*, 25(1), 57–72. <https://doi.org/10.21831/cp.v0i1.392>
- Putri, Y. E., & Kartika, L. (2016). Evaluasi efektivitas pelatihan marketing skills pada PT XYZ. *KOLEGIAL*, 4(2), 11–22.
- Rohani, E. (2015). Analisis kepuasan peserta pelatihan pertolongan pertama gawat darurat obstetri dan neonatal (PPGDON) di balai pengembangan tenaga kesehatan (BPTK) Mataram menggunakan metode servqual. *Media Bina Ilmiah*, 9(2), 37–45.
- Sahraini, S., & Madya, S. (2015). Model evaluasi internal kompetensi guru bahasa Inggris (Model\_EIKGBI) SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(2), 156–167. <https://doi.org/10.21831/pep.v19i2.5576>
- Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2002). *Evaluation models: Viewpoints on educational and human services evaluation*. New York, NY: Kluwer Academic Publishers.
- Utomo, A. P., & Tehupeiry, K. P. (2014). Evaluasi pelatihan dengan metode Kirkpatrick analysis. *Jurnal Telematika*, 9(2), 37–41.
- Vonny, R. P. E. (2016). Pengaruh pelatihan, fasilitas kerja dan kompensasi terhadap kepuasan kerja karyawan pada PT United Tractors cabang Manado. *Jurnal Berkala Ilmiah Efisiensi*, 16(3), 407–418.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: SAGE.
- Zahro, S., & Wu, M. (2016). Implementing of the employees training evaluation using Kirkpatrick's model in tourism industry - A case study. *International Journal of Innovation and Applied Studies*, 17(3), 1042–1049.